EXTRACTING INSIGHTS FROM DIGITAL PUBLIC HEALTH DATA USING ARTIFICIAL INTELLIGENCE

EDITED BY: Yu-Dong Zhang, Hong Lin, João Manuel R. S. Tavares and
Steven Fernandes
PUBLISHED IN: Frontiers in Public Health and
Frontiers in Computer Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-873-2
DOI 10.3389/978-2-83250-873-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EXTRACTING INSIGHTS FROM DIGITAL PUBLIC HEALTH DATA USING ARTIFICIAL INTELLIGENCE

Topic Editors:
Yu-Dong Zhang, University of Leicester, United Kingdom
Hong Lin, University of Houston–Downtown, United States
João Manuel R. S. Tavares, University of Porto, Portugal
Steven Fernandes, Creighton University, United States

Citation: Zhang, Y.-D., Lin, H., Tavares, J. M. R. S., Fernandes, S., eds. (2022). Extracting Insights From Digital Public Health Data Using Artificial Intelligence. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-873-2

Table of Contents

frontiers in Public Health

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/fpubh.2021.697850

Prediction of Mental Health in Medical Workers During COVID-19 Based on Machine Learning

Xiaofeng Wang[1], *Hu Li*[1], *Chuanyong Sun*[1,2]*, *Xiumin Zhang*[3], *Tan Wang*[1], *Chenyu Dong*[1] *and Dongyang Guo*[1]

[1] *Northeast Asian Research Center, Jilin University, Changchun, China,* [2] *Kuancheng Health Commission, Changchun, China,* [3] *Department of Social Medicine and Health Management, School of Public Health, Jilin University, Changchun, China*

Mental health prediction is one of the most essential parts of reducing the probability of serious mental illness. Meanwhile, mental health prediction can provide a theoretical basis for public health department to work out psychological intervention plans for medical workers. The purpose of this paper is to predict mental health of medical workers based on machine learning by 32 factors. We collected the 32 factors of 5,108 Chinese medical workers through questionnaire survey, and the results of Self-reporting Inventory was applied to characterize mental health. In this study, we propose a novel prediction model based on optimization algorithm and neural network, which can select and rank the most important factors that affect mental health of medical workers. Besides, we use stepwise logistic regression, binary bat algorithm, hybrid improved dragonfly algorithm and the proposed prediction model to predict mental health of medical workers. The results show that the prediction accuracy of the proposed model is 92.55%, which is better than the existing algorithms. This method can be used to predict mental health of global medical worker. In addition, the method proposed in this paper can also play a role in the appropriate work plan for medical worker.

Keywords: COVID-19, mental health, prediction, machine learning, artificial intelligence, neural network, public health

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Siyuan Lu,
University of Leicester,
United Kingdom
Dimas Lima,
Federal University of Santa
Catarina, Brazil

Correspondence:
Chuanyong Sun
chuanyongsun@hotmail.com

Specialty section:
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: *20 April 2021*
Accepted: *16 August 2021*
Published: *07 September 2021*

Citation:
Wang X, Li H, Sun C, Zhang X,
Wang T, Dong C and Guo D (2021)
Prediction of Mental Health in Medical
Workers During COVID-19 Based on
Machine Learning.
Front. Public Health 9:697850.
doi: 10.3389/fpubh.2021.697850

INTRODUCTION

Although the definition of mental health is not uniform in academic circles, the research significance of mental health is self-evident. Mental health has been widely used in psychology (1), sociology (2), psychiatry (3), pedagogy (4, 5), genetics (6), and other fields.

Currently, some representative scales are usually used to measure mental health, such as Self-reporting Inventory (SCL-90) (7), Minnesota Multiphasic Personality Inventory (MMPI) (8), Self-Rating Anxiety Scale (SAS) (9), Self-Rating Depression Scale (SDS) (10), Eysenck Personality Questionnaire (EPQ) (11), the Sixteen Personality Factor Questionnaire (16PF) (12). Above scales are widely used internationally because they are guided by various psychological theories and can transform abstract mental health concepts into observable specific indicators. However, some shortcomings are not considered in the scales mentioned above. First, the different emphasis of scale measurement leads to the differences in the evaluation criteria because many factors need to be considered in the measurement of mental status.

Second, the existing way of answering the scales is self-evaluation, which inevitably makes the respondent hold something back. Third, a lot of time is spent in obtaining the results of the scale for judging mental status in emergency situations. Although the diagnosis and intervention of mental symptoms are significant, prevention is even more important. Therefore, using existing information to predict mental health is of great significance.

Mental health prediction is conductive to detecting mental disorders in advance, reducing the incidence of serious mental illnesses, and facilitating the health system to provide people with targeted health care services (13). In particular, the mental health of medical workers is seriously threatened by the global spread of COVID-19. These workers are prone to anxiety and depression (14). United Nations Secretary-General António Guterres indicated in "Message on COVID-19 and the demand for action of mental health" (15) that various mental health services must be shifted to the community and must be included in the all-people medical plan. Based on a survey conducted by WHO, COVID-19 pandemic has caused the disruption of major mental health services in 93% of countries worldwide (16). However, there are urgent demand for mental health services in many countries. In addition, the delta variants have appeared in at least 98 countries and regions, and continue to mutate and evolve. Almost all new cases of CIVID-19 are the delta variants (17), and the delta variants are becoming the main epidemic strain in many countries. The delta variant pandemic is likely to further exacerbate the fears of the public and medical workers. Therefore, predicting the potential psychological symptoms of medical workers contributes to the mental health of medical personnel, and helps maintain the high efficiency of global medical institutions.

The existing mental health prediction methods are divided into statistical model methods and artificial intelligence algorithms.

Among the statistical model methods that are used for mental health prediction, structural equation models are widely used (18–20). Moving average methods are also commonly used in health prediction. Autoregressive Integrated Moving Average model (ARIMA) (21–23) and Exponential Smoothing (ES) (24, 25) are the representative methods of the moving average model. Negative binomial model (NBM) (26, 27) and fractional polynomials (28) also provide new mentalities for predicting health. However, the statistical analysis of the data can be achieved by the above methods, but the inherent relationship between the characteristic variables and the prediction results cannot be identified by the above methods. Therefore, the accuracy of mental health prediction based on statistical models is low.

For the purpose of improving the accuracy of mental health prediction, machine learning technology has been used in the mental health prediction research since the 1980s. Basavappa et al. (29) proposed a depth-first search method according to reverse search strategy in 1996, which is used to diagnose depression or dementia. Basavappa et al. developed an expert system based on the subjects' behavior, cognition, symptoms of emotion, and neuropsychological assessment results. Gil. and Manuel (30) come up with a system according to Artificial Neural Network (ANN) and Support Vector Machine (SVM) in 2009, which is used to diagnose Parkinson's disease. The system improves the accuracy of diagnosis and reduces the cost of diagnosis. Seixas et al. (31) come up with a model called Bayesian Network (BN) in 2014, which is used to diagnose dementia and Alzheimer disease. The experimental results show that compared with most other well-known classifiers, the BN decision model has better performance. Dabek and Caban (32) proposed a neural network model in 2015, which is used to assess the possibility of suffering from psychological illnesses.

However, three problems are not solved in the algorithms mentioned above. First, statistical models are difficult to tackle the impact of random interference factors on mental health because of their limitations. Therefore, statistical models cannot reflect the high uncertainty of mental health and the non-linear relationship between feature variables and prediction results, which leads to their low prediction accuracy. Second, existing machine learning methods that are used for mental health prediction only focus on prediction accuracy without considering the impact of feature variables on the importance of mental health. Therefore, the influence weight cannot be determined according to the degree to which feature variables have effects on the prediction results. As a result, the above algorithms cannot provide a theoretical basis for the health department to work out psychological intervention plans for medical workers. Third, a large number of irrelevant or redundant features are usually included in the datasets that are used for mental health prediction. Statistical model methods merely choose significant features rather than important features, which cannot eliminate the irrelevant and redundant features in dataset that influence prediction results. Consequently, the mental health prediction methods based on statistical model not only have low prediction accuracy, but also waste computing time.

To deal with the above problems, the article proposes a novel mental health prediction algorithm called the Improved Global Chaos Bat Back Propagation Neural Network (IGCBA-BPNN). The purpose of this article is to monitor the mental health of medical workers in time to reduce the incidence of mental illness of medical workers, and to rationalize the distribution of global public health resources. Therefore, IGCBA-BPNN is applied to the mental health prediction of Chinese medical workers. The experimental results show that, compared with the existing mental health prediction methods, IGCBA-BPNN not only improves the accuracy of mental health prediction, but also selects the fewest feature variables.

The contribution of this paper is the proposal of a new mental health prediction algorithm. The proposed algorithm can predict more effectively the mental health of medical workers during COVID-19, and at the same time provides a theoretical basis for global public health departments to work out psychological intervention plans.

The remaining content of this article is arranged as followed: in section Materials and Methods, we introduced the data and methods of this research. In section Results, the effectiveness of proposed algorithm is evaluated. At last, the discussion and conclusion are illustrated in detail.

MATERIALS AND METHODS

Data Preparation

Using dataset from the "Mental Health Status of Medical Workers During COVID-19" survey conducted in Changchun, Jilin Province, China from June 1, 2020 to June 7, 2020, this paper predicts the mental health of Chinese medical workers during COVID-19. The subjects of above survey are medical workers who participated in epidemic prevention and control. According to the population status and the characteristics of geographical distribution, we selected 150 grass-roots medical units from 220 grass-roots medical units in the Changchun city and then randomly selects 35 medical workers in each grass-roots medical unit. The questionnaire is conducted online and 5,260 questionnaires were obtained in this survey. Based on research need, 152 unqualified samples were eliminated and the final sample size is 5,108. There are 32 variables in the questionnaire. In the process of designing the questionnaire, we collected as much as possible the basic information of the subjects and the variables information that may affect mental status of medical workers during COVID-19. Studies have shown that the measurable factors affecting mental status mainly include the five respects of demography (33, 34), family (35, 36), employment (37, 38), lifestyle (39, 40), and work/living environment related to COVID-19 (41, 42). Based on the results of the existing literature and the actual situation of medical workers during COVID-19, 32 factors were decided.

The description of variables is presented in **Table 1**. The data and its description are published on GitHub (https://github.com/Hu-Li/mental-health-dataset).

This study had been reviewed and approved by the Ethics Committee of the School of Public Health, Jilin University. This study does not involve questions about the identity of the respondents. An informed consent page was provided on the first page of the questionnaire for confirmation. All participants voluntarily joined this study with informed consent.

TABLE 1 | The description of variables.

Factors	Variables	Variable type
Demography	Gender	Unordered
	Age	Numeric
	Place of residence	Unordered
	Town or country	Unordered
	Education	Ordered
	Marital status	Unordered
	Chronic disease	Unordered
Family	The only child	Unordered
	Have minor children or not	Unordered
	Whether the minor child is an only child	Unordered
	Primary caregiver for children	Unordered
	Primary caregiver for elderly parents	Unordered
	Annual family income	Numeric
	Current job is supported by family	Unordered
Employment	Occupation	Unordered
	Post	Unordered
	Working years	Numeric
	Work units nature	Unordered
	Title	Unordered
	Employment type	Unordered
	Monthly income	Numeric
	Changes in work intensity	Unordered
	Working hours per week	Numeric
	Satisfaction level with the protective measures	Unordered
	Psychological training	Unordered
Lifestyle	Usual sleep time	Numeric
	Resting place	Unordered
	Frequency of exercise	Unordered
Work/living environment related to COVID-19	Have COVID-19 patients or not in the workplace	Unordered
	In close contact with COVID-19 patients in the workplace	Unordered
	Have COVID-19 patients or not in the living place	Unordered
	The work unit is a designated treatment point or not	Unordered

Feature Selection

Bat Algorithm

The Bat Algorithm (BA) (43) proposed by Yang is widely used in many fields because of its simplicity, fast convergence speed and few parameters. The bat algorithm has been used by many scholars for feature selection (44, 45). The excellent performance of the bat algorithm has also been verified in comparison with other most well-known algorithms such as genetic algorithm (GA) and particle swarm optimization (PSO) (46). Bat algorithm uses echolocation principles to simulate the predation process of bats. Bat algorithm is also an effective search method, and it is used to search for the global optimal solution. Original bat algorithm has three ideal hypotheses so as to simulate the predation behavior of bats:

First, bats use echolocation to perceive the distance between themselves and the target, and they can effectively distinguish targets and obstacles. Second, the ith bat flies randomly at a speed v_i in the space position x_i, and searches for targets with frequency f_i, wavelength λ and loudness A_i. Bats adjust the rate of emission of pulse $r(r \in [0, 1])$ according to the distance between themselves and prey. Third, loudness changes from maximum $A_{\max}$ to minimum $A_{\min}$.

Based on the above three ideal hypotheses, in the search space, the calculating equations of the frequency, velocity and position of bats as follows:

$$f_i = f_{min} + (f_{max} - f_{min}) \times \beta \tag{1}$$

$$v_i^{t+1} = v_i^t + (x_i^t - x_*) \times f \tag{2}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{3}$$

where f_i is the pulse frequency of the ith bat, and $f_{\min}$ and $f_{\max}$ are the minimum and maximum value of the pulse frequency, respectively, β is a random number within [0,1], v_i^{t+1} is the flight speed of the ith bat at the $t + 1$th iteration, v_i^t is the flight speed of the ith bat at the tth iteration, x_i^t is the position where the ith bat stays at the tth iteration, x_i^{t+1} is the position where the ith bat stays at the $t + 1$th iteration, x_* is the optimal position of the bat in the current population.

In the process of searching for prey, the initial ultrasonic loudness of bats is large, but the emission rate is low. This helps bats search for prey in the entire space. When a bat finds prey, the loudness of volume that the bat emits is gradually reduced, and the rate of emission of pulse is gradually increased. Through the above adjustments, bats can more accurately determine the location of prey. The rate of emission of pulse and the loudness of volume that the bat emits are calculated as follows:

$$r_i^{t+1} = r_i^0[1 - \exp(-\gamma \times t)] \tag{4}$$

$$A_i^{t+1} = \alpha \times A_i^t \tag{5}$$

where r_i^{t+1} is the pulse emission rate of the ith bat at the $t + 1$th iteration, r_i^0 is the maximum of pulse emission rate of the ith bat, $\gamma(\gamma > 0)$ is the enhancement coefficient of the pulse frequency, A_i^{t+1} and A_i^t are the loudness of volume that the ith bat emits at the $t + 1$th iteration and at the tth iteration, respectively, $\alpha(\alpha \in [0, 1])$ is the attenuation coefficient of the pulse loudness.

However, the bat algorithm is easy to fall into the local optimum, and the prediction accuracy of bat algorithm is low. The population initialization of the bat algorithm is randomly generated and does not have the ability to cover the entire solution space, which greatly affects the performance of the bat algorithm.

Improved Global Chaos Bat Algorithm

In order to overcome the shortcomings of the bat algorithm, Global Chaos Bat Algorithm (GCBA) (47) is introduced to eliminate redundant features and irrelevant features in the dataset. As a heuristic optimization algorithm, GCBA is used for feature selection. At first, in the initial stage, the chaotic map method is introduced to ensure the bat population traverse the entire solution space as much as possible. The chaotic map method also conducive to enriching the population diversity. Then, a fitness function based on accuracy and feature subset length is proposed to calculate the score of the feature subset after each update. Finally, GCBA selects the feature subset with the highest score from all feature subsets through the score calculated, which eliminate irrelevant features and redundant features from all feature variables.

To further improve the performance of GCBA, Improved Global Chaos Bat Algorithm (IGCBA) with higher accuracy and better performance is proposed, in which a nonlinear function based on the number of iterations is designed to balance IGCBA's exploitation and exploration capabilities. In the early stage of IGCBA, the algorithm is inclined toward the exploration capability. Global information is fully utilized to enable IGCBA to traverse the entire solution space as much as possible. In the later stage of IGCBA, the algorithm is inclined toward exploitation capability. Partial information is fully utilized to enable IGCBA to obtain the better solution through further exploitation.

Currently, the logistic method is widely used as a chaotic map method. The initial population generated by this method is diverse and can traverse the entire solution space. Therefore, in this paper, the initialization of the population is finished by using an improved logistic mapping method, and its mathematical model (48) is:

$$y_i^{d+1} = |1 - 2 \times \left(y_i^d\right)^2| \tag{6}$$

where $y_i^d(i = 1, 2, \cdots N, d = 1, 2, \cdots D)(y_i^d \in [0, 1])$ is the chaotic variable, N is the amount of bat population, and D is the dimension of initial population. Then, the position x_i^d of the bat individual in the solution space is obtained by inverse mapping of y_i^d. The calculating equation of x_i^d is:

$$x_i^d = l_i + (u_i - l_i)y_i^d \tag{7}$$

where l_i and u_i are the minimum and maximum value of the variable range, respectively.

The local optimum position of the bat and the global optimum position of the population are recorded when the position of each bat is updated. The position of the ith bat at the $t + 1$th iteration can be calculated as follows:

$$x_i^{t+1} = x_i^t + v_i^{t+1} C_1 r_1 \left(P_i - x_i^t\right) + C_2 r_2 \left(P_g - x_i^t\right) \tag{8}$$

where P_i is the local optimal position of the ith bat, P_g is the global optimal position of the bat population, r_1 and r_2 are two random numbers within [0,1].

C_1 is the control coefficient that balances the global exploration capability of IGCBA, represents the degree to which the historical optimal position of a bat individual has effect on the current state of the bat. The larger the C_1 is, the more the algorithm focuses on exploitation capability. C_2 is the control coefficient that balances the local exploitation capability of IGCBA, represents the degree to which the historical optimal position of the bat population has effect on the current state of the bat. The larger the C_2 is, the more the algorithm focuses on exploration capability.

In the preliminary stage of algorithm, it is necessary to traverse the entire solution space as much as possible to ensure that the algorithm does not converge prematurely. Therefore, in the early stage of the algorithm, C_2 should be as large as possible and C_1 as small as possible; in the later stage of the algorithm, C_1 should be as large as possible and C_2 as small as possible. In this way, the algorithm can get better performance. According to the above analysis, the calculating equation of C_1 and C_2 as follows:

$$C_1 = \begin{cases} e^{-(\frac{T}{2}-t)/10} + 0.1, & 0 \le t < 40 \\ 0.0095 \times t - 0.0980, & 40 \le t < 70 \\ 4.8 + \frac{20}{\mathrm{loge}\,(t+70)}, & 70 \le t < 100 \end{cases} \tag{9}$$

$$C_2 = \begin{cases} 0.9 - e^{-(\frac{T}{2}-t)/10}, & 0 \le t < 40 \\ 0.0095 \times t + 0.9120, & 40 \le t < 70 \\ \frac{-20}{\mathrm{loge}\,(t+70)} - 3.8, & 70 \le t < 100 \end{cases} \tag{10}$$

where t represents the current iteration times, T represents the maximum iteration times.

When initializing the bat population, we use a matrix of size $N \times D$. N is the number of bat population, D is the number of features. In this paper, a transfer equation is used to perform discrete binary operations on the bat's position. The transfer equation is:

$$S(x_i^d(t)) = \frac{1}{1 + e^{-x_i^d(t)}} \tag{11}$$

where $x_i^d(t)$ is the position of the ith bat individual in the dth dimension at the tth iteration.

The updating equation of position of the bat individual is:

$$x_i^d\,(t) = \begin{cases} 0, & rand < S(x_i^d(t)) \\ 1, & rand \ge S(x_i^d(t)) \end{cases} \tag{12}$$

where $rand$ is a random number within [0,1].

When the ith bat's position in the dth dimension at the tth iteration is 0, this bat will not be selected. When the ith bat's position in the dth dimension at the tth iteration is 1, this bat will be selected.

Back Propagation Neural Network

Back Propagation Neural Network (BPNN) is particularly suited for solving the non-linear problems (49), so it is widely used in the field of health prediction (50). In the process of back propagation of prediction errors, the connection weights and bias are constantly adjusted. Finally, the output predicted by BPNN is constantly close to the expected output.

Before using BPNN for prediction, the network needs to be trained. Through training, the network will have associative memory and predictive capabilities. The main steps of the BPNN training process are:

Step 1: Initialize the network. Based on the input and output sequence (X, Y), the number of the input layer nodes s and the output layer nodes m can be determined. The number of hidden layers and the number of the hidden layer nodes l are given by experience. The connection weight $w_{hj}(h = 1, 2, \cdots s; j = 1, 2, \cdots l)$ between the input and the hidden layer, the connection weight $w_{jk}(j = 1, 2, \cdots l; k = 1, 2, \cdots, m)$ between hidden and the output layer, the hidden layer bias value a_j and the output layer bias value b_k are initialized. Given the learning rate η, the activation function $g(x)$. In order to solve non-linear problems, the activation function usually uses the Sigmoid function, which is defined as follows:

$$g\,(x) = \frac{1}{1 + e^{-x}} \tag{13}$$

Step 2: The output of the hidden layer. The output H_j of the hidden layer is calculated based on the input vector X, ω_{hj} and a_j.

$$H_j = g(\sum\nolimits_{h=1}^{n} \omega_{hj} x_h + a_j) \tag{14}$$

Step 3: The output of the output layer. The prediction output O_k of BPNN is calculated based on H_j, ω_{jk} and b_k.

$$O_k = \sum\nolimits_{j=1}^{l} H_j \omega_{jk} + b_k \tag{15}$$

Step 4: Calculate prediction error. The prediction error of pth simple E_p is calculated based on prediction output of pth simple O_{pk} and expected output of pth simple Y_{pk}.

$$E_p = \frac{1}{2} \sum\nolimits_{k=1}^{m} (Y_{pk} - O_{pk})^2 \tag{16}$$

Step 5: Calculate the reverse transmission value. The reverse transmission value of output layer δ_k, and the reverse transmission value of hidden layer δ_j are calculated as follows:

$$\delta_k = O_{pk}(1 - O_{pk})(Y_{pk} - O_{pk}) \tag{17}$$

$$\delta_j = H_j(1 - H_j) \sum\nolimits_{k=1}^{m} \delta_k \omega_{jk} \tag{18}$$

Step 6: Update the weight. η is the learning rate, and the weight ω_{hj} and ω_{jk} are updated as follows:

$$\omega_{hj} = \omega_{hj} + \eta \delta_j x_h \tag{19}$$

$$\omega_{jk} = \omega_{jk} + \eta \delta_k H_j \tag{20}$$

Step 7: Update the bias value. The bias value a_j and b_k are updated based on δ_j and δ_k.

$$a_j = a_j + \eta \delta_j \tag{21}$$

$$b_k = b_k + \eta \delta_k \tag{22}$$

Step 7: Determine whether the algorithm iteration is over, if not, return to step 2.

Improved Global Chaos Bat Back Propagation Neural Network

Figure 1 illustrates the process of IGCBA-BPNN. First, initialize all variables. Second, IGCBA is used for feature selection to select a feature subset that can represent as much

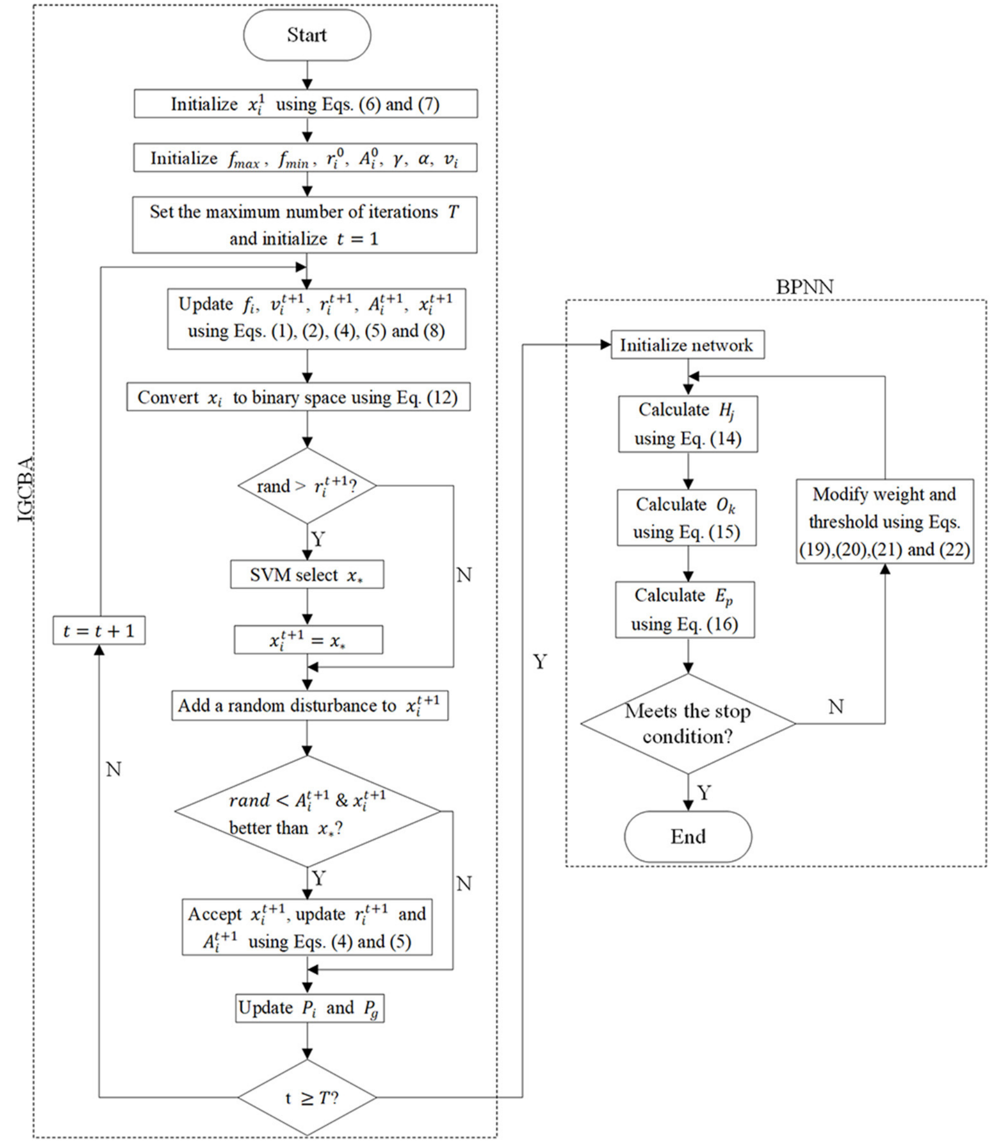


FIGURE 1 | The flowchart of IGCBA-BPNN.

information as possible of the original features and as few numbers as possible. Existing research has proved that compared with other classifiers, SVM has higher classification accuracy (51) and better stability (52). Therefore, SVM is used to judge the quality of the feature subset selected by IGCBA. Third, the features selected by IGCBA are used as the input of the BPNN to reduce the model complexity of BPNN.

RESULTS

Parameter Settings

Table 2 shows the parameter settings of feature selection algorithms. In binary bat algorithm (BBA) (46), GCBA, IGCBA, A is the loudness of volume that the bat emits and is set to 1.5, r is the rate of emission of pulse and is set to 0.5, $f_{\max}$ is the maximum value of the pulse frequency and is set to 1, and $f_{\min}$ is the minimum value of the pulse frequency and is set to 0. In GCBA, C_1 is the control coefficient, represents the degree to which the historical optimal position of a bat individual has effect on the current state of the bat. C_1 is set to 1.49618. C_2 is the control coefficient, represents the degree to which the historical optimal position of the bat population has effect on the current state of the bat. C_2 is set to 1.49618. In hybrid improved dragonfly algorithm (HIDA) (53), s and a are the separation weight and the alignment weight, respectively, and they are both set to 0.1. c is the cohesion weight and is set to 0.7. f and e are the food factor and the enemy factor, respectively, and they are both set to 1. w

TABLE 2 | The parameter settings of feature selection algorithms.

Algorithm	Parameter	Meaning	Value	References
BBA	A	Loudness	1.5	(36)
	r	The rate of emission of pulse	0.5	
	$f_{\max}$	The maximum value of frequency	1	
	$f_{\min}$	The minimum value of frequency	0	
HIDA	s	Separation weight	0.1	(38)
	a	Alignment weight	0.1	
	c	Cohesion weight	0.7	
	f	Food factor	1	
	e	Enemy factor	1	
	w	inertia weight	0.9	
GCBA	A	Loudness	1.5	(37)
	r	The rate of emission of pulse	0.5	
	$f_{\max}$	The maximum value of frequency	1	
	$f_{\min}$	The minimum value of frequency	0	
	C_1	The control coefficient	1.49618	
	C_2	The control coefficient	1.49618	
IGCBA	A	Loudness	1.5	–
	r	The rate of emission of pulse	0.5	
	$f_{\max}$	The maximum value of frequency	1	
	$f_{\min}$	The minimum value of frequency	0	
IG-bBOA	N	Number of butterflies	10	(54)
	p	Transition probability	0.8	
	a	Power exponent	0.1	
	C	Sensory modality	0.01-0.25	
	α	Increasing classification accuracy	0.99	
	β	Reducing the number of features	0.001	
	δ	Increasing the mean of mutual information	0.009	
HLBDA	s	Separation weight	0.1	(55)
	a	Alignment weight	0.1	
	c	Cohesion weight	0.7	
	f	Food factor	1	
	e	Enemy factor	1	
	w	inertia weight	0.9	
	pl	Personal learning rate	0.4	
	gl	Personal learning rate	0.7	

is the inertia weight and is set to 0.9. In information gain binary butterfly optimization algorithm (IG-bBOA) (54), N represents the number of butterflies and is set to 10, p is the transition probability and is set to 0.8, a is the power exponent and is set to 0.1, C is the sensory modality and is set to 0.01-0.25. α, β, and δ are set to 0.99, 0.001, and 0.009, respectively. In hyper learning binary dragonfly algorithm (HLBDA) (55), the parameters of s, a, c, f, e, and w are consistent with HIDA. The pl is the personal learning rate and is set to 0.4, and gl is the global learning rate and is set to 0.7.

After combining BPNN with SR, BBA, HIDA, GCBA, IGCBA, IG-bBOA, and HLBDA, the relevant parameters are set in **Tables 3**, **4**. q is the number of the hidden layer and is set to 1. p is the training goal and is set to 1,000. g is the training goal and is set to 1e-4. η is the learning rate and is set to 0.08. l is the number of the hidden layer nodes, and as a matter of experience, it is often set to half of the number of input layer nodes. The number of hidden layer nodes of SR-BPNN-4, BBA-BPNN-4, BBA-BPNN-8, HIDA-BPNN-4, HIDA-BPNN-16, GCBA-BPNN-4, GCBA-BPNN-16 and IGCBA-BPNN-4, IG-bBOA-BPNN-4, IG-bBOA-BPNN-10, HLBDA-BPNN-4, and HLBDA-BPNN-14 is set to 4, 4, 8, 4, 16, 4, 9, 4, 4, 10, 4, and 14, respectively.

Experiment Results

We make experiments to compare the IGCBA algorithm with stepwise regression (SR) (56), BBA, HIDA, GCBA, IG-bBOA, and HLBDA methods on the survey dataset in this section. At the same time, we also perform experiments to compare the

TABLE 3 | Common parameter settings in BPNN.

Parameter	Meaning	Value
q	Hidden layer	1
p	Training times	1,000
g	Training goal	1e-4
η	Learning rate	0.08

TABLE 4 | The number of hidden layer nodes in BPNN.

Algorithm	Value
SR-BPNN-4	4
BBA-BPNN-4	4
BBA-BPNN-8	8
HIDA-BPNN-4	4
HIDA-BPNN-16	16
GCBA-BPNN-4	4
GCBA-BPNN-9	9
IGCBA-BPNN-4	4
IG-bBOA-BPNN-4	4
IG-bBOA-BPNN-10	10
HLBDA-BPNN-4	4
HLBDA-BPNN-14	14

TABLE 5 | Comparison of prediction accuracy of different algorithms.

Algorithm	Number of features	The prediction accuracy
SR	10	87.45%
SR-BPNN-4		88.43%
BBA	15	88.43%
BBA-BPNN-4		88.62%
BBA-BPNN-8		89.02%
HIDA	32	88.63%
HIDA-BPNN-4		90.78%
HIDA-BPNN-16		91.18%
GCBA	18	88.49%
GCBA-BPNN-4		90.20%
GCBA-BPNN-9		90.98%
IG-bBOA	21	88.49%
IG-bBOA-BPNN-4		90.00%
IG-bBOA-BPNN-10		90.39%
HLBDA	29	88.66%
HLBDA-BPNN-4		90.39%
HLBDA-BPNN-14		90.98%
IGCBA	9	88.51%
IGCBA-BPNN-4		92.55%
IGCBA-KNN		87.52%
IGCBA-DT		79.56%

IGCBA-BPNN algorithm with SR-BPNN, BBA-BPNN, HIDA-BPNN, GCBA-BPNN, IG-bBOA-BPNN and HLBDA methods on the survey dataset. Given that BPNN, K-Nearest Neigbour (KNN) (57) and decision tree (DT) (58) are important methods for classification, we also add the comparison results of BPNN with KNN and DT. **Table 5** shows the experimental results.

Compared with SR, BBA, HIDA and GCBA, HIDA and HLBDA have the highest prediction accuracy followed by IGCBA. However, the number of features finally found by IGCBA is 23 and 20 fewer than HIDA and HLBDA, respectively. Besides, the number of features selected by IGCBA is also less than other methods. By comparing the performance of the feature selection algorithms, it can be proved that IGCBA can reduce the irrelevant and redundant features in the original features as much as possible without reducing the prediction accuracy of the classifier.

The prediction accuracy of SR-BPNN-4 is 0.98% higher than that of SR. The prediction accuracy of BBA-BPNN-4 and BBA-BPNN-8 is 0.19 and 0.59% higher than that of BBA, respectively. The prediction accuracy of HIDA-BPNN-4 and HIDA-BPNN-16 is 2.15 and 2.55% higher than that of HIDA, respectively. The prediction accuracy of GCBA-BPNN-4 and GCBA-BPNN-9 is 1.71 and 2.49% higher than that of GCBA, respectively. The prediction accuracy of IG-bBOA-BPNN-4 and IG-bBOA-BPNN-10 is 1.51 and 1.90% higher than that of IG-bBOA, respectively. The prediction accuracy of HLBDA-BPNN-4 and HLBDA-BPNN-14 is 1.73 and 2.32% higher than that of HLBDA. The prediction accuracy of IGCBA-BPNN-4 is 4.04% higher than IGCBA. The above experimental results prove that compared with the feature selection algorithms, the feature selection algorithms combined with BPNN can improve the prediction accuracy.

The prediction accuracy of SR-BPNN-4, BBA-BPNN-4, BBA-BPNN-8, HIDA-BPNN-4, HIDA-BPNN-16, GCBA-BPNN-4, GCBA-BPNN-9, IG-bBOA-BPNN-4, IG-bBOA-BPNN-10, HLBDA-BPNN-4, and HLBDA-BPNN-14 is 88.43, 88.62, 89.02, 90.78, 91.18, 90.20, 90.98, 90.00, 90.39, 90.39, and 90.98%, respectively. The prediction accuracy of IGCBA-BPNN-4 is 92.55%, which is 4.12, 3.93, 3.53, 1.77, 1.37, 2.35, 1.57, 2.55, 2.16, 2.16, and 1.57% higher than that of SR-BPNN-4, BBA-BPNN-4, BBA-BPNN-8, HIDA-BPNN-4, HIDA-BPNN-16, GCBA-BPNN-4, GCBA-BPNN-9, IG-bBOA-BPNN-4, IG-bBOA-BPNN-10, HLBDA-BPNN-4, and HLBDA-BPNN-14. The experimental results of combining each feature selection algorithm with BPNN prove that IGCBA-BPNN-4's performance is better than other algorithms. At the same time, the prediction accuracy of IGCBA-KNN and IGCBA-DT is 87.52 and 79.56%, respectively. The prediction accuracy of IGCBA-BPNN-4 is 5.03 and 12.99% higher than that of IGCBA-KNN and IGCBA-DT, respectively. It can be proved that BPNN is better than KNN and DT for classification on survey dataset. Therefore, IGCBA-BPNN-4 model has good applicability in predicting the mental health of medical workers in public health events.

For the purpose of better verifying the superior convergence performance of the IGCBA algorithm on the test dataset, **Figure 2** shows the convergence performance of the six algorithms. By directly plotting the classification accuracy curve

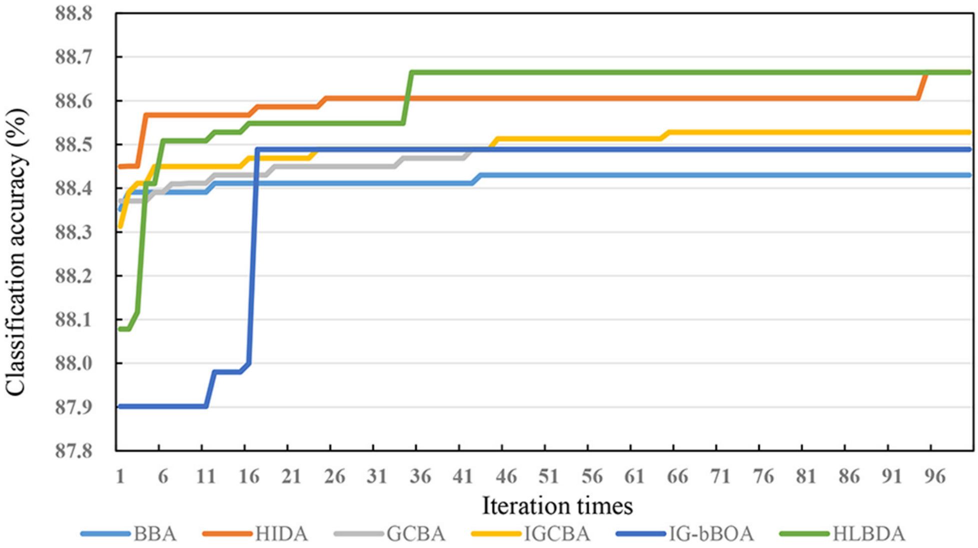


FIGURE 2 | Convergence curves of the six algorithms on the survey dataset.

with the iteration times, we can see that the classification accuracy increases monotonously at each iteration until level off. **Figure 2** shows that GCBA converges faster than BBA. From **Figure 2**, it can be analyzed that GCBA does not solve the shortcoming that BBA falls into the local optimal solution easily. IGCBA falls into the local optimal solution at the 46th iteration, and it jumped out of the local optimal solution at the 66th iteration. Although the ability of IGCBA to jump out of the local optimal solution is not as good as HIDA and HLBDA, it is significantly better than IG-bBOA, BBA and GCBA.

Figure 3 shows that the alteration trend of the number of features selected by the six algorithms in the survey dataset with the number of iterations. Since the non-linear equation balances the exploitation and exploration capabilities of IGCBA, IGCBA has strong exploitation capabilities in the later stage. Therefore, IGCBA finds fewer features at the 65th iterations. Particularly, although the prediction accuracy of HIDA and HLBDA in **Figure 2** is 0.12 and 0.12% higher than that of IGCBA, the number of features finally found by IGCBA is 23 and 20 fewer than HIDA and HLBDA. Combining **Figures 2**, **3**, the experimental results show that IGCBA has strong exploitation ability and superior performance in the later optimization stage.

Analysis of the Degree of Feature Variables on Mental Health

Mean Impact Value (MIV) is currently considered to be one of the best algorithms for evaluating the correlation between input variables and output variables. Sorting the variables according to MIV's absolute value can determine the degree of influence of input variables on network output variables. The symbol of the MIV value represents the relative direction, and the relative importance of the impact is represented by MIV's absolute value.

The IGCBA-BPNN-4 prediction model eliminates irrelevant and redundant features in the original dataset, decreases model running time, and improves the prediction accuracy of the classifier. At the same time, the feature variables that affect mental health are sorted according to the degree of their importance. **Table 6** shows that IGCBA-BPBB-4 selects a total of nine feature variables that affect mental health. The nine feature variables are "Have patients with COVID-19 or not in the living place," "age," "employment type," "Have patients with COVID-19 or not in the workplace," "the work unit is a designated treatment point or not," "changes in work intensity," "usual sleep time," "place of residence," and "marital status." We analyze the factors affecting mental health according to their degree of importance.

Variables in statistics are divided into numerical variables and categorical variables. When considering the impact of input variables on output variables, the direction of the symbol is only meaningful for numerical variables, and has no meaning for categorical variables. Since the variables in this article are mostly categorical variables, the positive or negative influence of the symbol is not considered in this analysis.

In the community transmission stage of the epidemic, according to a study that cluster transmission occurs in multiple communities and families. On average, each patient transmits the infection to 2.2 people (59). When relatives, friends, and nearby people in the living place are determined to be suspected or confirmed cases, people will have psychological problems such as fear and anxiety due to fear of infection.

Patients with COVID-19 are mostly elderly people. Under normal circumstances, the deterioration of body function with age decreases the health levels of the elderly. The elderly are more vulnerable to the threat of diseases because their immune system is relatively weak. In the "Questions and Answers About COVID-19 and the Elderly" on the WHO official website, a

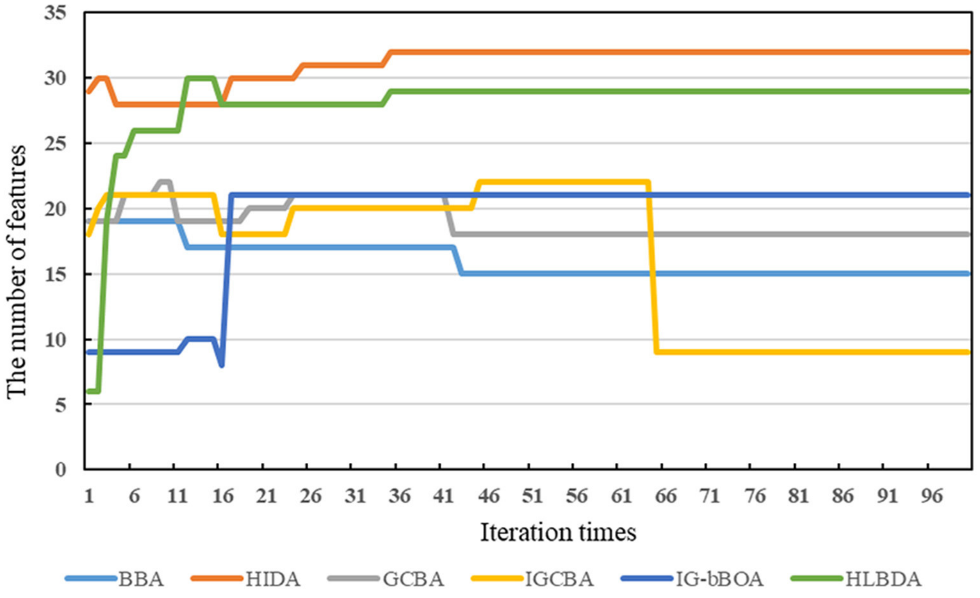


FIGURE 3 | The feature numbers curves of the six algorithms on the survey dataset.

clear answer is also given to the question "Who is at risk of severe illness," that is, the elderly and all ages of people who are diagnosed with diseases such as hypertension, heart disease, lung disease, diabetes or cancer are more likely to suffer from severe illness than others (60).

Differences in employment type lead to differences in the psychological status of medical workers. In contrast with formal medical personnel, temporarily hired medical personnel may show a stronger sense of anxiety and fear during COVID-19. On the one hand, due to the absence of both manpower capital and social capital, temporary medical workers are more likely to engage in low-tech and labor-intensive jobs. The work pressure caused by high labor intensity make easily temporary medical workers prone to anxiety and hostility. On the other hand, most temporary medical workers are exposed to such a severe epidemic for the first time. They lack the work experience and sufficient mental preparation to deal with severe infectious diseases. At the same time, due to the lack of objective cognition of COVID-19, they are in a highly alert state at work, and their anxiety and fear are more prominent.

There are patients with COVID-19 in the workplace, especially the workplace is a designated treatment point for COVID-19, which will have a greater impact on the mental status of medical workers. In face of high-intensity work pressure and the risk of being infected, medical workers are more likely to become a high-risk group with psychological symptoms. Less sleep and poor sleep during COVID-19 can cause sleep disorders, and sleep disorders are often accompanied by symptoms such as depression, tension, anxiety, hostility and irritability (61). For people who do not have a spouse, they cannot get timely help when they encounter difficulties and need a good listener. They are prone to anxiety and depression (62). The farther the place of residence is from the city center, the lower the population density. It is difficult for COVID-19 to spread rapidly in rural areas (63), and people living in rural areas have less fear of COVID-19 than people living in cities.

DISCUSSION

According to the above observations, we can make a conclusion that the performance of IGCBA-BPNN-4 is better than other algorithms. First, BPNN learns the non-linear relationship between feature variables and prediction results, which improves the accuracy of mental health prediction. The results in **Table 5** indicate that the accuracy of the feature selection algorithms combined with BPNN is higher than that of the feature selection algorithms without BPNN, with an average increase of 2.46%. Particularly, the accuracy of IGCBA-BPNN-4 is 4.04% higher than that of IGCBA. Second, the value calculated by MIV is used as the influence weight, which assesses the extent to which feature variables contribute to mental health. It can be seen from **Table 6** that through the calculation of MIV, the nine feature variables that affect mental health are sorted by their importance. The top three important factors affecting mental health are "whether there are patients with COVID-19 in the workplace," "age" and "employment type." The result corresponds with our expectations. Third, GCBA eliminates irrelevant and redundant features in the original features, which

TABLE 6 | Feature variables that affect mental health: sorted by importance.

Feature	MIV value	Sorted by importance
Have patients with COVID-19 or not in the living place	−0.11158	1
Age	0.09567	2
Employment type	0.09161	3
Have patients with COVID-19 or not in the workplace	−0.03947	4
The work unit is a designated treatment point or not	0.03899	5
Changes in work intensity	−0.03493	6
Usual sleep time	−0.02933	7
Place of residence	0.01904	8
Marital status	0.01328	9

reduces BPNN's complexity. The results in **Table 5** indicate that GCBA reduces the number of features in the survey dataset from 32 to 18. Although GCBA selects more features than SR and BBA, it has higher prediction accuracy. Fourth, the non-linear equation in IGCBA balances the exploitation and exploration capabilities of IGCBA, which accelerates the convergence speed of IGCBA and prevents IGCBA from falling into a local optimal solution. It can be seen from **Figures 2**, **3** that IGCBA does not fall into the local optimal solution due to its certain exploration capabilities in the later stage. As a result, IGCBA obtains a feature subset that can represent as much information as possible of the original features and as few features as possible. The number of features selected by IGCBA is only half of the number of features selected by GCBA. Besides, the prediction accuracy rate of IGCBA is higher than that of GCBA.

It should be pointed out that although many people have been vaccinated against COVID-19, the COVID-19 epidemic is far from over due to the spread of mutant strains. COVID-19 directly endangers people's lives, and it is extremely important to diagnose COVID-19 quickly and accurately. The latest method proposed by Wang et al. (64, 65) may help diagnose COVID-19 more quickly and effectively. In the fight against COVID-19, when the psychological symptoms of medical workers are discovered and intervened in time, the work efficiency of the entire health system will be improved. The algorithm proposed in this article can more effectively predict the mental health of medical staff, and the research results can also be directly used by global public health departments. However, several limitations also exist in our research. First, the data in this article was obtained through an online survey, and this research is an observational study. As a result, self-report problems and recall biases are inevitable to some extent. Secondly, mental health is affected by personal, family, economic, social environment and other factors. The factors affecting mental health in this article are incomprehensive. Finally, some parameters that are set manually are used in our algorithm. The parameters of the neural network are given by experience rather than obtained from adaptive changes or learning. We will solve this problem in future work.

CONCLUSIONS

The accuracy of existing mental health prediction methods is low because the relationship between the feature variables and the prediction results is non-linear and the prediction dataset contains a lot of irrelevant and redundant features. At the same time, current mental health prediction methods cannot estimate the extent to which the feature variables are important to the prediction results. Therefore, this paper proposes IGCBA-BPNN. First, BPNN is introduced to deal with the non-linear problem between prediction results and feature variables, which improves the accuracy of mental health prediction. Second, MIV is introduced to calculate the influence weight, which assesses the extent to which feature variables contribute to mental health. Third, GCBA is introduced to eliminate redundant and irrelevant features in the original features, which reduces the model complexity of BPNN and improves the performance of BPNN. Fourth, a non-linear equation is designed in IGCBA to speeds up the convergence speed of IGCBA and prevents IGCBA from falling into a local optimal solution. Experiment results show that the performance of IGCBA-BPNN is better than existing algorithms. The IGCBA-BPNN prediction model can obtain good results in mental health prediction.

However, IGCBA only reduces BPNN's input dimension. The BPNN's structure is not improved, and the parameters in the BP network is not optimized. Therefore, how to ascertain the number of neural network nodes is an important challenge in the future.

In a word, with the development of swarm intelligence algorithms and neural network technology, the methods based on swarm intelligence algorithms combined with neural networks are playing an increasingly significant role in the field of prediction. In the future health prediction research, the prediction method based on swarm intelligence algorithm combined with neural network will have a wider application prospect.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in an online repository: https://github.com/Hu-Li/mental-health-dataset.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the School of Public Health, Jilin University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

XW and HL came up with the original idea. HL and TW designed this study and provided research methods. CS and XZ completed the data collection and performed the statistical analysis. TW conducted the experiments. XW supervised the research. HL drafted the manuscript. XW, HL, DG, and CD improved the manuscript. All authors contributed to the article and approved the final version.

ACKNOWLEDGMENTS

We sincerely thank all participants for their support in this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.697850/full#supplementary-material

REFERENCES

1. Angehrn A, Sapach MJNT, Ricciardelli R, MacPhee RS, Anderson GS, Carleton RN. Sleep quality and mental disorder symptoms among Canadian public safety personnel. *Int J Environ Res Public Health*. (2020) 17:2708. doi: 10.3390/ijerph17082708
2. Lee MH, Seo MK. Community integration of persons with mental disorders compared with the general population. *Int J Environ Res Public Health*. (2020) 17:1596. doi: 10.3390/ijerph17051596
3. Tang J, Wu L, Huang HL, Feng J, Yuan YF, Zhou YP, et al. Back propagation artificial neural network for community Alzheimer's disease screening in China. *Neural Regen Res*. (2013) 8:270-6. doi: 10.3969/j.issn.1673-5374.2013.03.010
4. Francis L, Barling J. Organizational injustice and psychological strain. *Can J Behav Sci*. (2005) 37:250-61. doi: 10.1037/h0087260
5. Ko HC. A sustainable approach to mental health education: an empirical study using Zhuangzi's self-adaptation. *Sustainability*. (2019) 11:3677. doi: 10.3390/su11133677
6. Pirooznia M, Seifuddin F, Judy J, Mahon PB, Potash JB, Zandi PP, et al. Data mining approaches for genome-wide association of mood disorders. *Psychiat Genet*. (2012) 22:55-61. doi: 10.1097/YPG.0b013e32834dc40d
7. Derogatis LR, Melisaratos N. The brief symptom inventory-an introductory report. *Psychol Med*. (1983) 13:595-605. doi: 10.1017/S0033291700048017
8. Hathaway SR, McKinley JC. A multiphasic personality schedule (Minnesota): I. construction of the schedule. *J Psychol*. (1940) 10:249-54. doi: 10.1080/00223980.1940.9917000
9. Zung WWK. A rating instrument for anxiety disorders. *Psychosomatics*. (1971) 12:371-9. doi: 10.1016/S0033-3182(71)71479-0
10. Zung WWK. A self-rating depression scale. *Arch Gen Psychiat*. (1965) 12:63-70. doi: 10.1001/archpsyc.1965.01720310065008
11. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Pers Indiv Differ*. (1985) 6:21-9. doi: 10.1016/0191-8869(85)90026-1
12. Cattell RB. Scree test for number of factors. *Multivar Behav Res*. (1966) 1:245-76. doi: 10.1207/s15327906mbr0102_10
13. Center for Global Development. *Mapping and Realigning Incentives in the Global Health Supply Chain*. Available online at: http://www.cgdev.org/doc/DemandForecasting/Principles.pdf (accessed September 24, 2020).
14. Cosic K, Popovic S, Sarlija M, Kesedzic I, Jovanovic T. Artificial intelligence in prediction of mental health disorders induced by the COVID-19 pandemic among health care workers. *Croat Med J*. (2020) 61:279-88. doi: 10.3325/cmj.2020.61.279
15. Inter-Agency Standing Committee. *António Guterres (UN Secretary-General) on COVID-19 and the Need for Action on Mental Health*. Available online at: https://interagencystandingcommittee.org/iasc-reference-group-mental-health-and-psychosocial-support-emergency-settings/antonio-guterres-un (accessed June 12, 2020).
16. World Health Organization. *The Impact of COVID-19 on Mental, Neurological and Substance Use Services: Results of a Rapid Assessment*. Available online at: https://www.who.int/docs/default-source/mental-health/ppt-who-covid19-mental-health-rapid-assessment-v10.pdf?sfvrsn=2f45b88a_2 (accessed August 20, 2020).
17. O'Dowd A. Covid-19: cases of delta variant rise by 79%, but rate of growth slows. *BMJ*. (2021) 373:n1596. doi: 10.1136/bmj.n1596
18. Yang X, Jin M, Zheng L. The prediction model of college students' life stressor on mental health. *China J Health Psychol*. (2018) 26:775-8. doi: 10.13342/j.cnki.cjhp.2018.05.001
19. Margraf J, Zhang XC, Lavallee KL, Schneider S. Longitudinal prediction of positive and negative mental health in Germany, Russia, and China. *PLoS ONE*. (2020) 15:e0234997. doi: 10.1371/journal.pone.0234997
20. Wilson MG, DeJoy DM, Vandenberg RJ, Richardson HA, McGrath AL. Work characteristics and employee health and well-being: test of a model of healthy work organization. *J Occup Organ Psych*. (2004) 77:565-88. doi: 10.1348/0963179042596522
21. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. *Emerg Med J*. (2012) 29:358-65. doi: 10.1136/emj.2010.103531
22. Champion R, Kinsman LD, Lee GA, Masman KA, May EA, Mills TM, et al. Forecasting emergency department presentations. *Aust Health Rev*. (2007) 31:83-90. doi: 10.1071/AH070083
23. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak*. (2003) 3:2. doi: 10.1186/1472-6947-3-2
24. Medina DC, Findley SE, Guindo B, Doumbia S. Forecasting non-stationary diarrhea, acute respiratory infection, and malaria time-series in Niono, Mali. *PLoS ONE*. (2007) 2:e1181. doi: 10.1371/journal.pone.0001181
25. Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecasting*. (2002) 18:439-54. doi: 10.1016/S0169-2070(01)00110-8
26. Soyiri I, Reidpath D, Sarran C. Determinants of asthma length of stay in London hospitals: individual versus area effects. *Emerg Health Threats J*. (2011) 4:143. doi: 10.3402/ehtj.v4i0.11179
27. Soyiri I, Reidpath D, Sarran C. Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors. *PLoS ONE*. (2011) 6:e27184. doi: 10.1371/journal.pone.0027184
28. Williams JS. Assessing the suitability of fractional polynomial methods in health services research: a perspective on the categorization epidemic. *J Health Serv Res Po*. (2011) 16:147-52. doi: 10.1258/jhsrp.2010.010063
29. Basavappa SR, Rao SL, Harish B. Expert system for dementia/depression diagnosis. *Nimhans J*. (1996) 14:99-106.
30. Gil D, Manuel DJ. Diagnosing Parkinson by using artificial neural networks and support vector machines. *J Comput Sci Technol*. (2009) 9:63-71. https://lup.lub.lu.se/record/1776690
31. Seixas FL, Zadrozny B, Laks J, Conci A, Saade DCM. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. *Comput Biol Med*. (2014) 51:140-58. doi: 10.1016/j.compbiomed.2014.04.010
32. Dabek F, Caban JJ. A neural network based model for predicting psychological conditions. In: *Proceedings of the 2015 8th International Conference on Brain Informatics and Health (BIH)*. London (2015). p. 252-61.
33. Liu CY, Yang YZ, Zhang XM, Xu XY, Dou QL, Zhang WW. The prevalence and influencing factors in anxiety in medical workers fighting COVID-19 in China: a cross-sectional survey. *Epidemiol Infect*. (2020) 148:e98. doi: 10.1017/S0950268820001107

34. Lei L, Huang XM, Zhang S, Yang JR, Yang L, Xu M. Comparison of prevalence and associated factors of anxiety and depression among people affected by versus people unaffected by quarantine during the COVID-9 epidemic in Southwestern China. *Med Sci Monitor*. (2020) 26:e924609. doi: 10.12659/MSM.924609
35. Covinsky KE, Newcomer R, Fox P, Wood J, Sands L, Dane K, et al. Patient and caregiver characteristics associated with depression in caregivers of patients with dementia. *J Gen Intern Med*. (2003) 18:1006-14. doi: 10.1111/j.1525-1497.2003.30103.x
36. Lu JJ, Kong JX, Song JS, Li L, Wang HM. The health-related quality of life of nursing workers: a cross-sectional study in medical institutions. *Int J Nues Pract*. (2019) 25:e12754. doi: 10.1111/ijn.12754
37. Gomez MAL, Sabbath E, Boden L, Williams JAR, Hopcia K, Hashimoto D, et al. Organizational and psychosocial working conditions and their relationship with mental health outcomes in patient-care workers. *J Occup Environ Med*. (2019) 61:e480-5. doi: 10.1097/JOM.0000000000001736
38. Maunder RG, Lancee WJ, Balderson KE, Bennett JP, Borgundvaag B, Evans S. Long-term psychological and occupational effects of providing hospital healthcare during SARS outbreak. *Emerg Infect Dis*. (2006) 12:1924-32. doi: 10.3201/eid1212.060584
39. Lopresti AL, Hood SD, Drummond PD. A review of lifestyle factors that contribute to important pathways associated with major depression: diet, sleep and exercise. *J Affect Disord*. (2013) 148:12-27. doi: 10.1016/j.jad.2013.01.014
40. Hoare E, Milton K, Foster C, Allender S. The associations between sedentary behaviour and mental health among adolescents: a systematic review. *Int J Behav Nutr Phy*. (2016) 13:108. doi: 10.1186/s12966-016-0432-4
41. Lai JB, Ma SM, Wang Y, Cai ZX, Hu JB, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. *JAMA Netw Open*. (2020) 3:e203976. doi: 10.1001/jamanetworkopen.2020.3976
42. Lu W, Wang H, Lin YX, Li L. Psychological status of medical workforce during the COVID-19 pandemic: a cross-sectional study. *Psychiat Res*. (2020) 288:112936. doi: 10.1016/j.psychres.2020.112936
43. Yang, XS. (2010) A New Metaheuristic Bat-Inspired Algorithm. In: González J.R., Pelta D.A., Cruz C., Terrazas G., Krasnogor N. (eds) *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*. Studies in Computational Intelligence, vol 284. Berlin, Heidelberg:Springer, p. 65–74. https://doi.org/10.1007/978-3-642-12538-6_6
44. Nakamura RYM, Pereira LAM, Costa KA, Rodrigues D, Papa JP, Yang XS. BBA: a binary bat algorithm for feature selection. In: *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*. Ouro Preto: IEEE (2012). p. 291-7. https://ieeexplore.ieee.org/abstract/document/6382769.
45. Rodrigues D, Pereira LAM, Nakamura RYM, Costa KAP, Yang XS, Souza AN. A wrapper approach for feature selection and optimum-path forest based on bat algorithm. *Expert Syst Appl*. (2014) 41:2250-8. doi: 10.1016/j.eswa.2013.09.023
46. Mirjalili S, Mirjalili SM, Yang X. Binary bat algorithm. *Neural Comput Appl*. (2014) 25:663-81. doi: 10.1007/s00521-013-1525-5
47. Cui X, Li Y, Fan J. Global chaotic bat optimization algorithm. *J Northeast Univ (Nat Sci)*. (2020) 41:488-91. doi: 10.12068/j.issn.1005-3026.2020.04.006
48. Kazimipour B, Li XD, Qin AK. A review of population initialization techniques for evolutionary algorithms. In: *Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC)*. Beijing (2014). p. 2585-92.
49. Grossi E, Buscema M. Introduction to artificial neural networks. *Eur J Gastroen Hepat*. (2007) 19:1046-54. doi: 10.1097/MEG.0b013e3282f198a0
50. Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. *Ann Roy Coll Surg*. (2004) 86:334-8. doi: 10.1308/147870804290
51. Huang CL, Chen MC, Wang CJ. Credit scoring with a data mining approach based on support vector machines. *Expert Syst Appl*. (2007) 33:847-56. doi: 10.1016/j.eswa.2006.07.007
52. Subasi A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Comput Biol Med*. (2013) 43:576-86. doi: 10.1016/j.compbiomed.2013.01.020
53. Cui X, Li Y, Fan J, Wang T, Zheng Y. A hybrid improved dragonfly algorithm for feature selection. *IEEE Access*. (2020) 8:155619-29. doi: 10.1109/ACCESS.2020.3012838
54. Sadeghian Z, Akbari E, Nematzadeh H. A hybrid feature selection method based on information theory and binary butterfly optimization algorithm. *Eng Appl Artif Intel*. (2021) 97:104079. doi: 10.1016/j.engappai.2020.104079
55. Too J, Mirjalili S. A hyper learning binary dragonfly algorithm for feature selection: a COVID-19 case study. *Knowl Based Syst*. (2021) 212:106553. doi: 10.1016/j.knosys.2020.106553
56. Lucero RJ, Lindberg DS, Fehlberg EA, Bjarnadottir RI, Li Y, Cimiotti JP, et al. A data-driven and practice-based approach to identify risk factors associated with hospital-acquired falls: applying manual and semi- and fully-automated methods. *Int J Med Inform*. (2019) 122:63-9. doi: 10.1016/j.ijmedinf.2018.11.006
57. Peng NB, Zhang YX, Zhao YH. A SVM-kNN method for quasar-star classification. *Sci China Phys Mech*. (2013) 56:1227-34. doi: 10.1007/s11433-013-5083-8
58. Bui DT, Pradhan B, Lofman O, Revhaug I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naive bayes models. *Math Probl Eng*. (2012) 2012:974638. doi: 10.1155/2012/974638
59. Chan JFW, Yuan SF, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. (2020) 395:514-23. doi: 10.1016/S0140-6736(20)30154-9
60. World Health Organization. *Coronavirus Disease (COVID-19): Risks and Safety for Older People*. Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-risks-and-safety-for-older-people (accessed September 22, 2020).
61. Metse AP, Fehily C, Clinton-McHarg T, Wynne O, Lawn S, Wiggers J, et al. Self-reported suboptimal sleep and receipt of sleep assessment and treatment among persons with and without a mental health condition in Australia: a cross sectional study. *BMC Public Health*. (2021) 21:463. doi: 10.1186/s12889-021-10504-6
62. Skapinakis P, Bellos S, Koupidis S, Grammatikopoulos I, Theodorakis PN, Mavreas V. Prevalence and sociodemographic associations of common mental disorders in a nationally representative sample of the general population of Greece. *BMC Psychiatry*. (2013) 13:163. doi: 10.1186/1471-244X-13-163
63. Eilersen A, Sneppen K. SARS-CoV-2 superspreading in cities vs the countryside. *APMIS*. (2021) 129:401-7. doi: 10.1111/apm.13120
64. Wang SH, Satapathy SC, Anderson D, Chen, SX, Zhang YD. Deep fractional max pooling neural network for COVID-19 recognition. *Front Public Health*. (2021) 9:1117. doi: 10.3389/fpubh.2021.726144
65. Wang SH, Zhang Y, Cheng X, Zhang X, Zhang YD. PSSPNN: PatchShuffle Stochastic Pooling Neural Network for an explainable diagnosis of COVID-19 with multiple-way data augmentation. *Comput Math Methods Med*. (2021) 2021:6633755. doi: 10.1155/2021/6633755

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 22 September 2021
doi: 10.3389/fcomp.2021.672867

Discovering Key Topics From Short, Real-World Medical Inquiries *via* Natural Language Processing

***A. Ziletti*[1]*, *C. Berns*[2], *O. Treichel*[3], *T. Weber*[3], *J. Liang*[4], *S. Kammerath*[5], *M. Schwaerzler*[1], *J. Virayah*[6], *D. Ruau*[1], *X. Ma*[7] and *A. Mattern*[4]**

[1]Department of Decision Science and Advanced Analytics, Bayer AG, Berlin, Germany, [2]Department of Data Science and Data Engineering, Areto Consulting Gmbh, Cologne, Germany, [3]Department of Product Platforms, Bayer AG, Berlin, Germany, [4]Department of Medical Information, Bayer AG, Berlin, Germany, [5]Department of Medical Affairs and Pharmacovigilance, Bayer AG, Berlin, Germany, [6]Department of Medical Affairs and Pharmacovigilance Digital Transformation, Bayer AG, Berlin, Germany, [7]Department of Integrated Evidence Generation and Business Excellence, Bayer AG, Berlin, Germany

Millions of unsolicited medical inquiries are received by pharmaceutical companies every year. It has been hypothesized that these inquiries represent a treasure trove of information, potentially giving insight into matters regarding medicinal products and the associated medical treatments. However, due to the large volume and specialized nature of the inquiries, it is difficult to perform timely, recurrent, and comprehensive analyses. Here, we combine biomedical word embeddings, non-linear dimensionality reduction, and hierarchical clustering to automatically discover key topics in real-world medical inquiries from customers. This approach does not require ontologies nor annotations. The discovered topics are meaningful and medically relevant, as judged by medical information specialists, thus demonstrating that unsolicited medical inquiries are a source of valuable customer insights. Our work paves the way for the machine-learning-driven analysis of medical inquiries in the pharmaceutical industry, which ultimately aims at improving patient care.

Keywords: natural language processing, machine learning, medical inquiries, clustering, medical information, topic discovery

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Angelo D'Ambrosio,
Freiburg University Medical Center, Germany
Hsin-Yao Wang,
Linkou Chang Gung Memorial Hospital, Taiwan

****Correspondence:***
A. Ziletti
angelo.ziletti@bayer.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science.

Received: *26 February 2021*
Accepted: *06 September 2021*
Published: *22 September 2021*

Citation:
Ziletti A, Berns C, Treichel O, Weber T, Liang J, Kammerath S, Schwaerzler M, Virayah J, Ruau D, Ma X and Mattern A (2021) Discovering Key Topics From Short, Real-World Medical Inquiries via *Natural Language Processing. Front. Comput. Sci. 3:672867. doi: 10.3389/fcomp.2021.672867*

INTRODUCTION

Every day pharmaceutical companies receive numerous medical inquiries related to their drugs from patients, healthcare professionals, research institutes, or public authorities from a variety of sources (*e.g.,* websites, e-mail, phone, social media channels, company personnel, telefax). These medical inquiries may relate to drug-drug-interactions, availability of drugs, side effects of pharmaceuticals, clinical trial information, product quality issues, comparison with competitor products, storage conditions, dosing regimen, and the like. On the one hand, a single medical inquiry is simply a question of a given person searching for a specific information related to a medicinal product. On the other hand, a plurality of medical inquiries from different persons may provide useful insight into matters related to medicinal products and associated medical treatments. Examples of these insights could be early detection of product quality or supply chain issues, anticipation of treatment trends and market events, improvement of educational material and standard answers/frequently asked question coverage, potential changes in treatment pattern, or even suggestions on new possible indications to investigate. From a strategic perspective, this information could enable organizations to make better decisions, drive organization results, and more broadly create benefits for the healthcare community.

However, obtaining high-level general insights is a complicated task since pharmaceutical companies receive copious amounts of medical inquiries every year. Machine learning and natural language processing represent a promising route to automatically extract insights from these large amounts of unstructured (and noisy) medical text. Natural language processing and text mining techniques have been widely used in the medical domain (Allahyari et al., 2017; Luque et al., 2019), with emphasis on electronic health records (Sun et al., 2017; Landi et al., 2020; Mascio et al., 2020; Kormilitzin et al., 2021). In particular, deep learning has been successfully applied to medical text, with the overwhelming majority of works in supervised learning, or representation learning (in a supervised or self-supervised setting) to learn specialized word vector representations (*i.e.* word embeddings) (Alsentzer et al., 2019; Beltagy et al., 2019; Neumann et al., 2019; Weng and Szolovits, 2019; Wu et al., 2019). Conversely, the literature on unsupervised learning for medical text is scarce despite the bulk of real-world medical text being unstructured, without any labels or annotations. Unsupervised learning from unstructured medical text is mainly limited to the development of topic models based on latent Dirichlet allocation (LDA) (Blei et al., 2003). Examples of applications in the medical domain are clinical event identification in brain cancer patients from clinical reports (Arnold and Speier, 2012), modeling diseases (Pivovarov et al., 2015) and predicting clinical order patterns (Chen et al., 2017) from electronic health records, or detecting cases of noncompliance to drug treatment from patient forums (Abdellaoui et al., 2018). Only recently, word embeddings and unsupervised learning techniques have been combined to analyze unstructured medical text to study the concept of diseases (Shah and Luo, 2017), medical product reviews (Karim et al., 2020), or to extract informative sentences for text summarization (Moradi and Samwald, 2019).

In this work, we combine biomedical word embeddings and unsupervised learning to discover topics from real-world medical inquiries received by Bayer™. A real-world corpus of medical inquiries presents numerous challenges. From an inquirer (*e.g.* healthcare professional or patient) perspective, often the goal is to convey the information requested in as few words as possible to save time. This leads to an extensive use of acronyms and abbreviations, sentences with atypical syntactic structure, occasionally missing verb or subject, or inquiries comprising exclusively a single noun phrase. Moreover, since medical inquiries come from different sources, it is common to find additional (not relevant) information related to the text source; examples are references to internal computer systems, form frames (*i.e.* textual instructions) alongside with the actual form content, lot numbers, email headers and signatures, city names. The corpus contains a mixture of layman and medical language depending (mostly) on the inquirer being either a patient or a healthcare professional. Style and content of medical inquiries vary quite substantially according to which therapeutic areas (*e.g.* cardiovascular vs oncology) a given medicinal drug belongs to.

As already mentioned, medical inquiries are short. More specifically, they comprise less than fifteen words in most cases. Standard techniques for topic modelling based on LDA (Blei et al., 2003) do not apply, since the main assumption - each document/text is a distribution over topics - clearly does not hold given that the text is short (Qiang et al., 2019). Approaches based on pseudo-documents (Mehrotra et al., 2013) or using auxiliary information (Phan et al., 2008; Jin et al., 2011) are also not suitable since no meaningful pseudo-document nor auxiliary information are available for medical inquiries. Moreover, these models aim to learn semantics (*e.g.* meaning of words) directly from the corpus of interest. However, the recent success of pretrained embeddings (Peters et al., 2018; Devlin et al., 2019) shows that it is beneficial to include semantics learned on a general (and thus orders of magnitude larger) corpus, thus providing semantic information difficult to obtain from smaller corpora. This is particularly important for limited data and short text settings. To this end, there has been recently some work aimed at incorporating word embeddings into probabilistic models similar to LDA (Dirichlet multinomial mixture model (Yin and Wang, 2014)) and that - contrary to LDA - satisfies the single topic assumption (*i.e.* one document/text belong to only one topic) (Nguyen et al., 2015; Li et al., 2016). Even though these models include (some) semantic information in the topic model, it is not evident how to choose the required hyperparameters, for example determining an appropriate threshold when filtering semantically related word pairs (Li et al., 2016). Concurrently to our work, document-level embeddings and hierarchical clustering have been combined to obtain topic vectors from news articles and a question-answer corpus (Angelov, 2020).

Here, we propose an approach - schematically depicted in **Figure 1** - to discover topics from short, unstructured, real-world medical inquiries. Our methodology consists of the following steps: medical inquiries are preprocessed (*via* lemmatization, stopword removal) and converted to vectors *via* a biomedical word embedding (scispacy (Neumann et al., 2019)), a dimensionality reduction is then applied to lower the dimensionality of the embedded vectors (*via* UMAP (McInnes et al., 2018a; McInnes et al., 2018b)), clustering is performed in this lower dimensional space to group together similar inquiries (*via* HDBSCAN (Campello et al., 2013; Melo et al., 2016; McInnes et al., 2017)). These clusters of similar inquiries are then merged based on semantic similarity: we define these (merged) clusters as topics. Topics are then quantitatively evaluated *via* two novel quantities: topic semantic compactness and name saliency, introduced in this work. Finally, for visualization purposes, another dimensionality reduction is applied to visualize topics in a topic map. This methodology is used to discovery topics in medical inquiries received by Bayer™ Medical Information regarding the oncology drug regorafenib.

METHODS

Machine Learning Approach to Discover Topics in Medical Inquiries

Text Representation

One of the main challenges of topic discovery in short text is sparseness: it is not possible to extract semantic information from

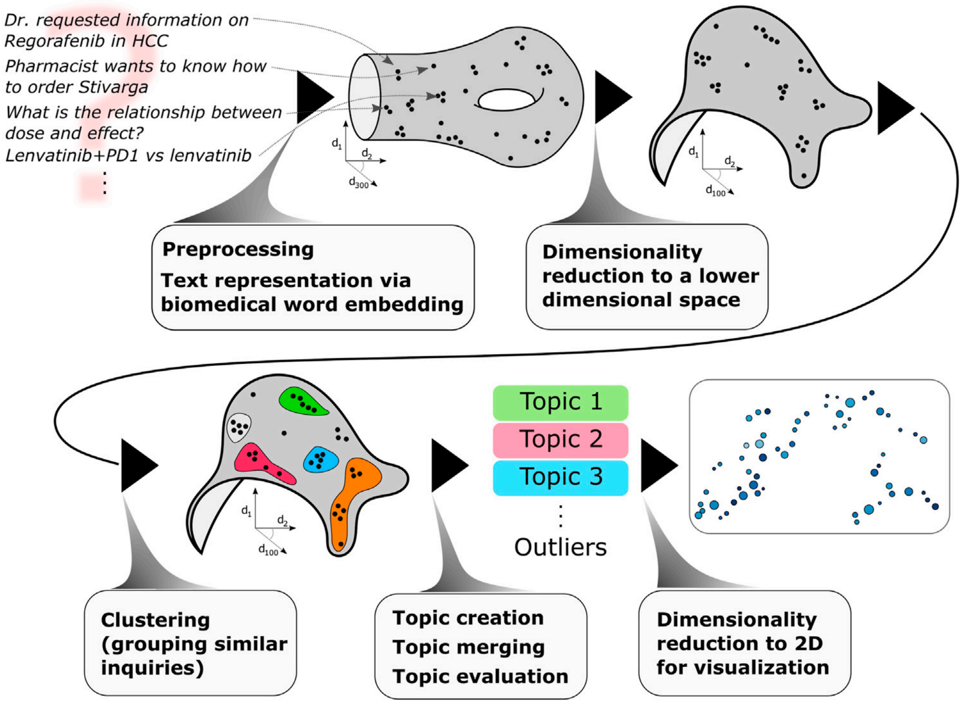


FIGURE 1 | From medical inquiries to medical topics *via* natural language processing and machine learning.

word co-occurrences because words rarely appear together since the text is short. In our case, the sparseness problem is exacerbated by two following aspects. First, the amount of data available is limited: most medicinal products receive less than 4,000 medical inquiries yearly. Second, medical inquiries are sent by patients as well as healthcare professionals (*e.g.* physicians, pharmacists, nurses): this leads to inquiries with widely different writing styles, containing a mixture of common and specialized medical text. The sparsity problem can be tackled by leveraging word embedding models trained on large corpora; these embeddings have been shown to learn semantic similarities directly from data, even for specialized biomedical text (Alsentzer et al., 2019; Beltagy et al., 2019; Lee et al., 2019; Neumann et al., 2019). Specifically, we use the scispaCy word embedding model (Neumann et al., 2019), which was trained on a large corpus containing scientific abstracts from medical literature (PubMed) as well as web pages (OntoNotes 5.0 corpus (Pradhan et al., 2013)). This assorted training corpus enables the model to treat specialized medical terminology and layman terms on the same footing, so that medical topics are discovered regardless of the writing style.

One of the main disadvantages of word vector (word2vec) models - like the (scispaCy) model used in this work - is their inability to handle out-of-vocabulary (oov) words: if a word appearing in the text is not included in the model vocabulary, it is effectively skipped from the analysis (*i.e.* a vector of all zeros is assigned to it). To tackle this issue, several models have been proposed, initially based on chargram level embeddings (FastText (Bojanowski et al., 2017)), and more recently contextual embeddings based on character (ELMO (Peters et al., 2018)), or byte pair encoding (Sennrich et al., 2016) representations (BERT (Devlin et al., 2019)). Even though other advancements - namely word polysemy handling and the use of attention (Vaswani et al., 2017) - were arguably the decisive factors, improvements in oov word handling also contributed in making ELMO and BERT the de facto gold standard for natural language processing, at least for supervised learning tasks.

Even though the use of contextual word embeddings is generally beneficial and can be readily incorporated in our approach (simply substituting the word representation), we notice that - given the large amount of noise present and the purely unsupervised setting - a word2vec model is actually advantageous for the task of extracting medical topics from real-word medical inquiries. Indeed, using a model with a limited yet comprehensive vocabulary (the scispaCy model used in this work includes 600 k word vectors) constitutes a principled, data-driven, efficient, and effective way to filter relevant information from the noise present in the corpus. This filtering is principled, and data driven because the words (and vectors) included in the model vocabulary are automatically determined in the scispaCy training procedure by optimizing the performance on biomedical text benchmarks (Neumann et al., 2019). This also leads to harmonization of the medical inquiry corpus by eliminating both non-relevant region-specific terms, and noise introduced by machine translation (words or expressions are sometimes not translated but simply copied still in the original language (Knowles et al., 2018)). Clearly, in

TABLE 1 | Illustrative comparison between standard and biomedical word embeddings.

Probe word	Most similar words (standard embedding)	Most similar words (biomedical embedding)
leukemia	cancer (0.68), cancers (0.65), tumor (0.65) tumors (0.64), chemotherapy (0.63), marrow (0.63) prognosis (0.61), malignant (0.61), anemia (0.60) diagnosed (0.60), pancreatic (0.59), ovarian (0.59)	leukaemia (0.97), leukemias (0.88), lymphoblastic (0.80) myelomonocytic (0.80), myelogenous (0.80), myeloid (0.80) promyelocytic (0.73), leukaemic (0.73), leukemic (0.72) blastic (0.67), blasts (0.67), therapy-related (0.66)
blood	urine (0.63), bleeding (0.62), liver (0.61) bloodstream (0.59), glucose (0.59), kidney (0.58) heart (0.58), kidneys (0.57), cholesterol (0.57) stomach (0.56), saliva (0.56), disease (0.56)	hematocrit (0.60), haematocrit (0.59), whole-blood (0.58) Arterial (0.57), pressure (0.55), heparinized (0.54) oncotic (0.53), hemoglobin (0.53), haemoglobin (0.52) venous (0.52), peripheral (0.52), venipuncture (0.51)
carcinoma	tumors (0.78), tumor (0.76), malignant (0.75) cancers (0.74), ovarian (0.71), pancreatic (0.69) lesions (0.67), cancer (0.66), prognosis (0.66) lung (0.65), prostate (0.64), leukemia (0.60)	carcinomas (0.90), adenocarcinoma (0.88), adenocarcinomas (0.79) squamous (0.76), well-differentiated (0.70), metastasizing (0.68) urothelial (0.68), tumours (0.68), cancers (0.68) cancer (0.68), non-metastatic (0.67), tumors (0.66)

The most similar words to the probe words blood, carcinoma, *and* leukemia *are shown for a standard and a biomedical word embedding. Values in parenthesis indicate the similarity with the corresponding probe word (maximum similarity is 1). The biomedical embedding model returns more specific and more medically relevant terms. The standard and biomedical embedding models are spaCy* en core web *lg and scispaCy* en core sci lg, *respectively.*

this context it is of paramount importance to use specialized biomedical embeddings so that the word2vec model has a comprehensive knowledge of medical terms despite its relatively limited vocabulary.

Table 1 presents a qualitative comparison of a standard embedding (*en core web lg*, trained on the Common Crawl) and a specialized biomedical embedding (scispaCy *en core sci lg*, trained also on PubMed). Specifically, for a given probe word (*i.e. leukemia*, *carcinoma*, *blood*), the words most semantically similar to it - measured by the cosine similarity between word vectors - are retrieved, together with their similarity with the probe word (shown in parenthesis, 1.0 being the highest possible similarity). It is evident that the biomedical embedding returns much more relevant and medically specific terms. For instance, given the probe word *leukemia*, the standard embedding returns generic terms like *cancer*, *tumor*, *chemotherapy* which are broadly related to oncology, but not necessarily to leukemia. In contrast, the biomedical embedding returns more specialized (and medically relevant) terms like *lymphoblastic*, *myelomonocytic*, *myelogenous*, *myeloid*, *promyelocytic*: acute lymphoblastic, chronic myelomonocytic, chronic myelogenous, adult acute myeloid, and acute promyelocytic are all types of leukemia.

Clustering Similar Medical Inquiries *via* Hierarchical Clustering

We have shown in the previous section that word embeddings provide a natural way to include semantic information (*i.e.* meaning of individual words) in the modeling. Medical inquiries comprise multiple words, and therefore a semantic representation for each inquiry needs to be computed from the word-level embeddings. We accomplish this by simply averaging the embeddings of the words belonging to the inquiry, thus obtaining one vector for each inquiry. Since these vectors capture semantic information, medical inquiries bearing similar meaning are mapped to nearby vectors in the high-dimensional embedding space. To group similar inquiries, clustering is performed in this embedding space, and for each medicinal product separately.

Before clustering is performed, a non-linear dimensionality reduction is applied to lower the dimensionality of the text representation, similar to Ref. 29. We utilize the UMAP algorithm (McInnes et al., 2018a; McInnes et al., 2018b) because of its firm mathematical foundations from manifold learning and fuzzy topology, ability to meaningfully project to any number of dimensions (not only two or three like t-SNE (van der Maaten and Hinton, 2008)), and computational efficiency. Reducing the dimensionality also considerably improves the clustering computational performance, greatly easing model deployment to production, especially for drugs with more than 5,000 inquiries.

Usually, it is not conducive to define an appropriate number of clusters *a priori*. A reasonable number of clusters depends on various interdependent factors: number of incoming inquiries, therapeutic area of the medicine, time frame of the analysis, and intrinsic amount of information (*i.e.* variety of the medical inquiries). For a given medicinal product, typically a handful of frequently asked questions covers a large volume of inquiries, accompanied by numerous low-volume and less cohesive inquiry clusters. These low-volume clusters often contain valuable information, which might not even be known to medical experts: their low volume makes it difficult to detect them *via* manual inspection. To perform clustering in the embedding space, we use the hierarchical, density-based clustering algorithm HDBSCAN (Campello et al., 2013; Melo et al., 2016; McInnes et al., 2017). As customary in unsupervised learning tasks, one needs to provide some information on the desired granularity, *i.e.* how fine or coarse the clustering should be. In HDSBCAN, this is accomplished by specifying a single, intuitive hyper-parameter (*min cluster size*). In our case, the objective is to obtain approximately 100 clusters so that the results can be easily analyzed by medical experts. Thus, the main factor in defining *min cluster size* is the number of inquiries for a given medicinal drug: the larger the medical inquiry volume, the larger the parameter *min cluster size*. Note that *min cluster size* is not a strict controller of cluster size (and thus how many clusters should be formed), but rather a guidance provided to the algorithm regarding the desired clustering granularity. It is also possible to combine different *min cluster size* for the same dataset, *i.e.* using a finer granularity for more recent inquiries, thus enabling the discovery of new topics when only few inquiries are received, at a price however of an increase in noise given the low data volume. Moreover, *min cluster size* is very slowly varying with data (medical inquiry) volume, which facilitate its determination (see **Supplementary Material**). At the end of this step, for each drug a set of clusters is returned, each containing a collection of medical inquiries. A given medical inquiry is associated to one topic only, in accordance with the single topic assumption.

In order to convey the cluster content to users, a name (or headline) needs to be determined for each cluster. To this end, the top-five most recurring words for each cluster are concatenated, provided that they appear in at least 20% of the inquiries belonging to that cluster; this frequency threshold is set to avoid to include in the topic name words that appear very infrequently but are still in the top-five words. Thus, if a word does not fulfill the frequency requirement, it is not included in the topic name (resulting in topic names with less than five words). By such naming (topic creation), the clusters are represented by a set of words, which summarize their semantic content.

Topic Merging and Topic Map Calculation

From this list of candidate topics, the vector representation for each word in the topic name is calculated; the topic name vector is then obtained by averaging the word vectors of the words present in the topic name. Topics are merged if their similarity evaluated as cosine similarity between their topic name vectors - is larger than a threshold. Threshold values range between 0.8 and 0.95 depending on the drug considered. This is done to limit the number of topics to be presented to medical experts. We favor this simple method over applying again HDSBCAN because the clustering would have to operate on very few datapoints (~100 topics). We also notice that HDBSCAN tends to group topics quite aggressively (even with *min cluster size* = 2), which would result in potentially loosing important information.

After the topics are merged, new topic names are generated according to the procedure outlined above. The final result is a list of topics defined by a given name, each containing a set of similar medical inquiries. The list of discovered topics is then outputted and presented to medical experts.

Since the goal is to extract as much knowledge as possible from incoming medical inquiries, a relatively large number of topics (typically around 100) is returned to medical experts for each medicinal product. To facilitate topic exploration and analysis, topics are visualized on a map that reflects the similarity between topics (**Figure 2A**): topics close to each other in this map are semantically similar. To obtain this semantic map, first topic vectors are computed by averaging the text representation of all inquiries belonging to a given topic; then, a dimensionality reduction to two dimensions *via* UMAP is performed.

Topic Evaluation: Topic Semantic Compactness and Name Saliency

Once topics are discovered, it is desirable to provide medical experts with information regarding the quality of a given topic.

The most popular topic evaluation metrics for topic modelling on long text are UCI (Newman et al., 2010) and UMass (Mimno et al., 2011). However, both UCI and UMass metrics are not good indicators for quality of topics in short text topic modelling due to the sparseness problem (Quan et al., 2015). In Ref. 44, a purity measure is introduced to evaluate short text topic modelling; however, it requires pairs of short and long documents (*e.g.* abstract and corresponding full text article), and thus it is not applicable here because there is no long document associated to a given medical inquiry. Indeed, evaluation of short text topic modelling is an open research problem (Qiang et al., 2019). An additional challenge is the absence of labels. Performing annotations would require substantial manual effort by specialized medical professionals and would be of limited use because one of the main goals is to discover previously unknown topics as new inquiries are received. The absence of labels precludes the use of the metrics based on purity and normalized mutual information proposed in Ref. Rosenberg and Hirschberg (2007), Huang et al. (2013), Yin and Wang (2014), Aletras et al. (2013). bring forward the valuable idea of using distributional semantic to evaluate topic coherence, exploiting the semantic similarity learned by word2vec models. Topic coherence is assessed by calculating the similarity among the top n-words of a given topic: semantically similar top n-words lead to higher topic coherence. If this might be in general desirable, in the case of discovering medical topics it is actually detrimental: interesting (and potentially previously unknown) topics are often characterized by top n-words which are not semantically similar. For example, a medical topic having as top 2-words *rivaroxaban* (an anticoagulant medication) and *gluten* is clearly relevant from a medical topic discovery standpoint. However, *rivaroxaban* and *gluten* are not semantically similar, and thus the metric proposed in Ref. 47 would consider this as a low coherence (and thus low quality) topic, in stark contrast with human expert judgment. Analogous considerations apply to the indirect confirmation measures in Roeder et al. (2015): words emerging in novel topics would have rarely appeared before in a shared context. For this reason, we introduce a new measure of topic compactness which takes into account the semantics of the inquiries, and does not require any labeled data. Specifically, we compute the similarity of all inquiries belonging to a given topic with each other (excluding self-similarity), sum the elements of the resulting similarity matrix, and divide by the total number of elements in this matrix. The topic semantic compactness γ^α of topic α reads

$$\gamma^\alpha = \sum_{i=1}^{|C^\alpha|} \sum_{\substack{j=1 \\ i \neq j}}^{|C^\alpha|} \frac{S(q_i, q_j)}{|C^\alpha|(|C^\alpha| - 1)} \tag{1}$$

where $|C^\alpha|$ is the cardinality of topic α (how many inquiries are in topic α), q_i (and q_j) is the word vector representing inquiry i

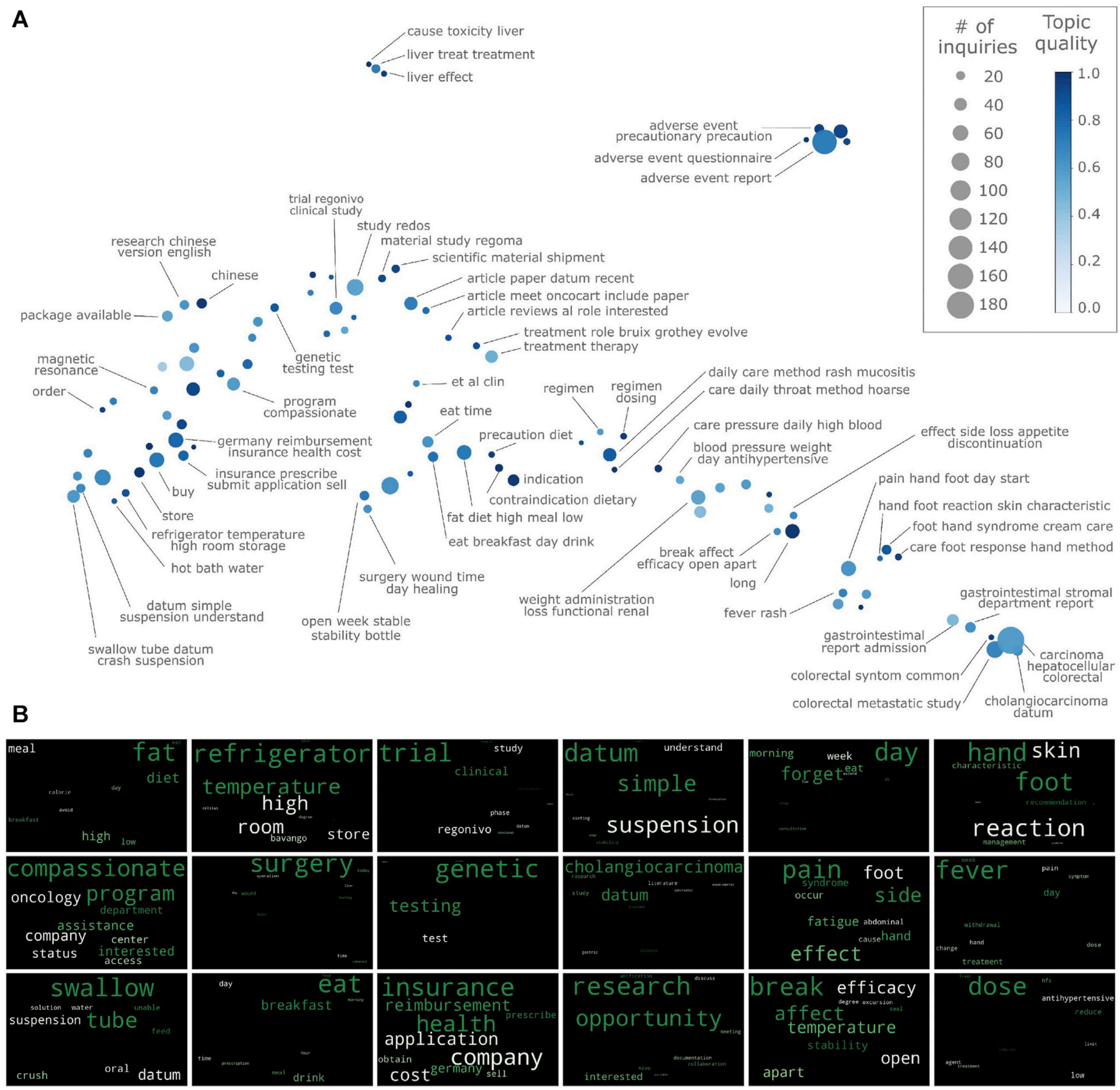


FIGURE 2 | Topic discovery for medical inquiries on regorafenib. **(A)** Topic semantic map. Axis are the non-linear dimensionality components and thus are not shown. **(B)** Wordclouds for selected topics.

(j), and S is a function quantifying the semantic similarity between inquiry q_i and q_j, taking values between 0 and 1 (S = 1 when q_i and q_j are identical, and S = 0 being the lowest possible similarity). Given the chosen normalization factor (*i.e.* the denominator in **Eq. 1**), $0 \leq \gamma^{\alpha} \leq 1$ and thus γ^{α} can be directly used as (a proxy for) topic quality score. The topic compactness maximum ($\gamma^{\alpha} = 1$) is attained if and only if every sentence (after preprocessing) contains exactly the same words. It is important to point out that γ^{α} automatically takes semantics into account: different but semantically similar medical inquiries would still have high similarity score, and thus would lead (as desired) to a high topic semantic compactness, despite these inquiries using different words to express similar content. Contrary to Ref. 47, the topic semantic compactness γ^{α} introduced in **Eq. 1** does not artificially penalize novel topics just because they associate semantically different words appearing in the same inquiry. To come back to the previous example, if numerous inquiries in a discovered topic contain the words *rivaroxaban* and *gluten*, the topic semantic compactness would be high (as desired), regardless from the fact that the top 2-words are not semantically similar since the similarity is evaluated at the inquiry level (by S (q_i,q_j) in **Eq. 1**).

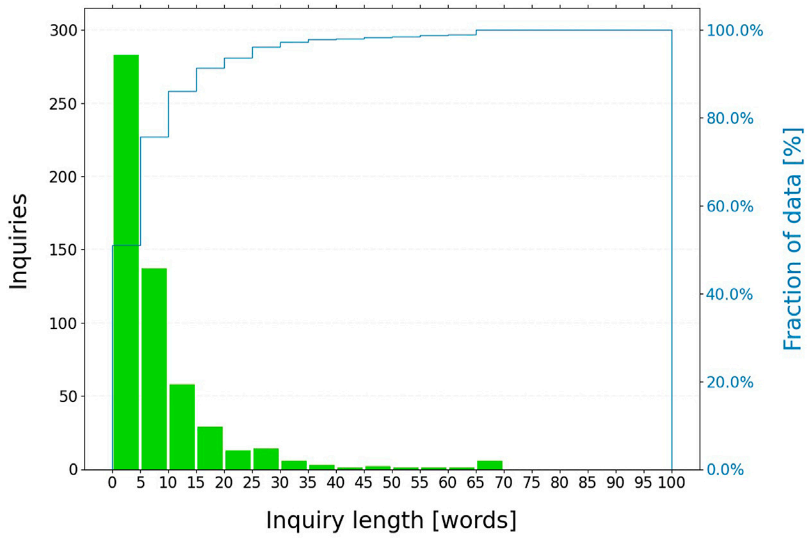


FIGURE 3 | Length distribution of medical inquiries on regorafenib. The dataset contains 3,880 medical inquiries; most of the inquiries (~90%) contains less than fifteen words after preprocessing.

The topic name is one of the main information shown to the users to summarize the semantic content of a discovered medical topic. It is therefore of interest to quantify how representative the name is for a given medical topic. This is tackled by answering the following question: how similar is the name with the inquiries grouped in the topic it represents? To this end, we calculate the name saliency τ^{α} for medical topic α by calculating the similarity of the word vector representing the topic name with the word vectors representing the inquiries in the topic, sum these similarity values, and divide by the total number of inquiries in the topic. This reads

$$\tau^{\alpha} = \sum_{i=1}^{|C^{\alpha}|} \frac{S(t^{\alpha}, q_i)}{|C^{\alpha}|} \tag{2}$$

where $|C^{\alpha}|$ is the cardinality of topic α (how many inquiries are in topic α), t^{α} is the word vector representing the name of topic α, and q_i is the vector representing inquiry i. This returns a score ($0 \leq \tau^{\alpha} \leq 1$) which quantifies how representative (salient) the name is for the topic it represents. As in the case of the topic semantic compactness, the name saliency τ^{α} takes natively semantics (*e.g.* synonyms) into account *via* $S(t^{\alpha}, q_i)$ in **Eq. 2**. In both **Eqs. 1**,**2**, the cosine similarity is used as similarity measure.

RESULTS

A Real-World Example of Topic Discovery: The Oncology Drug Regorafenib

As a real-world example of topic discovery, we present the results for medical inquiries on the oncology drug regorafenib (Bekaii-Saab et al., 2019). Regorafenib is an oral multikinase inhibitor which inhibits various signal pathways responsible for tumor growth.

In this work, all unsolicited medical inquiries received by Bayer™ worldwide in the time frame July 2019-June 2020 are considered (3,880 medical inquiries, see **Figure 3**). All non-English inquiries are translated to English using machine translation. These inquiries are then pre-processed: acronyms and abbreviations are resolved; non-informative phrases, words or patterns are removed; text is tokenized and lemmatized. Additional details are provided in **Supplementary Material**. Then, the topic discovery algorithm introduced above is applied with *min cluster size* = 6 and the UMAP dimensionality reduction to 100 components.

The semantic map with the discovered topics is shown in **Figure 2A**. These topics span a relatively large variety of themes, ranging from interactions with food and adverse drug reactions to purchase costs and literature requests. The topics are judged as meaningful and medically relevant by medical information specialists, based on their expert knowledge of the medicinal product.

Topics are also specific: the unsupervised learning approach allows information to emerge directly from the data, without recurring to predefined lists of keywords or classes, as required when using ontologies or supervised learning. An example of a very specialized topic for inquiries on scientific literature is *treatment role bruix grothey evolve*: 12 requests related to the review article on the treatment of advanced cancer with regorafenib published on February 2020 (Grothey et al., 2020). Other examples are the five topics *fat diet high meal low*, *eat breakfast day drink*, *precaution diet*, *eat time*, *contraindication diet*. Even though all these topics relate to nutrition, they are addressing different aspects. It is quite advantageous that they are identified as distinct since medical recommendations will likely differ across these five topics.

Thanks to the inclusion of semantics *via* word embedding, the algorithm is able to group together inquiries having similar meaning, even though the actual words in them are distinct. For instance, the topic *pain side effect foot fatigue* comprises 21 inquiries on medical issues (which may or may not be related with the medicine), in which the following words appear: *pain* (seven times), *side effect* (six times) *nausea* (three times), *fatigue* (five times), *dysphonia* (two times). The algorithm is able to cluster these inquiries together because similar inquiries are mapped close to each other in the high dimensional semantic space where clustering is performed. This is corroborated by the relatively high similarity score between the terms appearing in these inquiries (pain-nausea: 0.66, nausea-fatigue:0.61, pain-fatigue:0.71, dysphonia-pain:0.55, dysphonia-fatigue:0.49), scores much higher than zero, zero being the score expected for unrelated terms(*cf.* pain-day:0.05, nausea-sun:0.08). Conversely, if there is a moderate number of inquiries on a specific medical matter, the algorithm is generally able to detect that signal, as in the case of mucositis and hoarse in the two topics *daily care method rash mucositis*, and *daily care method rash hoarse*.

As shown in **Figure 2A**, the automatically generated topic names provide a reasonably good insight into the semantic content of their respective topics. However, one needs to be mindful that the topic might - and usually will - contain additional information of relevance. To convey this information in a simple yet effective way to the users, wordclouds are generated for each topic; examples of wordcloud are shown in **Figure 2B**. For example, in the wordcloud of topic *compassionate program* (**Figure 2B**, 1st column-2nd row), concepts not included in the topic name (*e.g. assistance*, *interested*, *access*, *status*) appear, thus giving further insight into the topic content. In some cases, even the wordcloud might not convey the topic meaning: users will then resort to manually inspect the inquiries belonging to the topic. For instance, the content of topic *chinese* is not clear, neither from the topic name nor from the wordcloud; however, inspection of the actual inquiries quickly reveals that they refer to the interaction between Chinese medicine and regorafenib (the word *medicine* does not appear since it is a stopword). Other examples are *al et clin* and *long*, which group together requests for scientific articles and product durability, respectively. Topic quality provides a useful guidance when exploring topics. If topic quality is close to one, medical inquiries in that topic are all very similar, and the topic name is expected to summarize the topic content well. Conversely, topics with low quality will contain inquiries that might differ quite substantially yet are similar enough to be clustered together by the algorithm. In these cases, manual inspection of the underlying medical inquiries may be a good strategy. From **Figure 2A**, it appears that smaller topics tend to have higher topic scores, although no clear trend emerges.

Finally, in addition of having similar inquiries within topics, the model captures semantic similarities between topics. This is apparent from **Figure 2A**: similar topics tend to be close to each other in the semantic map. Even though this feature does not influence the topic discovered, from a user perspective it provides a clear advantage when exploring topics (*e.g.* compared to reading them from as a simple list).

DISCUSSION

This study introduces an unsupervised machine learning approach to automatically discover topics from medical inquiries. After the initial (one-time) effort for preprocessing (*e.g.* abbreviation definition, stopword refinement) and hyper-parameters determination, the algorithm runs without requiring any human intervention, discovering key topics as medical inquiries are received. Topics can be discovered even if only a small number of inquiries is present, and are generally specific, thus enabling targeted, informed decisions by medical experts. Being completely unsupervised, the algorithm can discover topics that were neither known nor expected in advance, topics which often are the most valuable. This is in stark contrast with ontology or supervised based approaches, where topics need to be defined *a priori* (as collections of keywords or classes), and incoming text can be associated only to these predefined lists of topics, thus hindering the discovery of *a priori* unknown topics. The machine learning approach introduced here does not use ontologies (which are costly and hard to build, validate, maintain, and difficult to apply when layman and specialized medical terms are combined), and instead it incorporates domain knowledge *via* specialized biomedical word embeddings. This allows to readily apply the topic discovery algorithm to different drugs, without the burden of having to develop specialized ontologies for each product or therapeutic area. Indeed, the algorithm is periodically analyzing hundreds of thousands of medical inquiries for sixteen Bayer™ medicinal products, encompassing cardiology, oncology, gynecology, hematology, and ophthalmology.

Our approach has several limitations. First, it can happen that a small fraction of inquiries associated to a given topic are actually extraneous to it, especially for semantically broad topics. This is because - due to the noise present in this real-world dataset - the soft clustering HDBSCAN algorithm must be applied with a low probability threshold for cluster assignment to avoid the majority of inquiries being considered as outliers (see **Supplementary Material**). Second, even though the topic names are generally quite informative, a medical expert needs to read the actual inquiries to fully grasp the topic meaning, especially if a decision will be made on the grounds of the discovered topics. This is however not burdensome because inspection is limited to the inquiries associated to a given topic (and not all inquiries). Last, some discovered topics are judged by medical experts based on their expert knowledge - so similar that they could have been merged in a single topic, but are considered distinct by the algorithm. In these cases, manual topic grouping might be required to determine the top topics by inquiry volumes. Still, these similar topics very often appear close to each other in the topic map.

Despite these limitations, this study demonstrates that medical inquiries contain useful information, and that machine learning can extract this information in an automatic way, discovering topics that are judged by medical information specialists as meaningful and valuable. The hope is that this will stimulate mining of medical inquiries, and more generally the use of natural language processing and unsupervised learning in the medical industry. Interesting future directions are the inclusion of *a priori* expert knowledge (*e.g.* a list of expected topics) while at the same time maintaining the ability to discover new and previously unknown topics, and grouping topics in meta-topics though a clustering hierarchy.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data is proprietary to Bayer AG. Requests to access the datasets should be directed to angelo.ziletti@bayer.com.

AUTHOR CONTRIBUTIONS

AZ. led and thereby ideated and implemented the topic discovery algorithm, and is the main author the manuscript. MS, CB, DR. provided valuable suggestions on the topic discovery algorithm. CB, OT, and TW. designed and implemented the software architecture and data engineering pipeline for the algorithm deployment. TW, JV, JL, SK, XM, AM, DR, and MS. provided the in-house resources for the study, supervised the overall project, and provided domain knowledge expertise. All authors revised and commented on the manuscript.

FUNDING

Funding for the study was provided by Bayer AG.

ACKNOWLEDGMENTS

AZ. thanks Robin Williams and Nikki Hayward from Bayer™ Medical Information for providing expert insightful and in-depth feedback on the results of topic discovery.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2021.672867/full#supplementary-material

REFERENCES

Abdellaoui, R., Foulquié, P., Texier, N., Faviez, C., Burgun, A., and Schück, S. (2018). Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach. *J. Med. Internet Res.* 20, e85. doi:10.2196/jmir.9222

Aletras, N., and Stevenson, M. (2013). Evaluating Topic Coherence Using Distributional Semantics.," in *IWCS*. Editors K. Erk and A. Koller. (The Association for Computer Linguistics), 13–22.

Allahyari, M., Pouriyeh, S. A., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., et al. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *arXiv* [Epub ahead of print]. CoRR.

Alsentzer, E., Murphy, J., Boag, W., Weng, W.-H., Jindi, D., Naumann, T., and McDermott, M. (2019). "Publicly Available Clinical BERT Embeddings," in Proceedings of the 2nd Clinical Natural Language Processing Workshop, Minneapolis, Minnesota, USA: Association for Computational Linguistics, 72–78.

Angelov, D. (2020). Top2vec: Distributed Representations of Topics. *arXiv* [Epub ahead of print].

Arnold, C., and Speier, W. (2012). "A Topic Model of Clinical Reports," in Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA: SIGIR '12Association for Computing Machinery), 1031–1032. doi:10.1145/2348283.2348454

Bekaii-Saab, T., Ou, F., Ahn, D., Boland, P., Ciombor, K., Heying, E., et al. (2019). Regorafenib Dose-Optimisation in Patients with Refractory Metastatic Colorectal Cancer (Redos): a Randomised, Multicentre, Open-Label, Phase 2 Study. *Lancet Oncol.* 20, 1070–1082. doi:10.1016/S1470-2045(19)30272-4

Beltagy, I., Lo, K., and Cohan, A. (2019). Scibert: A Pretrained Language Model for Scientific Text," in *EMNLP/IJCNLP (1)*. Editors K. Inui, J. Jiang, V. Ng, and X. Wan. (Hong Kong, China: Association for Computational Linguistics), 3613–3618.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Machine Learn. Res.* 3, 993–1022.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Tacl* 5, 135–146. doi:10.1162/tacl_a_00051

Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013). Density-based Clustering Based on Hierarchical Density Estimates. *Lect. Notes In Comput. Sci.* 7819, 160–172. doi:10.1007/978-3-642-37456-2_14

Chen, J. H., Goldstein, M. K., Asch, S. M., Mackey, L., and Altman, R. B. (2017). Predicting Inpatient Clinical Order Patterns with Probabilistic Topic Models vs Conventional Order Sets. *JAMIA* 24 (3), 472–480. doi:10.1093/jamia/ocw136

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). {Bert}: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North {A}merican Chapter of the Association for Computational Linguistics: Human Language Technologies (Minneapolis, Minnesota: Association for Computational Linguistics), 4171–4186.

Grothey, A., Blay, J.-Y., Pavlakis, N., Yoshino, T., and Bruix, J. (2020). Evolving Role of Regorafenib for the Treatment of Advanced Cancers. *Cancer Treat. Rev.* 86, 101993. doi:10.1016/j.ctrv.2020.101993

Huang, R., Yu, G., Wang, Z., Zhang, J., and Shi, L. (2013). Dirichlet Process Mixture Model for Document Clustering with Feature Partition. *IEEE Trans. Knowl. Data Eng.* 25 (8), 1748–1759. doi:10.1109/tkde.2012.27

Jin, O., Liu, N. N., Zhao, K., Yu, Y., and Yang, Q. (2011). "Transferring Topical Knowledge from Auxiliary Long Texts for Short Text Clustering," in Proceedings of the 20th ACM International Conference on Information and Knowledge Management, New York, NY, USA: CIKM '11Association for Computing Machinery, 775–784. doi:10.1145/2063576.2063689

Karim, M. R., Beyan, O., Zappa, A., Costa, I. G., Rebholz-Schuhmann, D., Cochez, M., et al. (2020). Deep Learning-Based Clustering Approaches for Bioinformatics. *Brief. Bioinform.* 22, 393–415. doi:10.1093/bib/bbz170

Knowles, R., and Koehn, P. (2018). "Context and Copying in Neural Machine Translation.," in *EMNLP*. Editors E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii. (Brussels, Belgium: Association for Computational Linguistics) 3034–3041. doi:10.18653/v1/d18-1339

Kormilitzin, A., Vaci, N., Liu, Q., and Nevado-Holgado, A. (2021). Med7: a Transferable Clinical Natural Language Processing Model for Electronic Health Records. *Artif. Intell. Med.* 118, 102086. doi:10.1016/j.artmed.2021.102086

Landi, I., Glicksberg, B. S., Lee, H.-C., Cherng, S., Landi, G., Danieletto, M., et al. (2020). Deep Representation Learning of Electronic Health Records to Unlock Patient Stratification at Scale. *NPJ Digit. Med.* 3, 96. doi:10.1038/s41746-020-0301-z

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2019). BioBERT: a Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics* 36, 1234–1240. doi:10.1093/bioinformatics/btz682

Li, H., Wang, H., Zhang, Z., Sun, A., and Ma, Z. (2016). "Topic Modeling for Short Texts with Auxiliary Word Embeddings," in Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA: SIGIR '16Association for Computing Machinery, 165–174. doi:10.1145/2911451.2911499

Luque, C., Luna, J. M., Luque, M., and Ventura, S. (2019). An Advanced Review on Text Mining in Medicine. *WIREs Data Mining Knowl. Discov.* 9 (3), e1302. doi:10.1002/widm.1302

Mascio, A, Kraljevic, Z, Bean, D, Dobson, R. J. B., Stewart, R., Bendayan, R., et al. (2020). "Comparative Analysis of Text Classification Approaches in Electronic Health Records," in *BioNLP*. Editors D. Demner-Fushman, K. B. Cohen, S. Ananiadou, and J. Tsujii, (Association for Computational Linguistics), 86–94. doi:10.18653/v1/2020.bionlp-1.9

McInnes, L., Healy, J., and Astels, S. (2017). Hdbscan: Hierarchical Density Based Clustering. *Joss* 2 (11), 205. doi:10.21105/joss.00205

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. cite arxiv: 1802.03426Comment: Reference implementation available at http://github.com/lmcinnes/umap (Accessed September 9, 2021).

McInnes, L., Healy, J., Saul, N., and Grossberger, L. (2018). Umap: Uniform Manifold Approximation and Projection. *Joss* 3 (29), 861. doi:10.21105/joss.00861

Mehrotra, R., Sanner, S., Buntine, W., and Xie, L. (2013). "Improving Lda Topic Models for Microblogs via Tweet Pooling and Automatic Labeling," in Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA: SIGIR '13Association for Computing Machinery, 889–892. doi:10.1145/2484028.2484166

Melo, D., Toledo, S., Mourão, F., Sachetto, R., Andrade, G., Ferreira, R., et al. (2016). Hierarchical Density-Based Clustering Based on GPU Accelerated Data Indexing Strategy, *Proced. Comput. Sci.* 80, 951–961. doi:10.1016/j.procs.2016.05.389

Mimno, D. M., Wallach, H. M., Talley, E. M., Leenders, M., and McCallum, A. (2011). "Optimizing Semantic Coherence in Topic Models," in *EMNLP* (Edinburgh, United Kingdom: ACL), 262–272.

Moradi, M., and Samwald, M. (2019). Clustering of Deep Contextualized Representations for Summarization of Biomedical Texts. *arXiv* [Epub ahead of print]. abs/1908.02286.

Neumann, M., King, Beltagy, L., and Ammar, W. (2019). "Scispacy: Fast and Robust Models for Biomedical Natural Language Processing," in *BioNLP@ACL*. Editors (D. Demner-Fushman, K. B. Cohen, S. Ananiadou, and J. Tsujii. (Florence, Italy: Association for Computational Linguistics), 319–327. doi:10.18653/v1/w19-5034

Newman, D., Lau, J. H., Grieser, K., and Baldwin, T. (2010). "Automatic Evaluation of Topic Coherence," " in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, Stroudsburg, PA, USA: Association for Computational Linguistics), 100–108.

Nguyen, D. Q., Billingsley, R., Du, L., and Johnson, M. (2015). Improving Topic Models with Latent Feature Word Representations. *Tacl* 3, 299–313. doi:10.1162/tacl_a_00140

Peters, Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). "Deep Contextualized Word Representations," in *NAACL-HLT*. Editors M. A. Walker, H. Ji, and A. Stent. (New Orleans, LA, United States: Association for Computational Linguistics), 2227–2237. doi:10.18653/v1/n18-1202

Phan, X.-H., Nguyen, L.-M., and Horiguchi, S. (2008). "Learning to Classify Short and Sparse Text and Web with Hidden Topics from Large-Scale Data Collections," in Proceedings of the 17th International Conference on World Wide Web, New York, NY, USA: WWW '08Association for Computing Machinery), 91–100.

Pivovarov, R., Perotte, A. J., Grave, E., Angiolillo, J., Wiggins, C. H., and Elhadad, N. (2015). Learning Probabilistic Phenotypes from Heterogeneous Ehr Data. *J. Biomed. Inform.* 58, 156–165. doi:10.1016/j.jbi.2015.10.001

Pradhan, S., Moschitti, A., Xue, N., Ng, H. T., Bjorkelund, A., Uryupina, O., Zhang, Y., and Zhong, Z. (2013). "Towards Robust Linguistic Analysis Using¨ OntoNotes," in Proceedings of the Seventeenth Conference on Computational Natural Language Learning, Sofia, Bulgaria (Association for Computational Linguistics), 143–152.

Qiang, J., Qian, Z., Li, Y., Yuan, Y., and Wu, X. (2019). Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. *arXiv* [Epub ahead of print]. abs/1904.07695.

Quan, X., Kit, C., Ge, Y., and Pan, S. J. (2015). Short and Sparse Text Topic Modeling via Self-Aggregation.," in *IJCAI*. Editors Q. Yang and M. J. Wooldridge, (Buenos Aires, Argentina: AAAI Press), 2270–2276.

Roeder, M., Both, A., and Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures.," in *WSDM*. Editors X. Cheng, H. Li, E. Gabrilovich, and J. Tang. (New York, NY, United States: ACM) 399–408.

Rosenberg, A., and Hirschberg, J. (2007). "V-measure: A Conditional Entropy-Based External Cluster Evaluation Measure," in Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (Prague, Czech Republic: EMNLP-CoNLL), 410–420.

Sennrich, R., Haddow, B., and Birch, A. (2016). "Neural Machine Translation of Rare Words with Subword Units," in *ACL (1)* (Berlin, Germany: The Association for Computer Linguistics). doi:10.18653/v1/p16-1162

Shah, S., and Luo, X. (2017). Exploring Diseases Based Biomedical Document Clustering and Visualization Using Self-Organizing Maps. 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 1–6. doi:10.1109/HealthCom.2017.8210791

Sun, W., Cai, Z., Liu, F., Fang, S., and Wang, G. (2017). "A Survey of Data Mining Technology on Electronic Medical Records," in 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 1–6.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data Using T-SNE. *J. Machine Learn. Res.* 9, 2579–2605.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention Is All You Need," in *Advances in Neural Information Processing Systems*. Editors I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (Long Beach, CA, United States: Curran Associates, Inc.), 5998–6008.

Weng, W.-H., and Szolovits, P. (2019). Representation Learning for Electronic Health Records. *arXiv* [Epub ahead of print]. abs/1909.09248.

Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., et al. (2019). Deep Learning in Clinical Natural Language Processing: a Methodical Review. *J. Am. Med. Inform. Assoc.* 27, 457–470. doi:10.1093/jamia/ocz200

Yin, J., and Wang, J. (2014). "A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering," in *KDD*. Editors S. A. Macskassy, C. Perlich, J. Leskovec, W. Wang, and R. Ghani. (New York, NY, United States: ACM), 233–242. doi:10.1145/2623330.2623715

Conflict of Interest: Financial support for the research was provided by Bayer AG. AZ, OT, TW, JL, SK, MS, JV, DR, and AM work for Bayer AG. CB works for Areto Consulting Gmbh. The authors report a patent application on Topic Modelling of Short Medical Inquiries submitted on April 21st, 2020 (application number EP20170513.4).

The authors declare that this study received funding from Bayer AG. The funder was involved in the study design, model development, data analysis, and the writing of this article.

frontiers
in Public Health

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fpubh.2021.646157

Prediction of Obstetric Patient Flow and Horizontal Allocation of Medical Resources Based on Time Series Analysis

Hua Li[1,2], *Dongmei Mu*[2,3]*, *Ping Wang*[2], *Yin Li*[2] *and Dongxuan Wang*[1]

[1] Department of Abdominal Ultrasound, First Affiliated Hospital of Jilin University, Changchun, China, [2] School of Public Health, Jilin University, Changchun, China, [3] Department of Clinical Laboratory, First Affiliated Hospital of Jilin University, Changchun, China

OPEN ACCESS

Edited by:
João Manuel R. S. Tavares,
University of Porto, Portugal

Reviewed by:
Xiangfu Zhao,
Yantai University, China
Fouzi Harrou,
King Abdullah University of Science and Technology, Saudi Arabia

****Correspondence:***
Dongmei Mu
moudm@jlu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *25 December 2020*
Accepted: *08 September 2021*
Published: *14 October 2021*

Citation:
Li H, Mu D, Wang P, Li Y and Wang D (2021) Prediction of Obstetric Patient Flow and Horizontal Allocation of Medical Resources Based on Time Series Analysis.
Front. Public Health 9:646157.
doi: 10.3389/fpubh.2021.646157

Objective: Given the ever-changing flow of obstetric patients in the hospital, how the government and hospital management plan and allocate medical resources has become an important problem that needs to be urgently solved. In this study a prediction method for calculating the monthly and daily flow of patients based on time series is proposed to provide decision support for government and hospital management.

Methods: The historical patient flow data from the Department of Obstetrics and Gynecology of the First Hospital of Jilin University, China, from January 1, 2018, to February 29, 2020, were used as the training set. Seven models such as XGBoost, SVM, RF, and NNAR were used to predict the daily patient flow in the next 14 days. The HoltWinters model is then used to predict the monthly flow of patients over the next year.

Results: The results of this analysis and prediction model showed that the obstetric inpatient flow was not a purely random process, and that patient flow was not only accompanied by the random patient flow but also showed a trend change and seasonal change rule. ACF,PACF,Ljung_box, and residual histogram were then used to verify the accuracy of the prediction model, and the results show that the Holtwiners model was optimal. R2, MAPE, and other indicators were used to measure the accuracy of the 14 day prediction model, and the results showed that HoltWinters and STL prediction models achieved high accuracy.

Conclusion: In this paper, the time series model was used to analyze the trend and seasonal changes of obstetric patient flow and predict the patient flow in the next 14 days and 12 months. On this basis, combined with the trend and seasonal changes of obstetric patient flow, a more reasonable and fair horizontal allocation scheme of medical resources is proposed, combined with the prediction of patient flow.

Keywords: forecast, patient follow, DRG, seasonally, trend, medical costs

INTRODUCTION

In recent years, with the implementation of China's two child policy, the number of pregnant women is increasing rapidly. In addition, with the improvement of people's living standards, patients' pursuit of high-quality medical resources is also rising (1). For hospitals that can provide high-quality medical services and higher medical skills, the number of patients they receive each year is increasing year by year (2). In view of the ever-changing flow of patients received by hospitals, how the government and hospital management plan and allocate medical resources has become an important problem that requires urgent attention (3).

Hospital patient flow, congestion, and long waiting time are considered to be increasingly serious worldwide problems in the field of health care (4). Improper treatment of the above problems by hospital management can directly affect the patient's medical experience, and even hinder people from being admitted to hospital due to the lack of hospital reception capacity, and at the same time, result in an overload of work for medical staff (5). By accurately predicting the time and scale of patients' visits, managers can make better use of their limited resources (6). Thus, improving the utilization rate of medical resources and improving the medical experience of patients.

In recent years, the prediction of hospital patient flow has attracted more and more attention from business and academic circles (7, 8). Patient flow prediction is based on historical patient arrival and flow data, using a time series analysis prediction model to predict the arrival time and flow of patients in the future (9). When hospitals perform poorly in emergency patient flow handling, they seriously affect the medical experience of emergency patients (10). Patients sometimes have to bypass the nearest emergency hospital and take other emergency hospitals, which indirectly leads to an increase in the incidence rate and mortality rate (11). There is a commitment to the study of patient flow prediction in the literature on emergency room services (12). Time series techniques such as seasonal autoregressive integral moving average and generalized autoregressive conditional heteroscedasticity were used to predict patient flow (13, 14), providing decision support for management decision-making level, and providing more emergency patients with more urgent medical experience. Jones et al. (15) compared the performance of the regression index, smoothing, seasonal ARIMA, and the artificial neural network in emergency flow prediction. Wargon et al. (16) used the SPSS software package to determine a regression model and used three-year training data from four different hospitals to evaluate its one-year prediction performance. Reis and Mandl (17, 18) used the SAS software package to fit the ARIMA model into the historical data of the emergency department of the hospital in the past 10 years, and the model was verified accurately according to the last 2 years of the data set. A study by Liu and Ma (19) outlined that rural medical services play an important role in protecting and promoting the health of rural populations, and studied the degree of satisfaction with rural medical services. The most satisfied areas outlined by outpatients and inpatients related to the attitude of medical services and the explanation of disease conditions. Waiting time and medical expenses were the areas with which outpatients and inpatients were dissatisfied. To achieve forward-looking bed management, Jilani et al. (20) and Wimsatt (21) put forward the fuzzy time series to predict the number of visits of emergency patients, and used historical data to predict the patient flow of emergency departments in four hospitals in the United Kingdom. This method is more accurate and does not need any seasonal and periodic adjustment.

In terms of medical resource allocation, DRG has experienced more than 30 years of development. There is a growing interest in using DRG payment to repay hospitalization expenses worldwide. Many scholars have studied the DRG method. The purpose of this method is to change the previous retrospective payment method based on medical income to the settlement method based on medical output and make reasonable compensation for the medical resources consumed by medical institutions in the process of diagnosis and treatment through the packaging mode, so as to reduce the unreasonable growth of health expenses.

Tan et al. (22) put forward a medical and health care compensation method combined with activity weight, which allocates funds to hospitals according to the number and combination of clinical activities. This method is completely based on the compensation classification method of Australian refining DRG. Each refined DRG is assigned a weighted activity unit and converted to cost value to determine the annual ongoing funding allocation for each hospital. Australian public hospitals operate according to the fine diagnosis related group (AR DRG) (23) and separate them into specific DRG groups according to medical diagnosis or surgical procedures, patient age, separation method, clinical complexity, and complications, to effectively ensure the efficiency and fairness of hospital resource allocation. Chien et al. (24) then studied the impact of the DRG payment scheme on low value preoperative examination services. Through differential difference analysis, it was concluded that chest X-ray, echocardiography and blood pressure test, and other low-value examinations have decreased. The study concluded that financial incentives alone may not be enough to reduce the provision of low-value medical care.

According to the complexity of treatment and the impact of treatment time on the cost of treatment, Endrich et al. (25) have proposed the DRG scheme for newborns. Most of the mixed grouping schemes only classify newborns according to their birth weight. This method is used to determine the threshold definition based on the analysis of the variation point of birth weight and gestational age, and carry out detailed DRG classification. Bertoli and Grembi (26) studied the impact of political factors on DRG. After the implementation of budget restrictions, when there are doctors among politicians, the average level of DRG decreases. These analysis results show that DRG standardized price systems are not immune to political pressure. Vuagnat et al. (27) studied the impact of the implementation of a case-based payment system on postoperative readmission. The results showed that the rate of readmission in the private and public sectors increased relatively steadily and did not seem to be affected by the case-based payment system. Panagiotopoulos et al. (28) evaluated the Greek version of the DRG compensation system, and compared the DRG price with the average actual cost of each group of research cases, and evaluated the difference between the average length of DRG stay and the actual length of stay. Then a new DRG price calculation model is proposed, which is based on the actual cost calculation process of each year, which is composed of relative weight factors and benchmark price.

Based on the above research on patient flow prediction and DRG resource allocation, this paper proposes a method of patient flow prediction and the horizontal allocation of medical resources combined with time series analysis, to provide decision support for government and hospital management. The main contributions are as follows: using the time series analysis model to analyze and predict the obstetric patient flow, we analyze the trend and seasonal changes of obstetric patient flow, and predict the patient flow in each month in the next year. The trend and seasonal changes of obstetric patient flow were then studied, and a more reasonable and fair horizontal allocation scheme of medical resources was proposed combined with the prediction of patient flow.

DATA AND METHODS

Data

The historical patient flow data from January 1, 2018, to February 29, 2020, in the Department of Gynecology and Obstetrics at the First Hospital of Jilin University, China, were used as the training set. We did not use data from before 2018 due to the lack of data. However, due to the outbreak of the new crown epidemic, the hospital has imposed a patient flow restriction on the obstetrics department, that is, the data after March 2020 are not of trend and seasonal variation, and the data in this period were not used. In order to realize the analysis and prediction of monthly patient flow, the data were classified and summarized to achieve the monthly patient flow, and then 26 data points were obtained.

Smoothing of Time Series

Time series data sets usually contain significant random or error data. In order to analyze the trend rule in the data, we hope to delete these random fluctuations by drawing a smooth curve. The simplest method is to draw smooth curves by SMA function. The four graphs in **Figure 1** represent the corresponding smooth curves of orders 1, 2, 3, and 5 respectively. It can be seen from the figure that these data have obvious seasonal and trend change rules.

Data Decomposition

In the time series data with seasonal factors, the data can be decomposed into trend factors, seasonal factors, and random factors. The trend component can capture the change of future cycles, the seasonal component can capture the seasonal change of a cycle, and the random / error component can capture the random change that cannot be explained by the trend or seasonal change.

The decomposition function is used to decompose the time series data, and the decomposed data graph is shown in **Figure 2**. It can be seen from the seasonal part of the figure that the flow of pregnant women patients has a peak in autumn and a trough in winter. In the trend part, the number of pregnant women outpatient services increased periodically with time. With time, the random fluctuations in the time series seem to be roughly unchanged, showing a normal distribution.

RESULTS

Monthly Flow Forecast

Prediction Model

We use the HoltWinter model in the forecast package to fit a simple exponential smoothing prediction model to predict the number of pregnant women outpatient services in the next year. The prediction results are shown in **Figure 3**. Among them, light gray and gray are the predicted values corresponding to 80 and 95% confidence intervals respectively.

Accuracy

Whether the prediction model can continue to be optimized is determined by whether there is a correlation between the prediction errors of continuous prediction, that is, if there is no correlation between the prediction errors, the prediction model cannot be further optimized. ACF,PACF,Ljung box were used most to analyze the correlation. Among these functions, the ACF in R is used to calculate the error of the correlation graph. To specify the maximum order you want to see, you can use the Lag.max Parameter. In this case, the correlation diagrams of the 1–20 order delay of the flow prediction error of the maternal patients are calculated, as shown in **Figures 4**, **6** below. **Figure 4** indicates that the autocorrelation system does not touch the confidence limit in order 1–20. It can be seen from **Figure 5** that the partial correlation coefficients do not touch the confidence limit. Detailed ACF cases for 80 and 95% are given in **Figure 6**.

To test whether the non-zero correlation is significant when the lag is 1–20 in order, we used the Ljung box test and passed the test in the R software Box.tex Function implementation. The maximum order can be achieved by setting the Box.tex lag parameter. The Ljung box test statistic is 17.4, and the p-value is 0.6. This is not enough to prove (almost no evidence) that the prediction error in the sample is non-zero autocorrelation in the lag order of 1–20.

In addition, to ensure that the prediction model is the best, it is also a good way to check whether the prediction error has a normal mean of 0 and constant variance. For each test point, the prediction error can only be calculated by subtracting the predicted value from the observed value. Therefore, the prediction errors of the existing time series can only be calculated. To test whether the variance of the prediction error in the whole sequence remains unchanged, that is, the normal distribution of service value-added, we draw the prediction error graph and the prediction error distribution histogram of the normal curve in a period of time. **Figure 7** shows that the random fluctuation above the time series tends to be a constant with the increase of time; that is, it is stable in the mean value and variance. As can be seen from **Figure 8**, the histogram shows that the distribution of prediction error is ∼0-centered and more or less normal distribution. Although it seems to be slightly to the right,

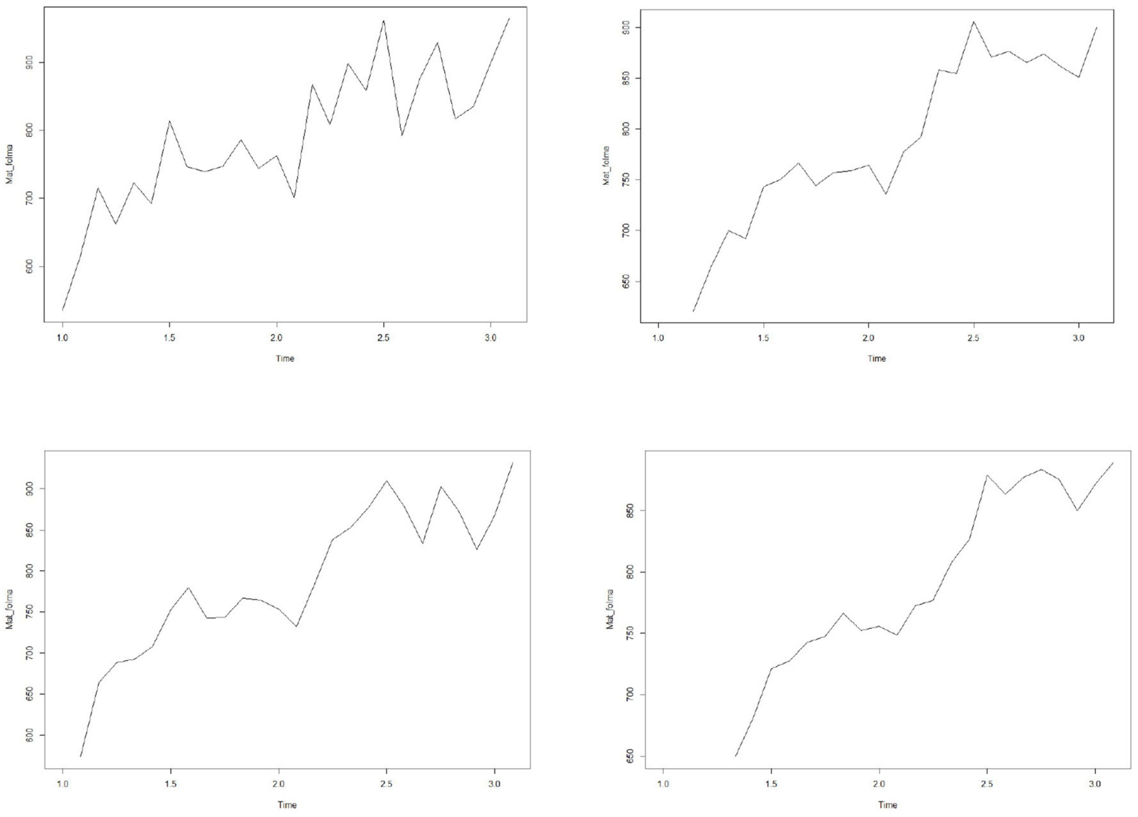


FIGURE 1 | Smoothed time series of patient flow.

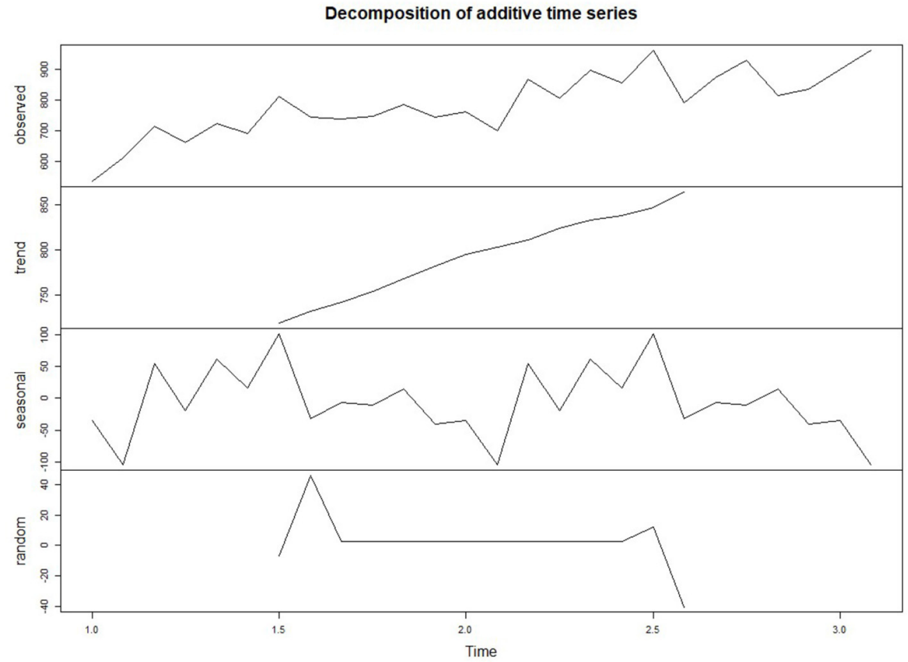


FIGURE 2 | Decomposed patient flow time series diagram.

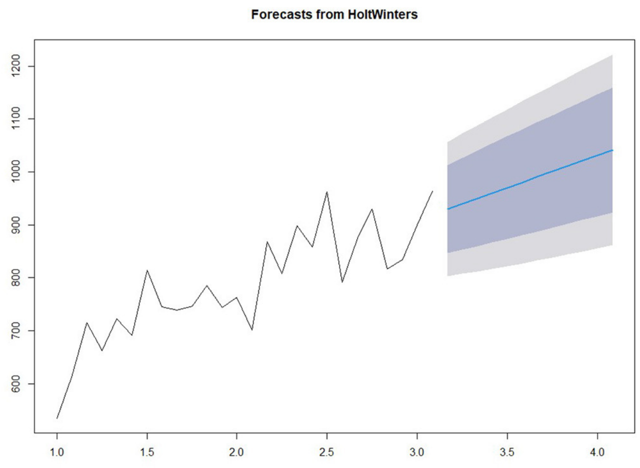


FIGURE 3 | Prediction of patient flow.

the tilt is relatively small, so the prediction error is approximately distributed with a mean of 0.

Because the continuous prediction error seems to have no autocorrelation and partial correlation, and the prediction error seems to be a normal distribution, the mean value is 0 and the variance is constant, meaning the prediction model can accurately predict the patient flow.

Daily Flow Forecast

Prediction Model

We used seven prediction models, such as XGBoost, SVM, RF, and NNAR, to predict the daily patient flow in the next 14 days. The basic principles of the seven models are introduced below.

The XGBoost (Extreme Gradient Boost) model is a special Gradient Boost decision tree that maximizes speed and efficiency (29, 30). XGBoost is essentially a tree-based approach combined with ensemble learning. The base tree is a classification regression tree. Similar to the local weighted linear regression algorithm, the tree-based regression algorithm is also a kind of local regression algorithm. By cutting the data set into pieces, each piece of data is modeled separately. But the difference is that a tree-based regression algorithm is a parameter-based learning algorithm. After training the model with the training data, once the parameters are determined, there is no need to change. A classification regression tree is a structure based on a decision tree, which can be used to solve both classification and regression problems. It is one of the first ten classical algorithms in the field of data mining.

XGBoost exports the enhancement tree by optimizing the objective function, which can be used to solve almost any objective function that can be written as a gradient. This includes things like ranking and Poisson regression, and random forests are difficult to implement. The XGBoost model is more sensitive to over-fitting if the data is noisy. Because trees are built in sequence, training often takes longer.

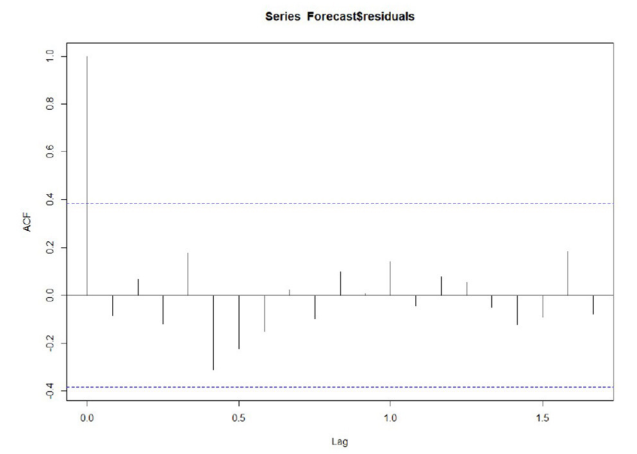


FIGURE 4 | Autocorrelation.

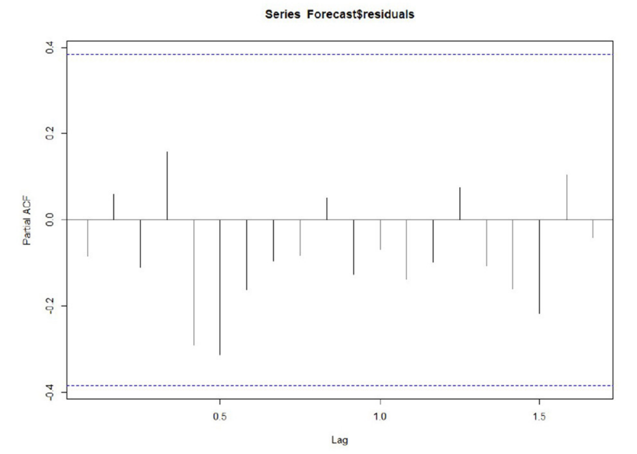


FIGURE 5 | Partial correlation graph.

The SVM model (31) transforms the source data into a higher dimensional space, where each input vector forms a sequence of seed points with a time delay. These vectors are then combined as key samples in a special way that allows the regression hyperplane to define the data distribution, and calculate with specified precision. These calculations represent the sum of all the samples called the "kernel," the integration function of the inputs. These functions can be linear or non-linear, and the guiding parameters affect the regression accuracy.

The advantages of SVM are as follows: it can deal with machine learning in the case of small samples, solve the problem of high-dimensional data, and avoid the problem of neural network structure selection and local minimum. This model has the following shortcomings: there is no universal solution for non-linear problems, so the Kernel function must be carefully selected. This model is also sensitive to missing data: when processing large-scale data, the memory resources of the computer are required to be high.

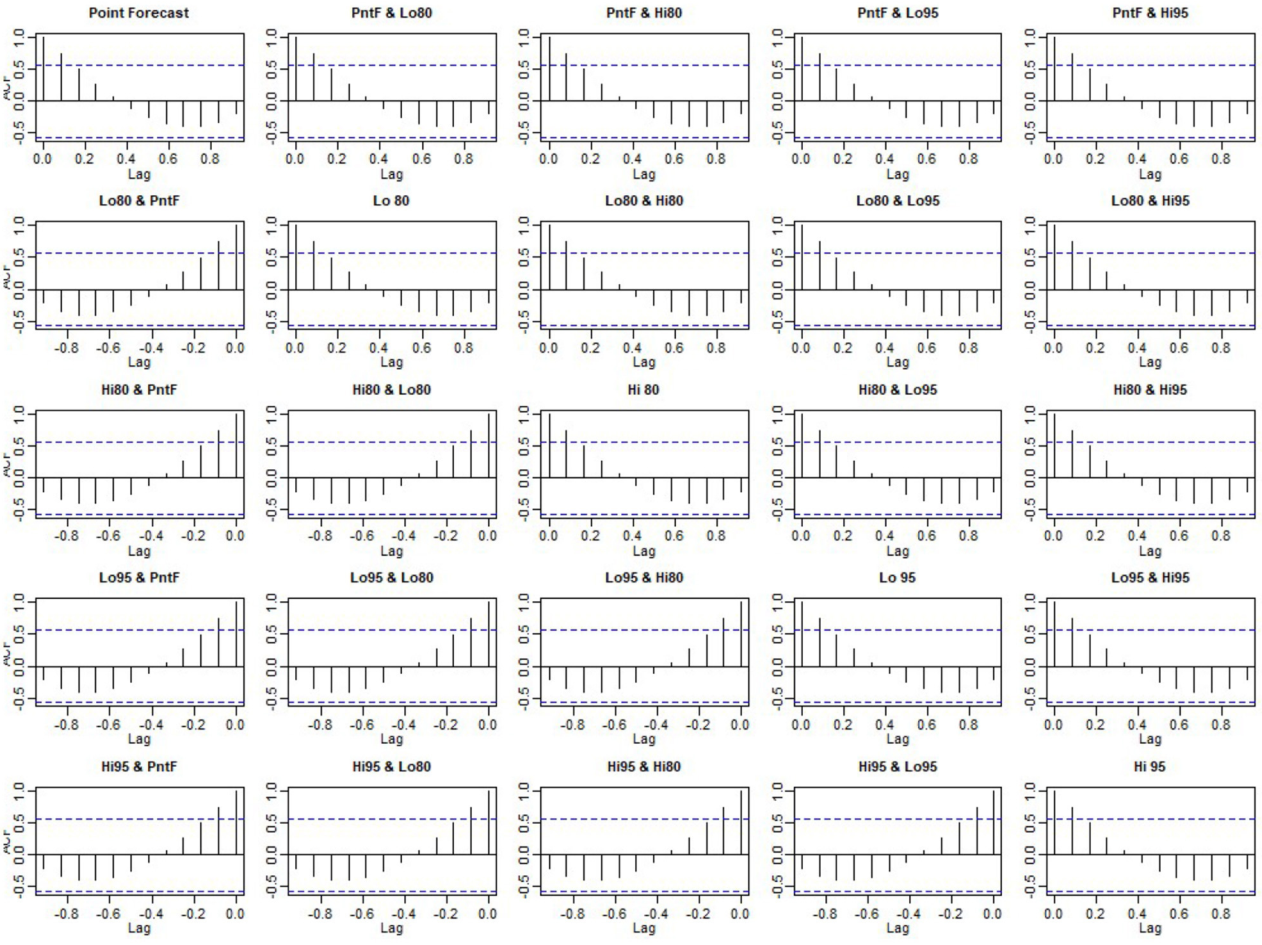


FIGURE 6 | Detailed autocorrelation diagram.

Random forest (32) is essentially a branch of machine learning called Ensemble Learning, which is a method of integrating many decision trees into a forest and using it to predict the final outcome. A Random Forest Model builds a forest in a random way. There are many decision trees in the forest, and there is no correlation between each decision tree in the random forest. Owning the forest, when a new input sample comes in, the decision tree in the forest makes its own judgment and determines to which category the sample should belong (for the classification algorithm). It then identifies which category is selected the most, and predicts which category the sample belongs to. A random forest can handle both quantities with attributes of discrete values and quantities with attributes of continuous values. In addition, the random forest can also be used for unsupervised learning clustering and outlier detection.

Model tuning is easier in a random forest than in XGBoost. In a random forest, we have two main parameters: the number of features to be selected for each node and the number of decision trees. Random forests are harder to configure than XGBoost. The main limitation of the random forest algorithm is that a large number of trees can slow down the algorithm for real-time prediction. For data containing categorical variables with different levels, the random forest favors those attributes with more levels.

NNAR is a model based on an artificial neural network and a prediction method based on a simple brain mathematical model (33). They allow complex non-linear relationships between response variables and their predictors. Neural networks can be considered as “neuron” networks organized by layers. The prediction variable (or input) constitutes the bottom layer, and the prediction (or output) constitutes the top layer. There may also be an intermediate layer containing “hidden neurons.” The simplest network does not contain a hidden layer, which is equivalent to linear regression. Once the middle layer with hidden neurons is added, the neural network will become non-linear.

NNAR can process temporal projections of large data sets. In addition, the prediction results are better for raw data with higher volatility. This model has many parameters, which start from random values during training, resulting in random factors in prediction. On the other hand, the network model needs to

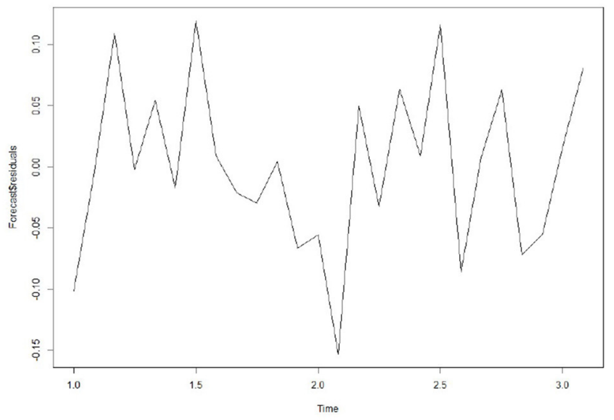


FIGURE 7 | Prediction residual diagram.

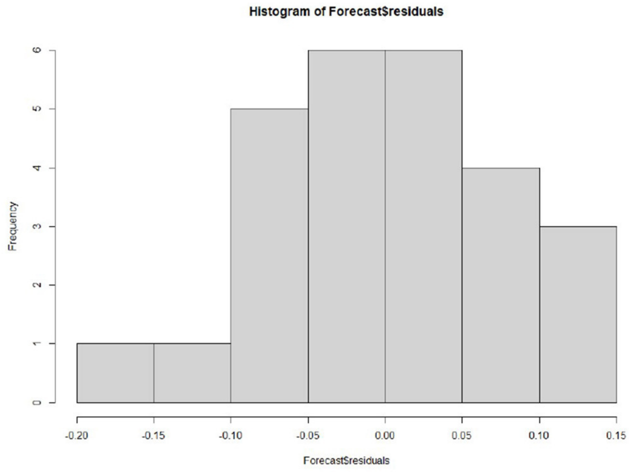


FIGURE 8 | Histogram of prediction residuals.

be trained many times with different random starting points, and the results are then averaged.

STL uses the Loess regression function to decompose the time series data into seasons, trends, and irregular components (34). What should be noted is that the prediction intervals ignore the uncertainty associated with the seasonal component. They are computed using the prediction intervals from the seasonally adjusted series, which are then reseasonalized using the last year of the seasonal component. The uncertainty in the seasonal component is ignored.

STL can handle any type of seasonal data, not just monthly and quarterly data. The smoothness of the trend cycle can also be controlled by the user. It is robust to outliers, that occasional outlier observations do not affect the estimation of trend periods and seasonal components. On the other hand, STL also has some disadvantages. In particular, it does not automatically handle changes in the trading day or calendar, it just provides additional decomposition.

If we combine autoregressive and moving average models, we will get a non-seasonal ARIMA model (35). In the autoregressive model, the linear combination of past values of variables is used to predict interesting variables in the future. The moving average model does not use the past value of the prediction variable in the regression but uses a past prediction error in the model similar to the regression.

The ARIMA model is very simple and requires only endogenous variables without recourse to other exogenous variables. The ARIMA model is recommended when the raw data does not fluctuate much. This model has the following disadvantages: it can only capture linear relations, but not non-linear relations. Temporal data is required to be stable, or stable through differential differentiation.

The Holt-winters method (36, 37) is a time series analysis and prediction method that is suitable for non-stationary series containing linear trends and periodic fluctuations. The exponential smoothing method is used to make the model parameters adapt to the changes of non-stationary series, and make a short-term prediction of the future trend. Holt winters method introduces winters cycle term based on the Holt model, which can be used to deal with the fluctuation behavior of a fixed cycle in time series such as monthly data, quarterly data, and weekly data.

Holt-winters is a method to predict time series by using cubic exponential smoothing. The cubic exponential smoothing algorithm can save the trend and seasonal information of time series data well. Through quadratic exponential smoothing, the overall trend information can be retained. The seasonal characteristics of time series are processed by cubic exponential smoothing. But this model is not suitable for long period series prediction.

In the above section of this paper, the basic principles of seven prediction models have been introduced. The following seven models are used to predict the daily patient flow over the next 14 days. In order to facilitate the display of predicted data, the training data was selected from 20 cycles of data. First, periodize the time series data, and the number of days selected for the period is 7. The prediction results of the five methods are shown in the **Figure 9**.

To accelerate the speed of gradient descent in machine learning and improve the prediction accuracy, the time series data are normalized. SVM and RF methods conduct training and prediction on the normalized data, and the results are shown in **Figure 10**. For details of the hyperparameter values of the above seven models, see **Appendix 2**.

Accuracy

In this section, ME, RMSE, MAE, and other indicators are used to describe the accuracy of the seven methods. The meanings of ME, RMSE, MAE, and other parameters are shown in **Appendix 1**. The values of each parameter in the training set are shown in **Table 1**. The XGBoost method has the least value on each index, that is, the model fits the training data best. Second, the RF method has a MAPE value of 11.79, which has a better performance than the other methods. The value of the ARIMA and SVR methods in each parameter are large, indicating that

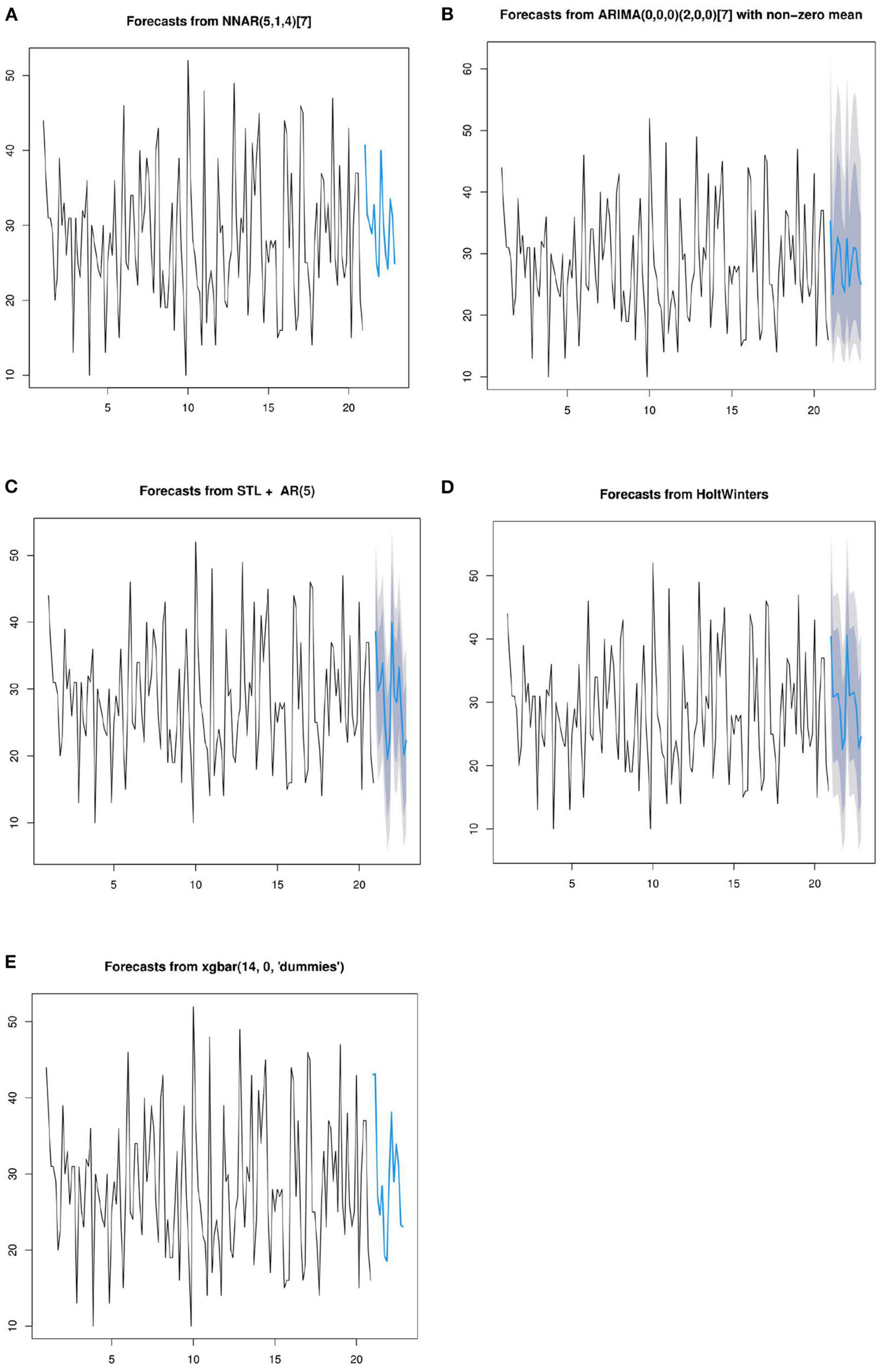


FIGURE 9 | Forecasts from NNAR**(A)**, ARIMA **(B)**, STL **(C)**, HoltWinters **(D)**, and XGBoost **(E)**.

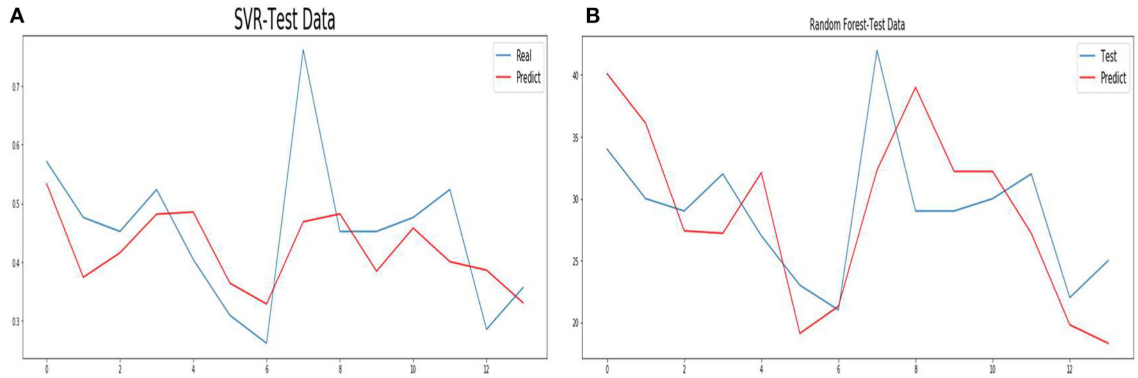


FIGURE 10 | Forecasts from SVR **(A)** and Random Forest **(B)**.

TABLE 1 | Accuracy index of training data.

	ME	RMSE	MAE	MPE	MAPE	ACF1	R2
NNAR	0	4.87	3.81	−5.57	15.53	0.02	0.54
STL	−0.13	6.68	5.35	−7.39	21.64	−0.01	0.44
ARIMA	−0.14	8.38	6.91	−10.58	27.63	0.07	0.14
HoltWinters	0.65	8.09	6.50	−5.76	25.24	0.06	0.36
XGBoost	0.02	0.26	0.18	−0.11	0.65	−0.10	0.99
SVR	0.58	11.6	9.28	1.67	38.40	–	0.30
RF	0.21	4.02	3.24	−0.60	11.79	–	0.75

TABLE 2 | Accuracy index of testing data.

	ME	RMSE	MAE	MPE	MAPE	R2
NNAR	−1.2	4.11	3.28	−5.56	11.92	0.63
STL	0.47	2.8	2.36	1.95	8.30	0.83
ARIMA	0.44	3.8	2.77	−0.02	9.53	0.48
HoltWintes	−1.12	2.4	1.93	−4.16	6.68	0.83
XGBoost	−0.52	6.52	4.90	−2.02	15.98	0.57
SVR	1.74	5.8	4.64	6.42	18.47	0.43
RF	0.40	4.54	3.04	2.64	10.09	0.35

these methods have the worst fitting on the training data. The maximum value of MAPE of the SVR method is 38.40; that is, the absolute difference of predicted data reaches 38.4% on average compared with all training data.

Table 2 shows the values of each parameter of the seven prediction models on the test set. From R2 and MAPE parameters, HoltWinters, and STL methods have higher prediction accuracy. The HoltWinters method has the best predictive performance with the highest R2 value of 0.833 of all the methods. The MPE value STL method is 1.95, while the MAPE value is 8.3, respectively. This indicates that most of the data predicted by this method are greater than the actual value. The MPE value of the HoltWinters method is −4.16, while the MAPE value is 6.68, respectively. This indicates that most of the data predicted by this method are smaller than the actual value. The MPE value of the ARIMA method is −0.02, while the MAPE value is 9.53, respectively. This indicates that half of the predicted data of the ARIMA method are greater than the actual value and half are less than the actual value.

Analysis and Discussion

The experimental results predicted by seven models are given above. In this section, combined with the advantages and disadvantages of the model and experimental data, the experimental results and the performance of the prediction model are further discussed. The HoltWinters and STL models achieve the best predictive performance. Their MAPE values are 6.68 and 8.3, and their R2 values are 0.833 and 0.832, respectively. The HoltWinters model has the best prediction performance, mainly because of its good processing of trends and non-stationary series of periodic fluctuations. This model is a method to predict the time series by using the cubic exponential smoothing method. Through quadratic exponential smoothing, the overall trend information can be retained. The seasonal characteristics of the time series are processed by the cubic exponential smoothing method. The prediction performance of the STL model is high mainly for the following reasons. This model can deal with any type of seasonal data, not just monthly and quarterly data, and has good performance for daily data. In addition, the model is very robust to outliers, and occasional outliers do not affect the prediction of trend cycles and seasonal components.

On the training set, XGboost and RF have the best fitting performance, with mAPE of 0.65 and 11.79 respectively. However, on the test set, the prediction performance is poor, and their R2 values are 0.57 and 0.35, respectively. The actual data of the hospital are used in this paper, and there is much noise in the data. XGBoost and RF models are sensitive to noise data during training, which leads to over-fitting of the model. SVR model has poor fitting performance in training data, and its MAPE value is 38.4. Its mAPE and R2 values are 18.47 and 0.43, respectively. This is mainly because SVR does not use the appropriate kernel function. Time series is a non-linear problem, so the method of choosing the kernel function has a great influence on model prediction performance. In addition, the model is sensitive to noise data. The MAPE and R2 values of the ARIMA model are 9.53 and 0.48, respectively. The actual data used in this paper fluctuated. When using the ARIMA model for prediction, the model cannot predict time series fluctuation data well, which leads to the model not achieving good performance.

DISCUSSION

This section of the paper first discusses the research status and shortcomings of DRG resource allocation methods at home and abroad, and then explores the progress and problems of medical insurance compensation in China. Then, according to the analysis and prediction method of patient flow, decision support for the government and hospital management is discussed. Finally, the limitations of the proposed method are analyzed.

DRG Status and Deficiencies

The DRG payment method shortens the length of hospital stay but brings increasing medical expenses and medical costs. Its impact on health care and health outcomes is controversial, especially in low- and middle-income countries. Policymakers should carefully consider each component of the DRG payment design according to policy objectives (38) so as to allocate medical resources efficiently and reasonably (39).

Clinical records ultimately determine the future funding of our healthcare system. Proper communication and education between medical staff and hospital coding staff are essential to ensure accurate documentation and accurate AR-DRG coding and to achieve the best and the most reasonable compensation in DRG mode. Lisbeth et al. (40) have studied whether the introduction of the DRG system affects the number of secondary diagnoses. During the study period, hospitals were divided into two groups: those with DRG payment systems and those without DRG systems. In all regional hospitals, the number of coded secondary diagnoses has increased, but the number of secondary diagnoses per case has also increased. After the implementation of the expected payment system, the secondary diagnoses of the hospitals have increased greatly. Aiming at the problem of misclassification frequently occurring in DRG, Suleiman et al. (41) have proposed the introduction of expert conjecture combined with the Bayesian model to improve the occurrence of DRG errors and significantly improve the efficiency of clinical coding review.

DRG payments may be slightly more efficient, but they face upgrades in medical record codes and damage the quality and fairness of medical care (42). Damage to health care rights and interests has been reported, especially for patients who are not supported by this payment plan (43). Zou et al. (44) evaluated the impact of DRG on health care and health outcomes in China and comprehensively described the expenditure, efficiency, quality, and fairness of health care. Montefiori et al. (45) studied the DRG program in relation to newborns. The results show that the efficiency of the DRG system is low. The cost of the same DRG patient varies greatly, and even though the cost of the patients with very low birth weight is about twice the reimbursement amount stipulated by the policy; when the newborn is full-term, the cost is 20,000 euros less than that reimbursed by DRG.

Kim et al. (46) have studied the performance of DRG based compensation for medical and health care expenditure. Payment based on DRG shortens the hospitalization time, and also changes the behavior of doctors collecting DRG code from outpatient, and increases medical expenses and medical costs continuously. To some extent, there is no overall budget for national medical insurance. How to plan the DRG global budget is an important problem that needs to be solved in further research.

Current Situation of China's Medical Insurance

Since the "new medical reform policy" was introduced in 2009, with the progress of medical reform year by year, the reform of medical security has achieved great results in many aspects, especially in terms of reimbursement content, reimbursement proportion, payment mode transformation, serious illness reimbursement systems, remote medical treatment systems, and the participation of commercial insurance. The new medical reform puts forward the short-term goal of effectively alleviating the difficulty and high cost of medical treatment, and the long-term goal of establishing and improving the basic medical and health system covering urban and rural residents, and providing safe, convenient, effective, and inexpensive medical and health services for the masses.

With the gradual development and improvement of medical security systems, the goal of universal medical insurance is possible, and the number of beneficiaries has been increasing in recent years. Because medical security is closely related to people's daily life, it occupies an extremely important part of the social security system. As the core resource of sustainable development of medical security, strengthening the supervision of basic medical insurance funds is of great significance to ensure the safe and sustainable operation of medical insurance funds.

To effectively curb the unreasonable growth of medical expenses and change the situation of "relying on drugs to support doctors," China has accelerated the pace of medical and health system reform. In January 2017, the National Development and Reform Commission, the Health and Family Planning Commission, and the Ministry of Human Resources and Social Security jointly issued a notice on promoting DRG payment, requiring all localities to further promote DRG payment reform, mainly to include gradually expanding the scope of DRG payment, reasonably determining specific diseases and charging standards, solidly completing the connection of DRG payment, and earnestly implementing these various reform policies.

The direction of China's medical reform is to implement DRG payment in a well-rounded way. From the hospital level, it is imperative to carry out DRG cost accounting. DRG cost accounting calculates the cost of diagnosis and treatment by superposition of medical project costs, drug costs, and separate charging material costs for the treatment of a single disease. Practice shows that DRG cost accounting can not only reduce the cost and optimize the diagnosis and treatment process, but also provide data support for government departments to formulate reasonable disease costs and promote the formation of cost-oriented operation mechanisms. Therefore, the introduction and strengthening of the DRG accounting method is key to furthering cost accounting work in hospitals under the situation of new medical reform, especially under the reform of the DRG payment mode.

Horizontal Allocation of Resources

As the medical center of Jilin Province in Northeast China, the first hospital of Jilin University ranks first in Jilin Province in terms of medical technology and abundant medical resources. In Jilin Province, medical institutions are divided into three categories: primary medical institutions, regional medical institutions, and tertiary medical institutions. Most medical institutions, especially regional and tertiary medical institutions, are public. The government has strict restrictions on the pricing of medical services. The loss of public hospitals is borne by the government financial department. All public medical institutions can enjoy the medical insurance policy.

At present, one of the main problems in the medical service system of Jilin province is that the total amount of medical resources is relatively insufficient and the utilization structure of medical resources is unreasonable. In this paper, we use the time series analysis model to analyze and predict the obstetric patient flow, analyze the trend and seasonal variation of obstetric patient flow, and predict the patient flow of each month in the next year. Based on the trend change, seasonal change, and patient flow of obstetric patients, a more reasonable and fair horizontal allocation plan of medical resources is given.

In this paper, the variation of patient flow trends was analyzed to provide decision support for government and hospital management. For diseases with an increasing tendency, the total amount of annual compensation can be increased according to the increase proportion. For diseases with a downward tendency, the total annual compensation will be reduced by floating. Then the total amount of compensation between different diseases can be adjusted horizontally. For hospital management, we can make a reasonable and efficient medical resource allocation plan according to the floating characteristics of patients every month.

The seasonal variation rule of patient flow obtained in this paper is helpful for hospital managers in providing more reasonable medical resource allocation plans for medical staff and patients. For patients, it provides patients with peak and low peak patient flow, thus providing a reference for patients in future hospitalization. For managers, it is conducive to resource allocation from the perspective of resource scheduling, to provide a better medical treatment experience for patients and guide medical staff to rest, study and work under the condition of limited resources. In the low peak period, some medical staff can take turns to take off work and study, thus providing medical staff with opportunities for rest and study, thus improving their life quality and medical quality. In the peak period, targeted medical staff should go to work in combination with the actual situation, which will improve the medical experiences of patients.

With the year-by-year increase of patient flow, the annual amount of medical insurance is less than the reimbursement amount of all patients in most cases. The existing medical insurance reimbursement system is based on the principle of first use first reimbursement, which often occurs at the end of each year that patients cannot reimburse or reimburse part of medical expenses. According to the monthly patient flow obtained by the prediction model, medical insurance compensation can be evenly covered for each patient in each month according to weight. Assuming that the predicted monthly patient flow is M_i ($i =$ 1,2,...,12), the average hospitalization cost of each patient is E, and the annual reimbursement amount of medical insurance is I. then the total patient flow in the next year is $M_1 + M_2 + ... + M_{12}$, and the medical insurance cost to be paid is $(M_1 + M_2 + ... + M_{12}) * E$. Based on the idea of horizontal resource allocation, according to the predicted patient flow, the amount of reimbursement for each person is $E * I / ((M_1 + M_2 + ... + M_{12}) * E)$, that is, $I / (M_1 + M_2 + ... + M_{12})$. The above horizontal resource allocation scheme based on flow prediction enables every patient to use medical insurance fairly, avoiding the problem that patients cannot use medical insurance fairly at the end of each year.

Limitations

The data used in this paper are only from a tertiary hospital in Jilin province, so the trend and seasonal variation of the analysis may not be suitable for the prediction of patient flow in all hospitals. In addition, in terms of patient flow prediction, the data used in this paper are <3 cycles, which also affects the accuracy of the prediction model in predicting future data. Finally, this paper only analyzes and predicts the flow model of patients in obstetrics, but it is also meaningful for other kinds of diseases. In future studies, it can be expanded to more hospitals and patients with more diseases. The analysis and prediction of patient flow in multiple hospitals and diseases will be more meaningful for the government and hospital management to provide decision support.

CONCLUSION

At present, patients' demand for medical treatment is not only expected to obtain timely and effective medical services at a reasonable cost but also expected to enjoy a higher level of medical services at a reasonable cost as they pay more attention to the service quality and service perception of the medical behavior itself. Seven models including XGBoost, SVM, RF, and NNAR were proposed to predict patient flow over the next 14 days and 12 months.

The analysis and prediction results obtained from the prediction model show that the obstetric inpatient flow is not a purely random process and that the patient flow is not only accompanied by the random patient flow but also presents the trend and seasonal changes. The trend change law is increasing year by year. The seasonal variation law is that the patient flow presents the trough in the first and fourth quarters of each year, and reaches the peak in the second and third quarters. The flow of patients shows a trough from Friday to Sunday and peaks from Monday to Thursday. The distribution of random flow is approximately normal.

The HoltWinters model is used to predict the monthly flow of patients over the next year. ACF, PACF, Ljung_box and residual histogram are used to verify the accuracy of the prediction model, and the results show that the Holtwiners model has reached the optimal level. Combined with the prediction of patient flow, a more reasonable and fair horizontal allocation of medical resources is proposed.

Seven models such as XGBoost, SVM, RF, and NNAR are used to predict the daily patient flow in the next 14 days. R2, MAPE, and other indicators were used to measure the accuracy of the 14-day prediction model. The results showed that Holtwiners and STL prediction models achieved high accuracy.

This method is helpful for optimizing the allocation of medical resources, reducing medical costs, and improving service efficiency and quality.

CODE AVAILABILITY

All code is open source with no restrictions and is available from https://github.com/hometownjlu/obstetric-patient-flow.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/hometownjlu/obstetric-patient-flow.

AUTHOR CONTRIBUTIONS

HL, DM, and DW designed and planned the study. HL, PW, and YL implemented the study and drafted the article. PW and YL analyzed and validated the data under the supervision of DM and DW. All authors participated in data analysis and result interpretation, carefully revised all contents of the manuscript, and critically reviewed and approved the submitted manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (71974074), the Science and Technology Development Plan of Jilin Province (20200301004RQ), and the Education Department of Jilin Province (20211103KJ).

ACKNOWLEDGMENTS

The authors wish to thank the obstetricians and midwives in the obstetrics department of Jilin University for their help in data analysis and the medical staff of the clinical research department for their valuable contributions to this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.646157/full#supplementary-material

REFERENCES

1. Huo CG, Zhang M, Ma FC. Factors influencing people's health knowledge adoption in social media: the mediating effect of trust and the moderating effect of health threat. *Libr Hi Tech*. (2018) 36:129–51. doi: 10.1108/LHT-04-2017-0074
2. Mira JJ, Lorenzo S, Navarro I. Hospital reputation and perceptions of patient safety. *Med Prin Pract*. (2014) 23:92–94. doi: 10.1159/000353152
3. Boerner H. A' Durable opportunity': ED overcrowding in the ACA ERA. *J Lab Phys*. (2016) 3:32–4. doi: 10.2105/AJPH.2014.302052
4. Elka A, Milena P, Wim G. Out-of-pocket patient payments for public health care services in bulgaria. *Front Public Health*. (2015) 3:175. doi: 10.3389/fpubh.2015.00175
5. Kheirbek RE, Beygi S, Zargoush M, Alemi F, Smith AW, Fletcher RD, et al. Causal analysis of emergency department delays. *Qual Manag Health Care*. (2015) 24:162–6. doi: 10.1097/QMH.0000000000000067
6. Janya M, Ross B, Roxanne SB. Continuous quality improvement and comprehensive primary healthcare: a systems framework to improve service quality and health outcomes. *Front Public Health*. (2018) 6:76. doi: 10.3389/fpubh.2018.00076
7. Qiu H, Luo L, Su Z, Zhou L,Chen YC. Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC Med Inform Decis Mak*. (2020) 20:83. doi: 10.1186/s12911-020-1101-8
8. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang LJ, Han WJ, et al. Effect of emergency department crowding on outcomes of admitted patients. *Ann Emerg Med*. (2013) 61: 605–11. doi: 10.1016/j.annemergmed.2012.10.026
9. Afilal M, Yalaoui F, Dugardin F, Amodeo L, Laplanche D, Blua P. Forecasting the emergency department patient's flow. *J Med Syst*. (2016) 40:175. doi: 10.1007/s10916-016-0527-0
10. Calegari R, Fogliatto FS, Lucini FR, Neyeloff J, Kuchenbecker RS, Schaan BD. Forecasting daily volume and acuity of patients in the emergency department. *Comput Math Methods Med*. (2016) 2016:3863268. doi: 10.1155/2016/3863268
11. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with Sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med*. (2016) 23:269–78. doi: 10.1111/acem.12876
12. Jiang S, Kwai-sang C, Tsui KL. A universal deep learning approach for modelling the flow of patients under different severities. *Comput Meth Prog Bio*. (2018) 154:191–203. doi: 10.1016/j.cmpb.2017.11.003
13. Juang WC, Huang SJ, Huang FD, Cheng PW, Wann SR. Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *BMJ Open*. (2017) 7:e018628. doi: 10.1136/bmjopen-2017-018628
14. Kadri F, Harrou F, Chaabane S, Tahon C. Time series modelling and forecasting of emergency department overcrowding. *J Med Syst*. (2014) 38:107. doi: 10.1007/s10916-014-0107-0
15. Spencer S, Jones A, Thomas R, Scott E, Shari J. Forecasting daily patient volumes in the emergency department. *Acad Emerg*. (2008) 15:159–70. doi: 10.1111/j.1553-2712.2007.00032.x
16. Wargon M, Brun-Ney D, Beaujouan L, Casalino E. No more winter crisis? Forecasting daily bed requirements for emergency department admissions to hospital. *Eur J Emerg Med*. (2018) 25:250–6. doi: 10.1097/MEJ.0000000000000451
17. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak*. (2003) 3:2. doi: 10.1186/1472-6947-3-2
18. Wei W, Jiang JJ, Liang H. Application of a combine model with autoregressive integrate moving average(arima) and generalized regressive neural network (grnn) in forecasting hepatitis incidence in heng county, china. *Plos One*. (2016) 11:e0156768. doi: 10.1371/journal.pone.0156768

19. Liu J, Mao Y. Patient satisfaction with rural medical services: a cross-sectional survey in 11 western provinces in china. *Int J Environ Res Public Health*. (2019) 16:3968. doi: 10.3390/ijerph16203968
20. Jilani T, Housley G, Figueredo G, Tang PS, Hatton J, Shaw D. Short and long term predictions of hospital emergency department attendances. *Int J Med Inform*. (2019) 129:167–74. doi: 10.1016/j.ijmedinf.2019.05.011
21. Wimsatt MA. Cross-jurisdictional sharing for emergency management-related public health: exploring the experiences of tribes and counties in California. *Front Public Health*. (2017) 5:254. doi: 10.3389/fpubh.2017.00254
22. Tan JY, Bennett R, Power J, Thompson L, Lwin Z. The wau (weighted activity unit) effect: evaluating the cost of DRG (diagnosis related group) coding. *Asia-Pac J Clin Onco*. (2015) 11:65. https://espace.library.uq.edu.au/view/UQ:366922
23. Hasanova R, Mentzakis E, Paolucci F, Shmueli A. Beyond DRG: the effect of socio-economic indicators on inpatient resource allocation in Australia. *Health Policy Technol*. (2018) 7:302–9. doi: 10.1016/j.hlpt.2018.07.001
24. Chien LC, Chou YJ, Huang YC, Shen YJ, Huang N. Reducing low value services in surgical inpatients in Taiwan: does diagnosis-related group payment work? *Health Policy*. (2020) 124:89–96. doi: 10.1016/j.healthpol.2019.10.005
25. Endrich O, Triep K, Torbica N, Nakas CT, Raio L. Changepoint analysis of gestational age and birth weight: proposing a refinement of diagnosis related groups. *Pediatr Res*. (2020) 87:910–6. doi: 10.1038/s41390-019-0669-0
26. Bertoli P, Grembi V. The political economy of diagnosis-related groups. *Soc Sci Med*. (2017) 190:38–47. doi: 10.1016/j.socscimed.2017.08.006
27. Vuagnat A, Yilmaz E, Roussot A, Rodwin V, Gadreau M, Bernard A, et al. Did case-based payment influence surgical readmission rates in France? A retrospective study. *BMJ Open*. (2018) 8:e018164. doi: 10.1136/bmjopen-2017-018164
28. Panagiotopoulos P, Maniadakis N, Papatheodoridis G, Pektasidis D. An evaluation of Diagnosis-Related Group (DRG) implementation focused on cancer DRGs in Greek Public Hospitals. *Pharmacoecon Open*. (2020) 4:61–9. doi: 10.1007/s41669-019-0146-z
29. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *ACM SIGKDD*. (2016). p. 785–94. doi: 10.1145/2939672.2939785
30. Wu Y, Hu H, Cai J, Chen R, Zuo X, Cheng H, et al. Machine learning for predicting the 3-year risk of incident diabetes in Chinese adults. *Front Public Health*. (2021) 9:626331. doi: 10.3389/fpubh.2021.626331
31. Choi HN, Dong HL. Bankruptcy prediction using ensemble SVM model. *J Korean Data Inform Sci Soc*. (2013) 24:1113–25. doi: 10.7465/jkdi.2013.24.6.1113
32. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
33. Talkhi N, Fatemi NA, Ataei Z, Nooghabi JM. Modeling and forecasting number of confirmed and death caused covid-19 in Iran: a comparison of time series forecasting methods. *Biomed Signal Process Control*. (2021) 66:102494. doi: 10.1016/j.bspc.2021.102494
34. Bergmeir C, Hyndman RJ, Benítez JM. Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. *Int J Forecasting*. (2016) 32:303–12. doi: 10.1016/j.ijforecast.2015.07.002
35. Dansana D, Kumar R, Adhikari JD, Mohapatra M, Le DN. Global forecasting confirmed and fatal cases of covid-19 outbreak using autoregressive integrated moving average model. *Front Public Health*. (2020) 8:580327. doi: 10.3389/fpubh.2020.580327
36. Holt CE. *Forecasting Seasonals and Trends by Exponentially Weighted Averages (O.N.R. Memorandum No. 52)*. Carnegie Institute of Technology. Pittsburgh USA (1957).
37. Winters PR. Forecasting sales by exponentially weighted moving averages. *Manage Sci*. (1960) 6:324–42. doi: 10.1287/mnsc.6.3.324
38. Wang H, Deng SH. A paper-text perspective: studies on the influence of feature granularity for Chinese short-text-classification in the big data era. *Electron Libr*. (2017) 35: 689–708. doi: 10.1108/EL-09-2016-0192
39. Fang R, Pouyanfar S, Yang Y, Chen SC, Iyengar SS. Computational health informatics in the big data age: a survey. *ACM Comput Surv*. (2016) 49:1–36. doi: 10.1145/2932707
40. Lisbeth S, Lindqvist R, Måns R. Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health care administrative data? *Health Policy*. (2003) 65:101–7. doi: 10.1016/S0168-8510(02)00208-7
41. Suleiman M, Demirhan H, Boyd L, Girosi F, Aksakalli V. Incorporation of expert knowledge in the statistical detection of diagnosis related group misclassification. *Int J Med Inform*. (2020) 136:104086. doi: 10.1016/j.ijmedinf.2020.104086
42. Wu D, Liang SB, Yu WT. Collaborative information searching as learning in academic group work. *Aslib J Inform Manag*. (2018) 70:2–27. doi: 10.1108/AJIM-03-2017-0063
43. Chen PS, Huang WT, Chiang TH. Applying heuristic algorithms to solve inter-hospital hierarchical allocation and scheduling problems of medical staff. *Int J Comput Int Sys*. (2020) 13:318–31. doi: 10.2991/ijcis.d.200310.004
44. Zou K, Li HY, Zhou D, Liao ZJ. The effects of diagnosis-related groups payment on hospital healthcare in China: a systematic review. *BMC Health Serv Res*. (2020) 20:112. doi: 10.1186/s12913-020-4957-5
45. Montefiori M, Pasquarella M, Petralia P. The effectiveness of the neonatal diagnosis-related group scheme. *PLoS ONE*. (2020) 15:e0236695. doi: 10.1371/journal.pone.0236695
46. Kim J, Yim E, Bae H. Performance of DRG-based reimbursement policy in national health insurance: eight years' experiences. *Value Health*. (2015) 18:A546. doi: 10.1016/j.jval.2015.09.1740

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fpubh.2021.766707

Using Satellite Images and Deep Learning to Identify Associations Between County-Level Mortality and Residential Neighborhood Features Proximal to Schools: A Cross-Sectional Study

Joshua J. Levy [1,2,3]*, *Rebecca M. Lebeaux* [1,2], *Anne G. Hoen* [1,2,4], *Brock C. Christensen* [1], *Louis J. Vaickus* [3] *and Todd A. MacKenzie* [1,4,5]

[1] Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States, [2] Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States, [3] Emerging Diagnostic and Investigative Technologies, Department of Pathology and Laboratory Medicine, Dartmouth Hitchcock Medical Center, Lebanon, NH, United States, [4] Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States, [5] The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Robertas Damasevicius,
Silesian University of Technology, Poland
Shihan Wang,
Utrecht University, Netherlands

***Correspondence:**
Joshua J. Levy
joshua.j.levy@dartmouth.edu

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *29 August 2021*
Accepted: *06 October 2021*
Published: *05 November 2021*

Citation:
Levy JJ, Lebeaux RM, Hoen AG, Christensen BC, Vaickus LJ and MacKenzie TA (2021) Using Satellite Images and Deep Learning to Identify Associations Between County-Level Mortality and Residential Neighborhood Features Proximal to Schools: A Cross-Sectional Study. Front. Public Health 9:766707. doi: 10.3389/fpubh.2021.766707

What is the relationship between mortality and satellite images as elucidated through the use of Convolutional Neural Networks?

Background: Following a century of increase, life expectancy in the United States has stagnated and begun to decline in recent decades. Using satellite images and street view images, prior work has demonstrated associations of the built environment with income, education, access to care, and health factors such as obesity. However, assessment of learned image feature relationships with variation in crude mortality rate across the United States has been lacking.

Objective: We sought to investigate if county-level mortality rates in the U.S. could be predicted from satellite images.

Methods: Satellite images of neighborhoods surrounding schools were extracted with the Google Static Maps application programming interface for 430 counties representing ∼68.9% of the US population. A convolutional neural network was trained using crude mortality rates for each county in 2015 to predict mortality. Learned image features were interpreted using Shapley Additive Feature Explanations, clustered, and compared to mortality and its associated covariate predictors.

Results: Predicted mortality from satellite images in a held-out test set of counties was strongly correlated to the true crude mortality rate (Pearson $r = 0.72$). Direct prediction of mortality using a deep learning model across a cross-section of 430 U.S. counties identified key features in the environment (e.g., sidewalks, driveways, and hiking trails) associated with lower mortality. Learned image features were clustered, and we identified 10 clusters that were associated with education, income, geographical region, race, and age.

Conclusions: The application of deep learning techniques to remotely-sensed features of the built environment can serve as a useful predictor of mortality in the United States. Although we identified features that were largely associated with demographic information, future modeling approaches that directly identify image features associated with health-related outcomes have the potential to inform targeted public health interventions.

Keywords: deep learning, satellite images, mortality, remote sensing, public health

INTRODUCTION

Life expectancy in the United States has increased dramatically over the past century from 48 years in 1900 to 80 years in 2019. However, the United States has experienced a drop in longevity over the past decade and now ranks 43rd in the world (1–5). Within the United States, crude mortality rates vary by more than 40%. Factors observed to be associated with mortality rates within the United States include disparities in socioeconomic status (6, 7) and health insurance coverage, (8) as well as obesity, smoking (9, 10), and drug use/opioid abuse (11).

Prior studies have attempted to infer characteristics of the underlying communities by characterizing land use using satellite imagery (12–14). Recent research has leveraged deep learning approaches to link the built environment to obesity (15, 16), socioeconomic status (17), poverty (18, 19), and other demographic factors (20, 21) by integrating lower level image features into higher order abstractions to make predictions (22, 23). In addition, the potential for deep learning-based image analysis to characterize a broad range of health exposures was recently reviewed (24). However, there is limited work that examines the relationship of mortality with the built environment in the United States at a large scale. The identification of information in satellite images associated with mortality predictions could potentially unlock previously unknown, yet related geographic and structural community characteristics that may be used to predict future mortality rates when demographic information is unavailable or possibly inform optimal county/state public health intervention strategies.

We hypothesized that county-scale level mortality rates could be predicted from satellite images. The goals of this study were to determine whether the analysis of satellite images by Convolutional Neural Networks (CNN) could be used to predict county-level mortality rates and uncover salient satellite image features associated with mortality. We also sought to determine if image features are related with county-level measures of income, education, age, sex, race, and ethnicity. The main objective of this study was to highlight a proof-of-concept deep learning application that presents strong baseline performance, enabling future work that can evaluate the potential for applying learnt features to the design of public health interventions in the built environment.

METHODS

Overview

Here, we will provide a brief overview of methods used to train, validate, test, interpret, and compare the selected machine learning models for the task of mortality prediction:

1. *Data collection*: We collected death certificate and county-level covariate/demographic information from private access census databases correspondent to 430 US counties, partitioning counties into training, validation and test sets. We downloaded 196 Google Maps satellite images per county.
2. *Predictive Modeling*: We trained, validated, and tested three separate models for comparison:
    a. A deep learning model which operates on images to predict the county-level mortality rate for each image. The predicted mortality rates for the images within each county are averaged with a trimmed mean to yield the final county-level mortality predictions.
    b. A linear regression model which can predict county-level mortality from county-level covariate information.
    c. A hybrid deep learning approach which can combine county-level covariates with image features to predict image-specific mortality rates, averaged (trimmed mean) across the county to yield county-level mortality rates.
    d. *Data Scaling Sensitivity Tests* to identify the amount of imaging information needed for optimal predictive accuracy for the image-only deep learning model.
3. *Interpretation Techniques* to identify important and potentially correspondent image and demographic predictors of mortality, using:
    a. Shapley additive explanations (*SHAP*) to generate a heatmap over each of the images to locate mortality associated image features.
    b. *Standardized regression coefficients* and *SHAP* to identify the most important demographic predictors of mortality from the regression model.
    c. *Unsupervised dimensionality reduction* to identify correspondence between the demographics and imaging predictors.

Graphical overviews of the aforementioned approaches may be found in **Figures 1**, **2** and in the **Supplementary Material**, section "Supplementary Overview of All Conducted Analyses."

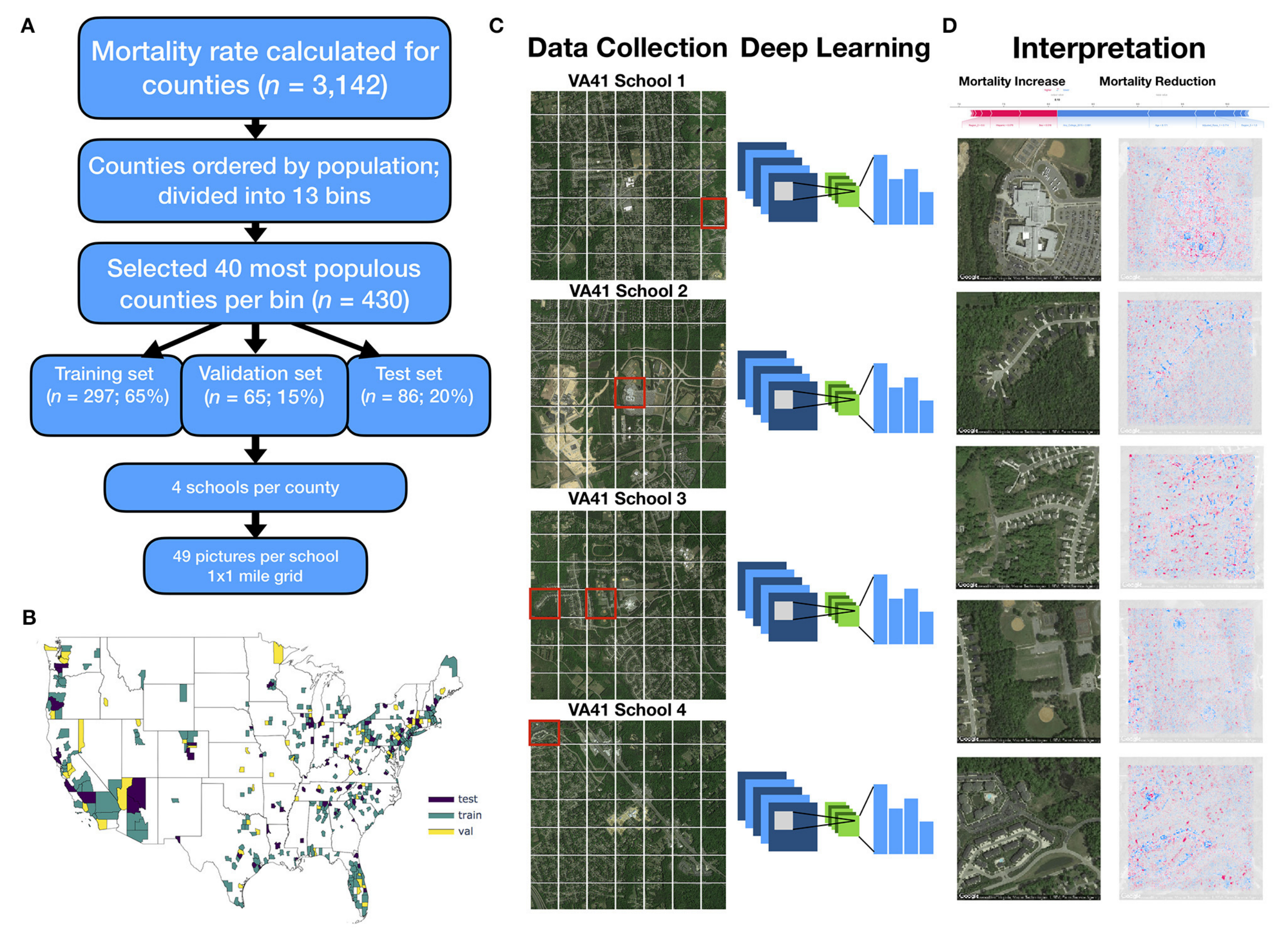


FIGURE 1 | Method overview: **(A)** flow diagram depicting selection of counties and sampling scheme for images; 13 mortality bins created from 1,000 most populous counties to capture diversity of mortality rates; **(B)** map of the United States colored by selection of training, testing and validation counties; **(C)** 4 sets of 1 × 1 mile grid of images (7 images by 7 images) were selected per each county, all images randomized and used to train deep learning model; visual depiction of deep learning: filters slide across each image and pick up on key image features, pooling layers reduce spatial dimensions of image, and output layers of neural network output one mortality rate prediction per image, $\hat{y}_i$; histogram visualization to the right of the neural network represents the set of mortality rate predictions $\{\hat{y}_i \mid i \in \{1, 2, \ldots, 195, 196\}\}$ corresponding to 196 images per county (49 images per school * 4 schools per county), which are averaged across using a trimmed mean to yield final mortality rates ($y_{\hat{county}} = \frac{\sum \hat{y}_i}{N}$); **(D)** SHAP is applied to neural network model and image to generate a heatmap of important image features across each image (red, positively associated with mortality; blue, negatively associated with mortality); these visualizations are compared to important covariates from linear model for each test county (visualized as force plot placed above the heatmap images; where presence of colors red/blue indicates positive/negative associations with mortality).

Data Collection

Extraction of County-Level Covariates

We used publicly available county-level mortality data. We used the CDC Wonder database to collect all available death data from 2015 for residents of the 50 United States and the District of Columbia, and matched death data with 2015 county Census population data from the Bureau of Economic Analysis, USDA ERS databases (25–27), and the Surveillance, Epidemiology, and Ends Results (SEER) Program (28). Crude mortality rates were calculated as the number of deaths in the county divided by the population in the county. Additional county level covariate information from 2015 was collected on age, sex, Hispanic status, race, income, education, and region (**Supplementary Table 1**).

Selection Criteria for Training, Validation, and Test Sets

We selected 430 counties from among 3,142 total United States counties in 2015 for inclusion in our study. This subset of representative counties was selected to decrease the computational burden and data storage resources. To reduce the variance of the mortality estimate and to limit potential bias from locale (e.g., rural, urban), we selected the top 1,000 most populous counties (restricting to more urban environments) and split these counties into 13 bins containing similar numbers of counties rank ordered by mortality. We selected up to 40 of the most populous counties in each bin. These selection criteria were developed to capture a greater proportion of total deaths and represent wider variation in mortality rates,

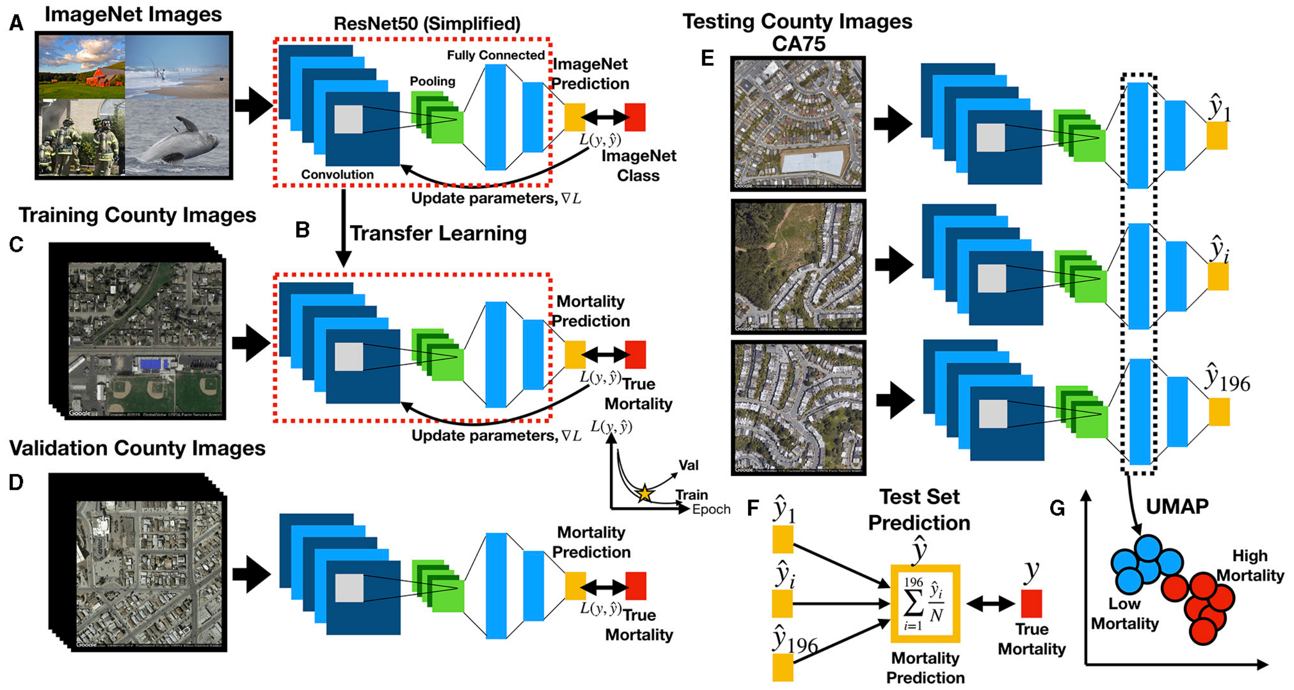


FIGURE 2 | Deep learning workflow: **(A)** ResNet50 model is first pre-trained on ImageNet images for the task of distinguishing 1,000 different everyday objects (e.g., farm, beach, firemen, whale); ResNet50 is comprised of multiple convolution and pooling layers (only one set shown for simplicity), which extract features and reduce the spatial dimensions of the image, respectively, and fully connected layers which represents objects at multiple levels of abstraction using vectors; prediction is compared to ground truth to estimate loss *L*; the gradient is backpropagated to update the model parameters; **(B)** transfer learning, where parameters from pretrained model (red dashed box) serve as the initial parameters for training the country-level mortality prediction model; output layer has been replaced with a layer to predict mortality; **(C)** training images pass through ResNet50 to predict county-level mortality, compared to true mortality to update model parameters; **(D)** during training, ResNet50 model predicts mortality from images from validation counties to estimate validation loss, which is compared to training loss over training epochs to determine when training should stop to avoid overfitting; **(E)** trained ResNet50 model predicts county-level mortality on 196 individual images (one mortality rate prediction per image) from each test county; **(F)** for each test county, 196 image-level mortality rate predictions are averaged using a trimmed mean to infer the overall county-level mortality rate; the overall county-level mortality rate is correlated to true county-level mortality rate; **(G)** image features from images of the test counties are extracted from ResNet50 fully connected layers in the form of feature vectors, which are used to generate UMAP plots to demonstrate how mortality and other corresponding covariates (extracted from the demographic data) separate/cluster based on the image features; Spectral Clustering is used to assign image features to blue and red clusters, representing low and high mortality related images, respectively.

ultimately resulting in the selection of the aforementioned 430 counties (**Supplementary Table 2**; in-depth description in **Supplementary Figure 1**). Of the selected counties, 279 counties were randomly placed into the training set (65%), 65 in the validation set (15%), and 86 in the test set (20%). The training set was used to update the parameters of the trained model while the validation set was used to limit the model from overfitting to the training data (**Figures 1A,B**, **2A–D**). The held-out test set represents an application of the modeling approach to unseen counties that had no role in the training of the model.

Acquisition of Imaging Data

Satellite imagery data was collected using the Google Static Maps application programming interface (API) to build our deep learning pipeline, similar to Maharana and Nsoesie (15). First, four schools from each county were randomly selected to serve as points of interest (POI) to sample nearby images. We used schools because they are typically placed in densely populated regions of the county (potentially serving as a representative proxy for residential neighborhoods of each county) and to reduce the computing space usage and required resources at our local computing cluster. We acknowledge that the selection of schools as a POI presents a limited view of the entire county. We obtained geographic coordinates for the schools in our study from the National Center for Education Statistics (29), and divided the 1 square mile area surrounding each school into an evenly spaced 7 images by 7 images grid (extracting 196 total images per county) (**Figure 1C**). The zoom level for each image was set to 17, and image dimensions were set to 400 pixels by 400 pixels to provide enough detail to make out street patterns and cover the space between the selected images of the grid. Eighty-four thousand two hundred eighty satellite images were downloaded using the Google Static Maps API. Images were downloaded between July and September 2019, representing a collection of images acquired April 2018 to December 2018.

This study was exempt from institutional review board approval as we accessed previously collected data and could not identify individuals. Our study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies where appropriate (30).

Predictive Modeling

Deep Learning Model Training

Deep neural networks have been used in a wide range of health-related applications (31, 32). Convolutional Neural Networks (CNN) slide filters across images to pick up on low level features such as edges or curves and then expand the visual field to pick up higher-order constructs [(33); **Figure 1C**]. The particular variant of CNN that was used to predict mortality on the satellite images was ResNet-50 (34), pretrained on the ImageNet database to recognize over 1,000 different objects (35), some of which may have features corresponding with our satellite image set (**Figures 2A,B**). We performed transfer learning, which applies knowledge gained from ImageNet to initialize the parameters of our model. These model parameters were updated to minimize the divergence between the true and predicted mortality rates using the negative Poisson log likelihood; training the model with this objective (models the outcome as a rate) allowed us to use each image to directly predict the mortality of its respective county. Training images were randomly cropped, rotated, resized, and flipped to improve the generalizability of the approach (**Figure 2C**). Predicted mortality rates from the individual images were averaged across the images for each county using a trimmed mean to comprise the final county-level prediction. A validation set of images from the validation counties were used to select optimal hyperparameters (e.g., learning rate, weight decay, early stopping criterion) to avoid memorization of the training data (**Figure 2D**). After a coarse hyperparameter search, the validation set was used to terminate the learning process at 5 training iterations and identify the ideal learning rate for the model, 1e-4. After training our model on images from 279 counties, we evaluated the model on images from the remaining test counties and averaged the predicted image-level mortality rates across each county using a trimmed mean to derive county mortality estimates (**Figures 2E,F**). The deep learning models developed using Python 3.6, utilized the PyTorch 1.3.0 framework and were trained using K80 Graphics Processing Units (GPUs).

Linear Regression on County-Level Covariates

A linear regression model to predict mortality using county level demographic characteristics was fit to data from both the training and validation sets, weighted by county population size (to reduce the variance of the parameter estimates), and evaluated on the test set as a comparison method versus the satellite image approach (**Supplementary Figure 2B**). We note that the goal of this study was to provide a benchmark for how well the CNN could predict mortality and to use covariates to help contextualize what the deep learning model is "seeing."

Hybrid Image-Covariate Deep Learning Approach

We also combined demographic information with satellite imagery data to test whether adjustment for covariates can improve prediction of mortality from satellite images (**Supplementary Figure 2C**, see **Supplementary Materials** section "Covariate Adjustment During Deep Learning Model Training and Evaluation").

Data Scaling Sensitivity Tests

A sensitivity analysis was utilized to decide the ideal number of schools and sampling area around the schools for the deep learning model (**Supplementary Figures 2E,F**). Preliminary tests from our modeling approach indicated that mortality prediction performance increases with the number of schools and area around the school sampled. However, performance saturates as the number of images in each county approaches 196, thereby warranting selection of images contained within a one-mile square area around schools to maximize the potential utility of assessing residential neighborhoods around schools as a POI and limit the amount of noise created beyond surveying neighborhoods in the immediate vicinity of schools. See **Supplementary Materials**, section "Effects of Sampling Larger Residential Areas Around Schools and Dataset Size on Predictions."

Model Interpretation

Image Interpretations With SHAP

Shapley additive feature explanations (SHAP) (36) is an analytical technique that explains complex models using a simpler surrogate model for each testing instance. We applied SHAP to the images of the test counties to form pixel-wise associations with increases or reductions in mortality; hotspots in these images denote important mortality-associated objects that the model has learned (**Figure 1D**). We also convolved learned CNN image filters over select images to further demonstrate which features of the built environment were utilized by deep learning model.

Important County-Level Covariates With SHAP and Standardized Regression Coefficients

SHAP was applied to identify county-specific important covariate predictors (**Figure 1D**) from the linear model, similar to how the SHAP image approach could identify important image-specific predictors. Overall effect estimates were reported using unstandardized regression coefficients. Standardized regression coefficients from the linear regression model served to explain the overall top mortality predictors.

Unsupervised Dimensionality Reduction to Identify Correspondent Demographic and Image Predictors

Correspondence between the covariate mortality predictors for each county and the image information was assessed through *embedding* of deep learning image features. Neural networks compress high-dimensional image data into lower dimensional representations in the process of making a prediction (**Figure 2G** and **Supplementary Figure 2D**). The output of an intermediate layer of the network was extracted from each image to form *embeddings*—reduced dimensional representations of the data in the form of vectors. These *embeddings* (represented by 1,000-dimensional vectors) could demonstrate how overall features of the images cluster and correspond with mortality or other demographics. We applied UMAP (37), an unsupervised dimensionality reduction technique, to visualize extract image features and identify clustering by images using scatterplots (one image is a point in the scatterplot). We overlaid the actual images themselves for each of the points, then true mortality and other covariates such as education and aging to visually demonstrate separation/clustering of important image-associated demographic and mortality characteristics as learned by the deep learning model. To examine specific associations between the pertinent image features (i.e., embeddings), we clustered image features using the Spectral Clustering approach (38) and found 10 clusters of images. We averaged the covariate information associated with each of the images across each grouping to yield characteristic covariate descriptors for each cluster. Weighted pairwise *t*-tests on mean differences between select covariates between all pairs of clusters demonstrated associations between extracted image features with covariate mortality predictors. As a direct means for assessing the relationship between deep learning predicted mortality and the covariates, we also regressed the image predicted mortality against each of these covariates.

RESULTS

The training, validation, and test counties included 1,721,052 deaths per 217,938,597 individuals in the 2015 U.S. population, a crude mortality rate of 7.90 deaths per 1,000 U.S. residents (**Supplementary Table 1**).

Predictive Modeling Results

Our deep learning model was able to accurately predict mortality, with a Pearson correlation coefficient of 0.72 ($R^2 = 51\%$) between the predicted and true mortality (**Figures 3A,C,D**) on a held-out test set. The linear regression model using county demographics also accurately predicted mortality; the R^2 was 90% (0.95 correlation) between covariates and the mortality rate (**Figure 3B**). Our covariate adjusted deep learning model appeared to improve mortality predictive performance (Test Set Pearson $r = 0.83$, $P < 0.001$; Validation Set Pearson $r = 0.87$, $P < 0.001$) but was unable to surpass the performance achieved using only the demographic characteristics (**Supplementary Materials**, section "Covariate Adjustment During Deep Learning Model Training and Evaluation") Sensitivity analyses over the included number of schools and area around schools indicated that mortality predictive performance tends to increase with the number of schools and area sampled around each school. However, predictive performance appeared to level-off after selection of 75 images per county (3 schools per county; 25 images per school; each image occupies $\frac{1}{49}$ square mile) (**Supplementary Materials**, section "Effects of Sampling Larger Residential Areas Around Schools and Dataset Size on Predictions").

Model Interpretation Results

Using SHAP to identify relevant features in satellite images across these test counties, we noticed that the model was able to associate common features associated with the socioeconomic status of that community to reductions in mortality. Generally, we were able to spot instances in which sidewalks, driveways, curved roads, hiking trails, baseball fields, and light-colored roofs were associated with reductions in mortality (**Figures 1D**, **4**). Conversely, we noticed instances where centerlines of large roads and shadows of buildings and trees were positively associated with mortality. To corroborate the evidence found using SHAP, we plotted small images representing patterns that the first convolutional layer had learned and then convolved each of these patterns with three select images (**Supplementary Figures 3–6**). Across three images from California, Virginia, and Georgia, filters such as 12, 29, and 43 focused on bright cues and were able to highlight driveways, sidewalks, walkways, and baseball fields, while filters 13 and 46 were able to pick up on contours and nuanced shape-based patterns.

From the county-level covariate linear model, we observed a reduction in 13 deaths per 100,000 individuals for an 1% increase in the proportion of those who attend college (regression coefficient $\beta = -12.6 \pm 0.8$), and increase of 214 deaths per 100,000 individuals for 5-year increases in average population age ($\beta = 2.1 \pm 0.1$); these were the most important predictors of county mortality (**Figures 5A–C** and **Supplementary Tables 3–6**). Increased Hispanic ($\beta = -6.6 \pm 0.5$), female ($\beta = 42.5 \pm 6.4$), or Asian race proportions ($\beta = -7.8 \pm 1.0$), and living in the western United States ($\beta = -1.1 \pm 0.3$) as compared to the other regions, were found to be protective county-level factors against mortality (**Figures 3A,B** and **Supplementary Tables 2–4**). County-level covariate mortality predictors are ranked in order of decreasing importance in **Figures 5A,B** and **Supplementary Tables 3–6**.

We identified demographic and mortality-related sources of significant variation between the 10 image feature clusters (**Figures 5C–E**, **6A**) established using UMAP and Spectral Clustering (**Figure 6B** and **Supplementary Table 7**). Images in cluster 7 were associated with the highest mortality rate of all of the clusters (mortality rate of 11 deaths per 1,000 individuals) and generally included counties from the southeastern United States (**Supplementary Figure 7**). Income and educational status were lower on average compared to the other clusters. We contrast cluster 7 with cluster 2, a small cluster with the third lowest mortality rates (6.6 deaths per 1,000 individuals) of the 10 clusters (**Figure 6** and **Supplementary Figures 8–11** and **Supplementary Table 7**) and identified variation within cluster 5; we include a brief discussion in the **Supplementary Materials**. Many of the covariate mortality predictors across the test set demonstrated strong associations with the deep learning predicted county-level mortality rates (**Supplementary Table 8**).

DISCUSSION

Here, we demonstrated the feasibility of using deep learning and satellite imagery to predict county mortality across the United States, extending methods and approaches from prior deep learning studies that linked the built environment and health-related factors. Furthermore, we established clusters of images that represented various demographic groups and interrogate these learned patterns to find additional corroborating evidence.

While the covariate model significantly outperformed our deep learning model and identified important predictors that corroborated with prior literature, the deep learning approach identified meaningful built features of the environment,

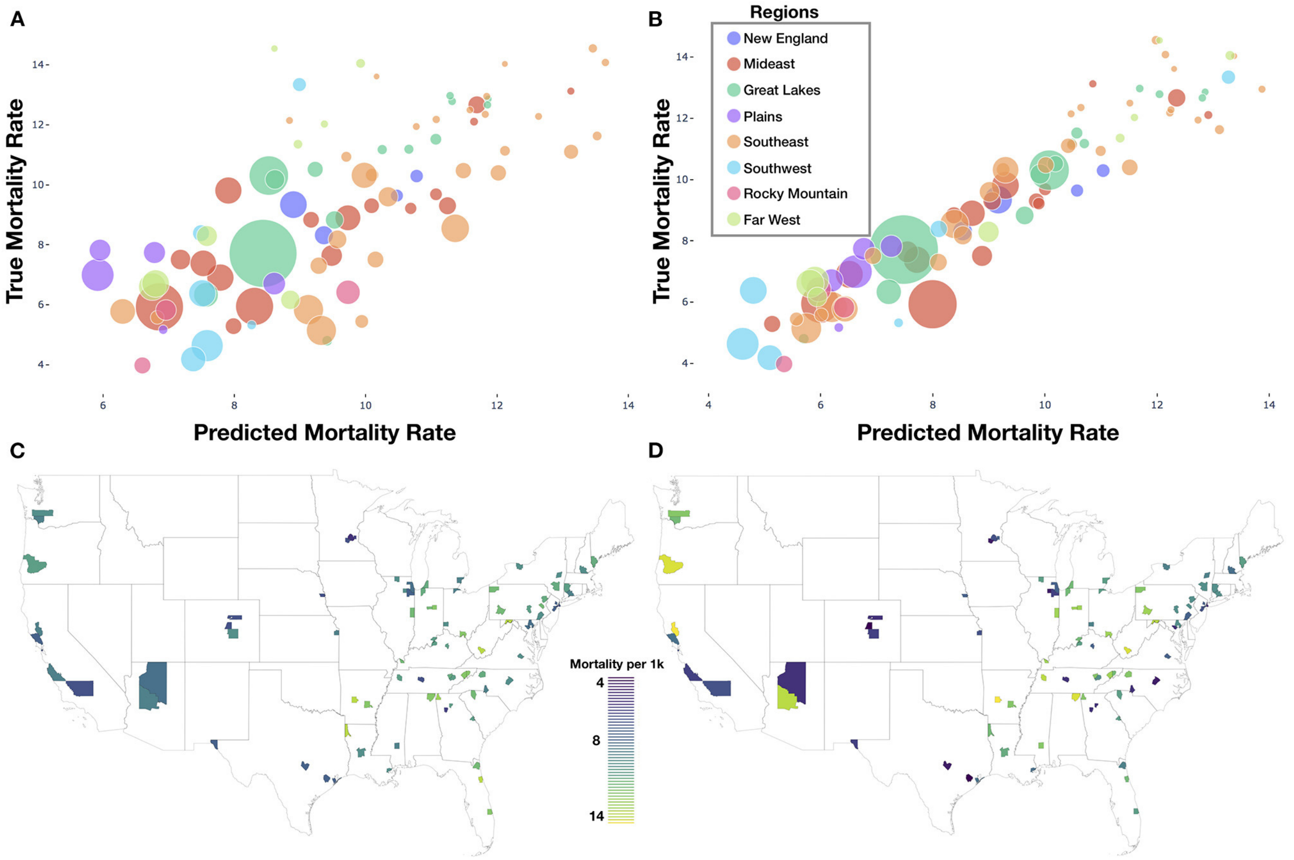


FIGURE 3 | Model results: **(A)** deep learning predicted vs. true mortality for each test county; **(B)** linear model predicted vs. true mortality for each test county; **(A,B)** legend contains look-up dictionary for naming of US regions featured in the rest of the study; size of bubble is correspondent to population size of county; **(C)** predicted mortality plotted geographically for test counties; **(D)** true mortality plotted geographically for test counties.

representing a benchmark for performance from which to compare future applications (39, 40).

We only sampled neighborhood characteristics found around schools. For instance, there is likely a greater number of sports fields and playgrounds surrounding schools than in random locations in the county. In many of the neighborhoods, regardless of mortality rates, there is likely more green space surrounding schools. While prior literature suggests green space is associated with positive health factors (41), the fact that this green space is nearly ubiquitous in neighborhoods near these schools may cause our model to down-weight these urban design factors. It also appeared that predictive performance increased by sampling the surrounding neighborhood around each school, suggesting that the surrounding neighborhood contains more important information associated with mortality than utilizing the schools alone.

The evidence of what our deep learning model found to be indicative of mortality can be corroborated with existing literature on associations of these image features with higher socioeconomic status, decreased obesity, and greater designs in urban planning (41–43). For instance, we found that sidewalks, driveways, curvy roads, hiking trails and baseball fields, amongst other factors, were related to lower mortality; these factors were also uncovered from inspection of learned convolutional filters (**Supplementary Figures 3–6**).

The assigned importance given to image features such as baseball fields does not imply that these components are necessary for urban design but does further elucidate a suite of modifiable community factors to design interventions. For instance, accessibility to trails, while indicative of a county with high socioeconomic status (44, 45), provides a convenient means to exercise but can also provide access to other portions of a community, allowing more social mobility and access to health care (46, 47).

Despite our strong findings, there were some limitations to our study. The design of this study was ecological with sampling at one time point for both images and mortality (cross-sectional). This precluded our ability to show temporality and that the built environment was causally associated with mortality. The images also were taken in 2018 while we assessed mortality in 2015. While we do not expect areas surrounding schools to change substantially, we cannot confirm the extent of land-use

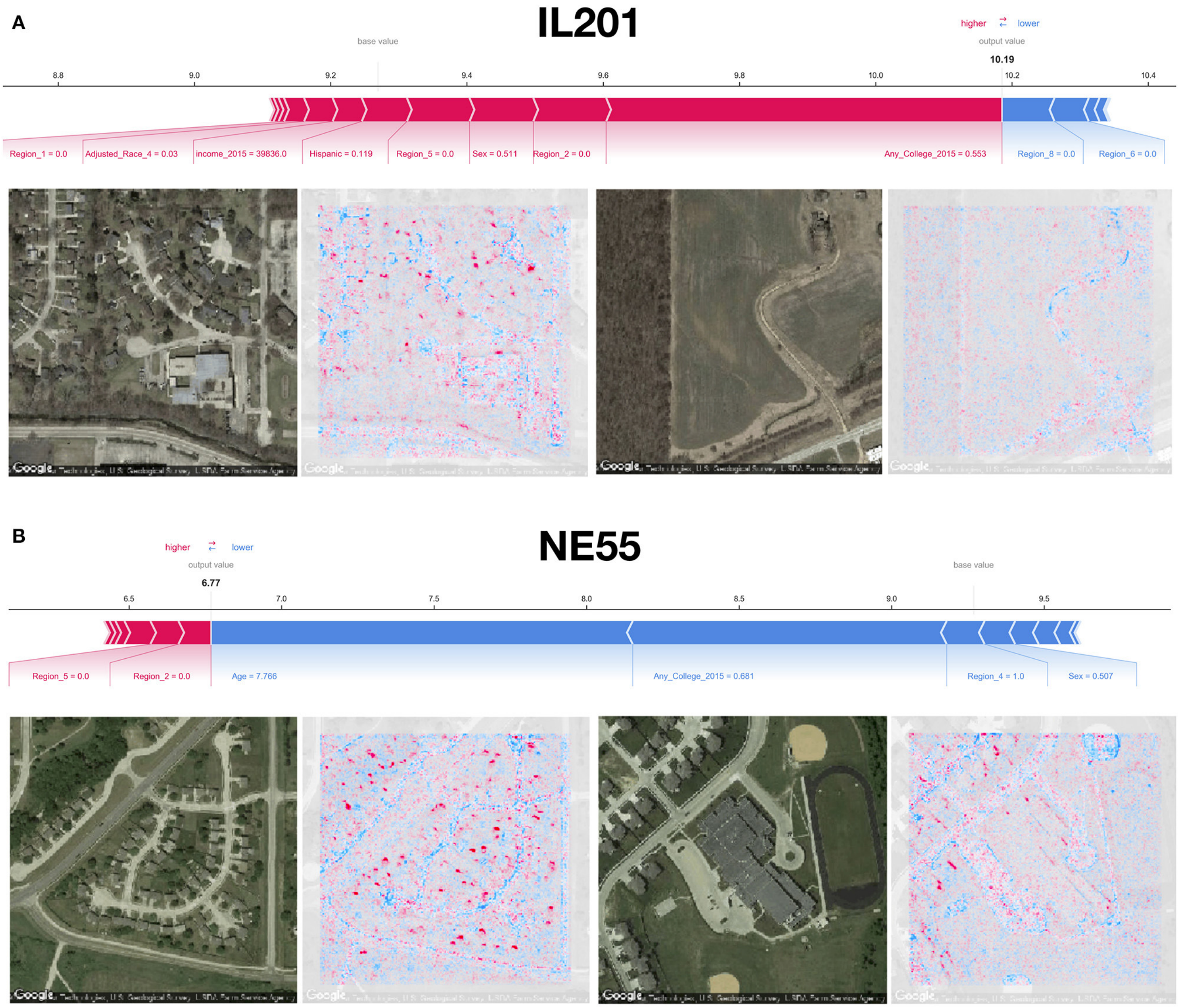


FIGURE 4 | SHAP image interpretations: **(A,B)** Shapley features extracted from both covariate and image models for select images from counties in Illinois and Nebraska (FIPS codes IL201 and NE55, respectively); blue coloring indicates features associated with reductions in mortality, red indicates association with increased mortality; additional image interpretations can be found in **Figure 1D**.

alterations between 2015 and 2018. We also acknowledge the shortfalls associated with utilizing counties as the primary spatial unit of analysis in this study. While county-level demographic and mortality information are more widely and freely available, mortality rates and demographic factors can vary widely within counties; thus, counties may not always represent the best unit of analysis for capturing the complete heterogeneity in population demographics and mortality rates. In future work, longitudinal image monitoring could be used to forecast increases in mortality rate and help determine optimal intervention strategies that complement county/state planning and development efforts. We assessed county mortality rates using a small percentage of schools from the county, making generalizations to counties included in the study and counties not in our dataset. We were able to predict crude mortality rates accurately within a subsample of these counties. However, our primary modeling approach could not delineate how much the corresponding model-identifiable neighborhood features were intermediates reflective of mortality associated demographics or were directly associated with mortality. Additionally, the deep learning model that combined county-level demographic and satellite imagery data did not surpass the performance of the demographic-only model. We incorporated county-level demographic data into images that have varying degrees of importance for the prediction of mortality; we suspect that adjusting and including the images that are less predictive of mortality may partially explain the performance of the demographic plus satellite image model.

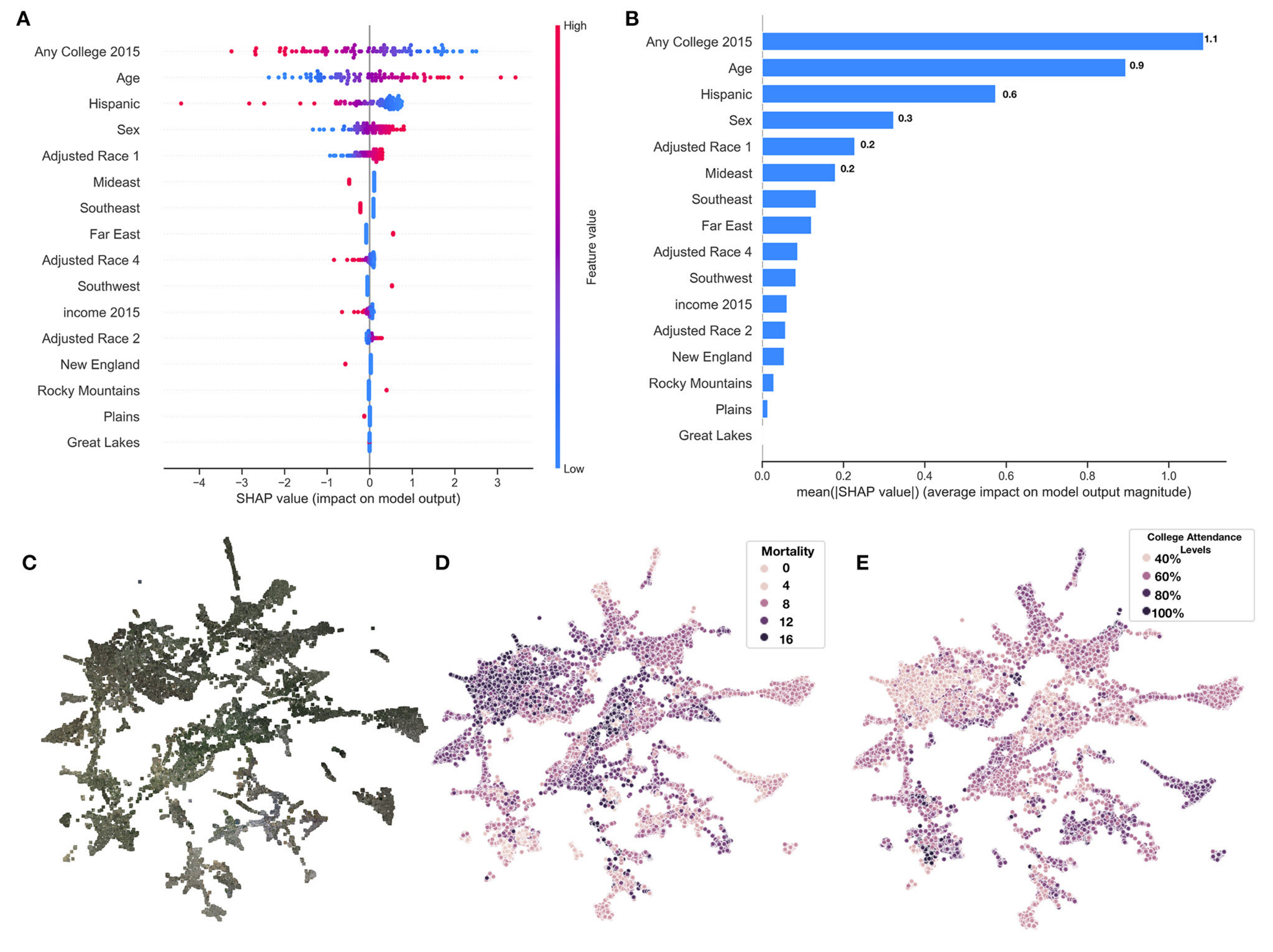
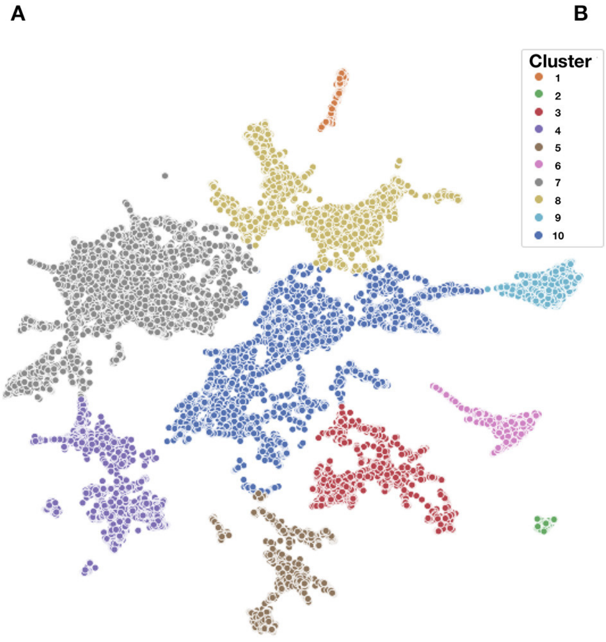


FIGURE 5 | SHAP summary of covariate predictors and image embeddings. **(A)** SHAP summary plot for covariate model evaluated on test set, covariates ranked by feature importance; **(B)** SHAP rankings of overall importance of covariates; **(C–E)** 2-D UMAP plot of image features derived from images using deep learning model: **(C)** actual images overlaying 2-D coordinates; **(D)** colored by true county mortality; **(E)** colored by average college attendance; we note here that our model has been explicitly trained to associate image features with mortality; the separability of college attendance over the image features is an artifact of the association between education status and mortality.

Higher correspondence between predicted and true mortality may be achieved by increasing the number of schools selected but tapers off with a larger sampling area, which is also suggestive that satellite images beyond a certain distance from a school are not as predictive of mortality. In addition, we may be able to identify more ubiquitous points of interest other than schools from which to sample. While schools were selected as a proxy for residential neighborhoods (prior research on impact of surrounding neighborhood of schools on health disparities), inclusion and exclusion of other points of interest (e.g., proximity to fast-food establishments) may result in more accurate models, which can potentially allow for in-depth hypotheses testing of the impact of urban planning on mortality and health disparities. Another opportunity is to integrate neighborhood land use information (e.g., residential/commercial) into the selection of images and adjustments during the modeling approach, though such approaches require accurate up-to-date mapping information (48–55).

While we acknowledge prior literature documenting the potential for shadow effects to confound aerial imagery analysis, we also note that our model was able to pick up on key factors associated with mortality by utilizing learned filters with shape, color, and intensity [(56, 57); **Supplementary Figures 1–4**]. Possible removal or augmentation of these shadows may cause the model to focus on other important characteristics pertinent to higher mortality prediction (56–58). Finally, although we sampled counties randomly, our sampling scheme preferred populous counties, and had assumed that these counties would solely contain suburban and residential land-use patterns. However, even populous counties contained rural areas that may have obfuscated our ability to sample ubiquitous land-use regions for the mortality prediction potentially biasing the results. Recent



B

Cluster	Predicted Mortality	True Mortality	Sex	Hispanic	Race White	Race Black	Race Asian
1	7.4	6.2	0.5	0.1	0.7	0.2	0.1
2	6.8	6.6	0.5	0.1	0.6	0.1	0.4
3	9.4	8.3	0.5	0.1	0.7	0.2	0.1
4	7.5	7.0	0.5	0.2	0.8	0.1	0.1
5	8.3	8.5	0.5	0.3	0.9	0.0	0.0
6	6.6	6.5	0.5	0.1	0.8	0.1	0.1
7	11.0	11.0	0.5	0.1	0.9	0.1	0.0
8	10.0	9.5	0.5	0.1	0.8	0.2	0.0
9	7.3	7.0	0.5	0.2	0.6	0.3	0.1
10	10.3	10.2	0.5	0.1	0.8	0.1	0.0

Cluster	Population	College Attendance (%)	Income 2015 (USD)	Age	Absolute Residual	Number Images
1	595 000	71.4	50 000	7.9	1.3	269
2	844 000	75.4	106 000	8.7	0.2	117
3	776 000	66.1	59 000	8.2	1.8	1502
4	677 000	64.6	59 000	7.9	1.4	1652
5	456 000	61.7	44 000	8.4	2.8	1205
6	731 000	73.1	62 000	8.0	1.8	691
7	206 000	54.9	43 000	8.6	1.3	4290
8	666 000	60.1	45 000	8.4	1.7	2465
9	1 138 000	58.0	52 000	8.2	0.6	686
10	275 000	58.0	45 000	8.5	1.7	3979

FIGURE 6 | Key covariate characteristics of found image clusters. **(A)** 2-D UMAP plots colored by cluster as estimated through Spectral Clustering; **(B)** averaged covariate statistics for each image cluster; average population and income estimates rounded to thousands place; statistics average county-level covariates, weighted by constituent images, covariates include: (1) average predicted mortality for the cluster, (2) average true mortality assigned to the cluster, (3) proportion female/male, (4) proportion Hispanic, (5) proportion white, black, or Asian, (6) average population, (7) proportion attend college, (8) average household income, (9) average age of population (in increments of 5-years per description in the **Supplementary Materials**; e.g., 7.9*5 years = 39.5 years old), (10) Absolute Residual between true and predicted mortality, (11) number of images for cluster/size of cluster.

deep learning works have focused on grasping health factors such as access to care in the rural setting (59) and could be employed in the context of mortality in the future.

The deep learning approaches used were able to achieve remarkable performance given technical challenges associated with sampling images over large geographic regions. Nor were covariate measures or temporality incorporated into the model. Different sampling techniques (60), feature aggregation measures, evaluating, or producing higher resolution images (61), direct estimation of neighborhood demographic factors, and segmentation of various land-use objects can potentially provide more accurate and interpretable models for studying health and disease (24, 62, 63). The images may have also been sampled during various points throughout the year, thus results may have been affected by seasonality, however, it is likely that the images had been collected randomly with respect to seasonality and the fact that the model was able to distinguish mortality rates further attests of the ability to learn features less tied to seasonality. Future applications could include deep learning methodologies that are able to account for effects of seasonality (64, 65). Transfer learning from other GIS-corroborated data may also improve the model's performance. Regardless, our approach is scalable and uses open access data enabling further exploration.

While the county-level covariate prediction model obtained higher accuracy vs. the deep learning model, a major limitation of the county-covariate approach pertains to evolving guidelines on the reporting of sex and race in public health research studies. County-level demographic characteristics were extracted from the Surveillance, Epidemiology, and Ends Results (SEER) Program, yet numerous publications have critiqued this reporting system for failing to incorporate sexual and gender minorities (e.g., gradations of sex identification, non-binary sex, LGBTQ+) and racial minority groups (e.g., description of individuals based on regional descent). Some of the criticisms of such reporting standards are that they enforce socially defined racial, ethnic, and gender divisions/constructs which further separates such groups (potentially contributing to health disparities they seek to study and reduce), disregards self-reported identification that transcends social structure imbued by support of such classifications, and reduces the potential to study additional meaningful health disparities between these minority groups (e.g., delineating health outcomes for individuals of Southeast Asian descent versus that of Asian Americans as a whole). These issues have made it difficult to study and document health outcomes for these minority groups. Conversely, some have pointed out that these reporting standards have still proven useful for studying health disparities (e.g., allostatic load and stress amongst minority groups and effect of segregation on health care access) in order to devise policies to alleviate these differences and that further subclassification may make it more difficult to assess meaningful differences. In response, standards have shifted toward asking multiple questions in demographic surveys which provide further clarification on self-identity and country of origin in addition to these coarse measures of sex and race. Conversations around the inclusion of sex and race in medical research studies are especially pertinent given persistent violence against minority groups that have more recently prompted a national conversation on such issues. In accordance with 2021 reporting standards on race and sex, we acknowledge the limitations of our findings with respect to these issues (using broad racial groups and binary sex identification), as such limitations reflect 2015 survey standards. While it is outside of the study scope to modify the analysis in response to data limitations, we have included a number of citations for the reader to explore facets of this important issue as part of a larger conversation on the role of sex and race in public health research studies (66–84).

Despite the study limitations, this approach to assess mortality using CNNs has much external applicability and room for improvement. Offshoots of this approach may further explore how the built environment affects mortality in more precise estimates, such as cities, or explore rural areas exclusively (59). Additionally, our approach demonstrates that built image features of the environment are correlated with demographic characteristics. Future mortality research that lacks the ability to attain particular covariate information due to feasibility or expense costs, could thus infer mortality using images alone or in combination with some known covariates. Applications beyond assessing mortality as an outcome could benefit from this approach and could use other landmarks as a sampling strategy for prediction instead of schools. Sampling from landmarks such as schools reduces the stochastic selection of images from large areas and potentially obviate the need to select and download huge swaths of images. However, if the researcher is able to overcome the resources required to download all of the images from within each county, future methodological advances should consider methods to assign importance scores to each satellite image or weight each county by their perceived macro-level importance for improved mortality prediction. Combining such methods with demographic information and contextualizing satellite images by neighboring images may yield models that ignore these demographic community factors. Ultimately, these tools could then be used by epidemiologists and policy makers to identify clusters of exposures or diseases such as cancer, arsenic exposure, or infectious diseases for more targeted interventions (24).

CONCLUSION

We found that mortality could be predicted from satellite imagery. Future use of deep learning and satellite imagery may assist in forming targeted public health interventions and policy changes.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Mortality death certificate data can be downloaded from CDC Wonder, per third-party request, available at the following URL: https://www.cdc.gov/nchs/data_access/cmf.htm. Google Maps Satellite Images are publicly available and can be downloaded using the Google Static Maps API. The code used to acquire and analyze the dataset and results can be found in the following GitHub repository: https://github.com/jlevy44/SatelliteCountyMortalityPrediction. Data for county-level covariate information is not available from our repository as permissions to access the data source are required. Requests to access these datasets should be directed to joshua.j.levy@dartmouth.edu.

AUTHOR CONTRIBUTIONS

JL, RL, and TM: concept and design, drafting of the manuscript, statistical analysis, and administrative, technical, or material support. TM: supervision. All authors participated in data acquisition, analysis, or interpretation, and critical revision to the manuscript for important intellectual content.

FUNDING

This work was supported by NIH grant R01CA216265. JL was supported through the Burroughs Wellcome Fund Big Data in the Life Sciences at Dartmouth. RL was funded under NIAID T32AI007519.

ACKNOWLEDGMENTS

The authors would like to thank Christopher L. Henry and Jorge F. Lima for discussion of image interpretations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.766707/full#supplementary-material

REFERENCES

1. Dyer O. US life expectancy falls for third year in a row. *BMJ.* (2018) 363:k5118. doi: 10.1136/bmj.k5118
2. Hamidi S, Ewing R, Tatalovich Z, Grace JB, Berrigan D. Associations between urban sprawl and life expectancy in the United States. *Int J Environ Res Public Health.* (2018) 15:861. doi: 10.3390/ijerph15050861
3. Muennig PA, Reynolds M, Fink DS, Zafari Z, Geronimus AT. America's declining well-being, health, and life expectancy: not just a white problem. *Am J Public Health.* (2018) 108:1626–31. doi: 10.2105/AJPH.2018.304585
4. Case A, Deaton A. Mortality and morbidity in the 21st century. *Brook Papers Econ Activity.* (2017) 2017:397–476. doi: 10.1353/eca.2017.0005
5. Woolf SH, Schoomaker H. Life expectancy and mortality rates in the United States, 1959-2017. *JAMA.* (2019) 322:1996–2016. doi: 10.1001/jama.2019.16932
6. Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States. *Evidence Natl Longitud Mortal Study. Ann Epidemiol.* (1996) 6:12–20; discussion: 21–2. doi: 10.1016/1047-2797(95)00090-9

7. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, et al. The association between income and life expectancy in the United States, 2001-2014. *JAMA*. (2016) 315:1750–66. doi: 10.1001/jama.2016.4226
8. Woolhandler S, Himmelstein DU. The relationship of health insurance and mortality: is lack of insurance deadly? *Ann Intern Med.* (2017) 167:424. doi: 10.7326/M17-1403
9. Carter BD, Abnet CC, Feskanich D, Freedman ND, Hartge P, Lewis CE, et al. Smoking and mortality — beyond established causes. *New Engl J Med.* (2015) 372:631–40. doi: 10.1056/NEJMsa1407211
10. Xu H, Cupples LA, Stokes A, Liu C-T. Association of obesity with mortality over 24 years of weight history: findings from the Framingham Heart study. *JAMA Netw Open.* (2018) 1:e184587. doi: 10.1001/jamanetworkopen.2018.4587
11. Case A, Deaton A. Rising morbidity mortality in midlife among White non-Hispanic Americans in the 21st Century. *Proc Natl Acad Sci USA*. (2015) 112:15078–83. doi: 10.1073/pnas.1518393112
12. Jensen JR, Cowen DC. Remote sensing of urban/suburban infrastructure and socio-economic attributes. In: *The Map Reader*. John Wiley & Sons, Ltd. (2011). p. 153–63. Available online at: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470979587.ch22 (accessed October 29, 2019). doi: 10.1002/9780470979587.ch22
13. Lo CP, Faber BJ. Integration of Landsat thematic mapper and census data for quality of life assessment. *Remote Sens Environ.* (1997) 62:143–57. doi: 10.1016/S0034-4257(97)00088-6
14. Tapiador FJ, Avelar S, Tavares-Corrêa C, Zah R. Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. *Int J Remote Sens.* (2011) 32:6437–56. doi: 10.1080/01431161.2010.512928
15. Maharana A, Nsoesie EO. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. *JAMA Netw Open.* (2018) 1:e181535. doi: 10.1001/jamanetworkopen.2018.1535
16. Nguyen QC, Sajjadi M, McCullough M, Pham M, Nguyen TT, Yu W, et al. Neighbourhood looking glass: 360º automated characterisation of the built environment for neighbourhood effects research. *J Epidemiol Commun Health.* (2018) 72:260–6. doi: 10.1136/jech-2017-209456
17. Suel E, Polak JW, Bennett JE, Ezzati M. Measuring social, environmental and health inequalities using deep learning and street imagery. *Sci Rep.* (2019) 9:6229. doi: 10.1038/s41598-019-42036-w
18. Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining satellite imagery and machine learning to predict poverty. *Science.* (2016) 353:790–4. doi: 10.1126/science.aaf7894
19. Tingzon I, Orden A, Sy S, Sekara V, Weber I, Fatehkia M, et al. *Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-Sourced Geospatial Information*. Karaj: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS) 5.
20. Keralis JM, Javanmardi M, Khanna S, Dwivedi P, Huang D, Tasdizen T, et al. Health and the built environment in United States cities: measuring associations using Google Street View-derived indicators of the built environment. *BMC Public Health.* (2020) 20:215. doi: 10.1186/s12889-020-8300-1
21. Nguyen QC, Khanna S, Dwivedi P, Huang D, Huang Y, Tasdizen T, et al. Using Google Street View to examine associations between built environment characteristics and U.S. health outcomes. *Prevent Med Rep.* (2019) 14:100859. doi: 10.1016/j.pmedr.2019.100859
22. Robinson C, Hohman F, Dilkina B. A deep learning approach for population estimation from satellite imagery. *arXiv:170809086 [cs]*. (2017). Available online at: http://arxiv.org/abs/1708.09086 (accessed October 29, 2019). doi: 10.1145/3149858.3149863
23. Gebru T, Krause J, Wang Y, Chen D, Deng J, Aiden EL, et al. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc Natl Acad Sci USA.* (2017) 114:13108–13. doi: 10.1073/pnas.1700035114
24. Weichenthal S, Hatzopoulou M, Brauer M. A picture tells a thousand…exposures: opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environ Int.* (2019) 122:3–10. doi: 10.1016/j.envint.2018.11.042
25. *Multiple Cause of Death Data on CDC WONDER*. Available online at: https://wonder.cdc.gov/mcd.html (accessed October 29, 2019).
26. *USDA ERS - County-Level Data Sets*. Available online at: https://www.ers.usda.gov/data-products/county-level-data-sets/ (accessed November 22, 2019).
27. *Personal Income by County, Metro, Other Areas*. U.S. Bureau of Economic Analysis (BEA). Available online at: https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas (accessed November 22, 2019).
28. *19 Age Groups - Standard Populations - SEER Datasets*. SEER. Available online at: https://seer.cancer.gov/stdpopulations/stdpop.19ages.html (accessed October 29, 2019).
29. *Geographic*. Available online at: https://nces.ed.gov/programs/edge/Geographic/SchoolLocations (accessed October 29, 2019).
30. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg.* (2014) 12:1495–9. doi: 10.1016/j.ijsu.2014.07.013
31. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interf.* (2018) 15:20170387. doi: 10.1098/rsif.2017.0387
32. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539
33. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc. (2012). p. 1097–105. Available online at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (accessed June 10, 2019).
34. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv:151203385 [cs]*. (2015). Available online at: http://arxiv.org/abs/1512.03385 (accessed July 31, 2019).
35. Deng J, Dong W, Socher R, Li L-J, Li K, Li FF. *ImageNet: a Large-Scale Hierarchical Image Database*. Miami, FL: IEEE CVRP (2009). p. 248–55. doi: 10.1109/CVPR.2009.5206848
36. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc. (2017). p. 4765–74. Available online at: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf (accessed June 10, 2019).
37. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv:180203426 [cs, stat]*. (2018). Available online at: http://arxiv.org/abs/1802.03426 (accessed March 5, 2019).
38. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* (2003) 15:1373–96. doi: 10.1162/089976603321780317
39. Muller A. Education, income inequality, and mortality: a multiple regression analysis. *BMJ.* (2002) 324:23. doi: 10.1136/bmj.324.7328.23
40. Montez JK, Zajacova A. Trends in mortality risk by education level and cause of death among US White Women from 1986 to 2006. *Am J Public Health.* (2013) 103:473–9. doi: 10.2105/AJPH.2012.301128
41. Gelormino E, Melis G, Marietta C, Costa G. From built environment to health inequalities: an explanatory framework based on evidence. *Prevent Med Rep.* (2015) 2:737–45. doi: 10.1016/j.pmedr.2015.08.019
42. Aneshensel CS, Harig F, Wight RG. Chapter 15 - Aging, neighborhoods, and the built environment. In: George LK, Ferraro KF, editors. *Handbook of Aging and the Social Sciences.* 8th ed. San Diego, CA: Academic Press (2016). p. 315–35. Available online at: http://www.sciencedirect.com/science/article/pii/B9780124172357000159 (accessed October 29, 2019). doi: 10.1016/B978-0-12-417235-7.00015-9
43. Rutt CD, Coleman KJ. Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX. *Prevent Med.* (2005) 40:831–41. doi: 10.1016/j.ypmed.2004.09.035

44. Wilson DK, Kirtland KA, Ainsworth BE, Addy CL. Socioeconomic status and perceptions of access and safety for physical activity. *Ann Behav Med.* (2004) 28:20–8. doi: 10.1207/s15324796abm2801_4
45. Engelberg JK, Conway TL, Geremia C, Cain KL, Saelens BE, Glanz K, et al. Socioeconomic and race/ethnic disparities in observed park quality. *BMC Public Health.* (2016) 16:395. doi: 10.1186/s12889-016-3055-4
46. Moore LV, Diez Roux AV, Evenson KR, McGinn AP, Brines SJ. Availability of recreational resources in minority and low socioeconomic status areas. *Am J Prev Med.* (2008) 34:16–22. doi: 10.1016/j.amepre.2007.09.021
47. Thornton RLJ, Glover CM, Cené CW, Glik DC, Henderson JA, Williams DR. Evaluating strategies for reducing health disparities by addressing the social determinants of health. *Health Aff.* (2016) 35:1416–23. doi: 10.1377/hlthaff.2015.1357
48. Hirt S. Home, sweet home: American residential zoning in comparative perspective. *J Plan Educ Res.* (2013) 33:292–309. doi: 10.1177/0739456X13494242
49. Hubbard T. *Limited Access to Healthy Foods: How Zoning Helps to Close the Gap*. Chapel Hill, NC: UNC Carolina Digital Library (2011).
50. Austin SB, Melly SJ, Sanchez BN, Patel A, Buka S, Gortmaker SL. Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments. *Am J Public Health.* (2005) 95:1575–81. doi: 10.2105/AJPH.2004.056341
51. Kestens Y, Daniel M. Social inequalities in food exposure around schools in an urban area. *Am J Prevent Med.* (2010) 39:33–40. doi: 10.1016/j.amepre.2010.03.014
52. Kwate NOA, Loh JM. Separate and unequal: the influence of neighborhood and school characteristics on spatial proximity between fast food and schools. *Prevent Med.* (2010) 51:153–6. doi: 10.1016/j.ypmed.2010.04.020
53. Bayer P, Ferreira F, McMillan R. A unified framework for measuring preferences for schools and neighborhoods. *J Politic Econ.* (2007) 115:588–638. doi: 10.1086/522381
54. Lareau A, Goyette K. *Choosing Homes, Choosing Schools*. New York, NY: Russell Sage Foundation (2014).
55. Mennis J. Increasing the accuracy of urban population analysis with dasymetric mapping. *Cityscape.* (2015) 17:115–26.
56. Kwatra V, Han M, Dai S. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In: *2012 IEEE International Conference on Computational Photography (ICCP)*. Seattle, WA: IEEE (2012). p. 1–8. doi: 10.1109/ICCPhot.2012.6215222
57. Corke P, Paul R, Churchill W, Newman P. Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Tokyo (2013). p. 2085–92. doi: 10.1109/IROS.2013.6696648
58. Bansal N, Akashdeep, Aggarwal N. Deep learning based shadow detection in images. In: Krishna CR, Dutta M, Kumar R, editors. *Proceedings of 2nd International Conference on Communication, Computing and Networking*. Singapore: Springer (2019). p. 375–82. doi: 10.1007/978-981-13-1217-5_37
59. Bruzelius E, Le M, Kenny A, Downey J, Danieletto M, Baum A, et al. Satellite images and machine learning can identify remote communities to facilitate access to health services. *J Am Med Inform Assoc.* (2019) 26:806–12. doi: 10.1093/jamia/ocz111
60. Jean N, Wang S, Samar A, Azzari G, Lobell DB, Ermon S. *Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data*. New Orleans, LA: AAAI (2018) doi: 10.1609/aaai.v33i01.33013967
61. Shermeyer J, Van Etten A. The effects of super-resolution on object detection performance in satellite imagery. *arXiv:181204098 [cs]*. (2019). Available online at: http://arxiv.org/abs/1812.04098 (accessed November 29, 2019). doi: 10.1109/CVPRW.2019.00184
62. Bischke B, Helber P, Folz J, Borth D, Dengel A. Multi-task learning for segmentation of building footprints with deep neural networks. In: *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei (2019). p. 1480–4. doi: 10.1109/ICIP.2019.8803050
63. Iglovikov V, Mushinskiy S, Osin V. Satellite imagery feature detection using deep convolutional neural network: a Kaggle competition. *arXiv:170606169 [cs]*. (2017). Available online at: http://arxiv.org/abs/1706.06169 (accessed November 29, 2019).
64. Fu H, Gong M, Wang C, Batmanghelich K, Zhang K, Tao D. *Geometry-Consistent Generative Adversarial Networks for One-Sided Unsupervised Domain Mapping*. Atlanta, GA: IEEE CVRP. 10.
65. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv:170310593 [cs]*. (2018). Available online at: http://arxiv.org/abs/1703.10593 (accessed November 29, 2019).
66. Salas LA, Peres LC, Thayer ZM, Smith RW, Guo Y, Chung W, et al. A transdisciplinary approach to understand the epigenetic basis of race/ethnicity health disparities. *Epigenomics.* (2021). doi: 10.2217/epi-2020-0080. [Epub ahead of print].
67. Lee C. "Race" and "ethnicity" in biomedical research: how do scientists construct and explain differences in health? *Soc Sci Med.* (2009) 68:1183–90. doi: 10.1016/j.socscimed.2008.12.036
68. Bhopal R, Donaldson L. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *Am J Public Health.* (1998) 88:1303–7. doi: 10.2105/AJPH.88.9.1303
69. Flanagin A, Frey T, Christiansen SL, AMA Manual of Style Committee. Updated guidance on the reporting of race and ethnicity in medical and science journals. *JAMA.* (2021) 326:621–7. doi: 10.1001/jama.2021.13304
70. Olufadeji A, Dubosh NM, Landry A. Guidelines on the use of race as patient identifiers in clinical presentations. *J Natl Med Assoc.* (2021) 113:428–30. doi: 10.1016/j.jnma.2021.02.005
71. Routen A, Pareek M, Khunti K. Reporting of race and ethnicity in medical and scientific journals. *JAMA.* (2021) 326:674. doi: 10.1001/jama.2021.9265
72. Flanagin A, Frey T, Christiansen SL, Bauchner H. The reporting of race and ethnicity in medical and science journals: comments invited. *JAMA.* (2021) 325:1049–52. doi: 10.1001/jama.2021.2104
73. Flanagin A, Christiansen S, Frey T. Reporting of race and ethnicity in medical and scientific journals. *JAMA.* (2021) 326:674–5. doi: 10.1001/jama.2021.9274
74. El-Sayed AM. Complex systems for a complex issue: race in health research. *AMA J Ethics.* (2014) 16:450–4. doi: 10.1001/virtualmentor.2014.16.6.stas1-1406
75. White K, Lawrence JA, Tchangalova N, Huang SJ, Cummings JL. Socially-assigned race and health: a scoping review with global implications for population health equity. *Int J Equit Health.* (2020) 19:25. doi: 10.1186/s12939-020-1137-5
76. Khunti K, Routen A, Banerjee A, Pareek M. The need for improved collection and coding of ethnicity in health research. *J Public Health.* (2021) 43:e270–2. doi: 10.1093/pubmed/fdaa198
77. Ioannidis JPA, Powe NR, Yancy C. Recalibrating the use of race in medical research. *JAMA.* (2021) 325:623–4. doi: 10.1001/jama.2021.0003
78. Mays VM, Ponce NA, Washington DL, Cochran SD. Classification of race and ethnicity: implications for public health. *Ann Rev Public Health.* (2003) 24:83–110. doi: 10.1146/annurev.publhealth.24.100901.140927
79. Clayton JA, Tannenbaum C. Reporting sex, gender, or both in clinical research? *JAMA.* (2016) 316:1863–4. doi: 10.1001/jama.2016.16405
80. Streed CG, Grasso C, Reisner SL, Mayer KH. Sexual orientation and gender identity data collection: clinical and public health importance. *Am J Public Health.* (2020) 110:991–3. doi: 10.2105/AJPH.2020.305722
81. *Race Recode Changes - SEER Documentation*. SEER. Available online at: https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/index.html (accessed October 1, 2021).
82. *SEER Data Dictionary for U.S. Population Estimates - SEER Population Data*. SEER. Available online at: https://seer.cancer.gov/popdata/popdic.html (accessed October 1, 2021).
83. Burkhalter JE, Margolies L, Sigurdsson HO, Walland J, Radix A, Rice D, et al. The national LGBT cancer action plan: a white paper of the 2014. National summit on cancer in the LGBT communities. *LGBT Health.* (2016) 3:19–31. doi: 10.1089/lgbt.2015.0118
84. Quinn GP, Sanchez JA, Sutton SK, Vadaparampil ST, Nguyen GT, Green BL, et al. Cancer and lesbian, gay, bisexual, transgender/transsexual, and

queer/questioning (LGBTQ) populations. *CA Cancer J Clin.* (2015) 65:384–400. doi: 10.3322/caac.21288

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

REVIEW
published: 23 November 2021
doi: 10.3389/fpubh.2021.641424

Understanding the Barriers and Facilitators to Sharing Patient-Generated Health Data Using Digital Technology for People Living With Long-Term Health Conditions: A Narrative Review

Emma Simpson[1], **Richard Brown**[2*], **Elizabeth Sillence**[2], **Lynne Coventry**[2], **Karen Lloyd**[3], **Jo Gibbs**[3], **Shema Tariq**[3] **and Abigail C. Durrant**[4]

[1] The NHS Business Services Authority, Newcastle upon Tyne, United Kingdom, [2] Department of Psychology, Northumbria University Newcastle, Newcastle upon Tyne, United Kingdom, [3] Institute for Global Health, University College London, London, United Kingdom, [4] Open Lab, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Priyamvada Paudyal,
University of Sussex, United Kingdom
Agnieszka Kitkowska,
Karlstad University, Sweden

***Correspondence:**
Richard Brown
richard6.brown@northumbria.ac.uk

Specialty section:
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: 14 December 2020
Accepted: 18 October 2021
Published: 23 November 2021

Citation:
Simpson E, Brown R, Sillence E, Coventry L, Lloyd K, Gibbs J, Tariq S and Durrant AC (2021) Understanding the Barriers and Facilitators to Sharing Patient-Generated Health Data Using Digital Technology for People Living With Long-Term Health Conditions: A Narrative Review.
Front. Public Health 9:641424.
doi: 10.3389/fpubh.2021.641424

Using digital technology to share patient-generated health data has the potential to improve the self-management of multiple long-term health conditions. Sharing these data can allow patients to receive additional support from healthcare professionals and peer communities, as well as enhance their understanding of their own health. A deeper understanding of the concerns raised by those living with long-term health conditions when considering whether to share health data *via* digital technology may help to facilitate effective data sharing practices in the future. The aim of this review is to identify whether trust, identity, privacy and security concerns present barriers to the successful sharing of patient-generated data using digital technology by those living with long-term health conditions. We also address the impact of stigma on concerns surrounding sharing health data with others. Searches of CINAHL, PsychInfo and Web of Knowledge were conducted in December 2019 and again in October 2020 producing 2,581 results. An iterative review process resulted in a final dataset of 23 peer-reviewed articles. A thorough analysis of the selected articles found that issues surrounding trust, identity, privacy and security clearly present barriers to the sharing of patient-generated data across multiple sharing contexts. The presence of enacted stigma also acts as a barrier to sharing across multiple settings. We found that the majority of literature focuses on clinical settings with relatively little attention being given to sharing with third parties. Finally, we suggest the need for more solution-based research to overcome the discussed barriers to sharing.

Keywords: data sharing, patient-generated health data, technology, long-term health conditions, trust, identity, privacy, security

INTRODUCTION

Over the last several decades there has been a substantial increase in life expectancy across the industrialized world due to advancements in digital technology and medicine, as well as successful public health initiatives (1, 2). Despite this achievement, an aging society has come with a rise in the prevalence of long-term health conditions (LTHCs) (3). Many LTHCs are supported by continuous self-monitoring and management. Advancements in digital technology have provided the opportunity for people to collect, manage and share personal health data to better manage their own health and achieve better health outcomes and quality of life. People living with LTHCs often record, monitor and manage personal health data, which encompasses a broad range of personal health information such as medication adherence, health and lifestyle practices and experiences of health, that patients may choose to share with others. These patient-generated health data (PGData) have the potential to improve the self-management of multiple conditions and, when shared with healthcare providers, improve the provision of care (4, 5).

There are multiple benefits to sharing PGData. Sharing these data can lead to a feeling of increased support when interacting with peer communities (others living with the same or similar condition), family or friends, as well as leading to better healthcare decision making in patients (6, 7). Using PGData from electronic devices has been shown to improve patient outcomes in a range of conditions such as diabetes, obesity, heart disease, and other chronic conditions (8). For example, in a study of cancer patients, the use of a digital app on an electronic tablet helped to improve patients' recall of symptoms and enabled the sharing of health information with clinicians (9). Cancer patients have also been reported to be willing to share PGData with cancer registries where they recognize the benefits for personal health management and population health (10). Patients who share PGData *via* digital platforms such as PatientsLikeMe report the greatest benefits to sharing as being able to learn more about their symptoms and to understand the side effects of their treatment (11). Furthermore, the increased sharing of PGData with third parties may allow big data public health practices to identify previously concealed patterns among the reported experiences of multiple LTHCs, which may help to optimize the delivery of care for individual patients (12, 13). Ultimately, the use of PGData in the management of health conditions enhances understanding and generates a holistic picture of one's personal health and disease management (14, 15).

There are a number of factors that facilitate the sharing of PGData, such as individual altruistic tendencies and the seeking of social support (16). Conversely, factors that are considered barriers to the sharing of PGData include poor health literacy and the perceived burden of having to manage data associated with one's condition(s) (17). The growing prevalence of digital technology in the transmission of personal health data would suggest that issues surrounding Trust, Identity, Privacy and Security (TIPS) are likely to be an increasing and evolving concern. For example, TIPS concerns have been found to be critical when seeking to facilitate the sharing of PGData among those living with HIV (18). This narrative review is conducted as part of a UK EPSRC funded programme ("INTUIT: Interaction Design for Trusted Sharing of Personal Health Data to Live Well with HIV", 2020) (19) examining TIPS concerns around the sharing of PGData primarily among those living with HIV, but also looks to investigate TIPS concerns among those living with a range of other LTHCs. The INTUIT project aims to contribute toward removing barriers to collecting and sharing PGData in order to improve the health and well-being of stigmatized populations. The sharing of PGData raises multiple TIPS concerns for those living with LTHCs and may hold particular significance for those with potentially stigmatized conditions due to fears of discrimination or other harmful consequences. People who anticipate experiences of stigma as a result of their LTHC(s) are likely to be more guarded when reporting their experiences of health, which may prevent them from receiving an appropriate level of care (20, 21). Therefore, understanding the role that both stigma and TIPS concerns play in the sharing of PGData with others, by those living with LTHCs, may help to promote effective data-sharing practices, potentially leading to improved delivery and self-management of care.

The potential benefits of PGData for understanding a range of health conditions and for optimizing delivery of care may help to support the rising prevalence of LTHCs. The use of PGData has the potential to transform the delivery of healthcare and to improve the management of countless LTHCs (4). However, cultivating an ecosystem that protects the interests of patients and builds confidence that healthcare systems will use personal information responsibly presents unique challenges to researchers, designers and policy makers working in digital health. To realize the benefits of PGData we must first understand the barriers and facilitators to sharing using digital technology for people living with LTHCs. To address this, we have conducted a narrative review of previous literature addressing TIPS concerns and the role of stigma in the sharing of PGData *via* digital technology by those living with LTHCs. The research questions directing this narrative review are (i) do TIPS concerns present a barrier to the successful sharing of PGData using digital technology by people living with LTHCs; and (ii) what is the impact of stigma on the sharing of PGData *via* digital technology by those living with LTHCs? By addressing these research questions, we aim to discuss barriers and facilitators to the effective sharing of PGData across multiple contexts: sharing with clinical staff, public health surveillance, researchers, peer communities, friends, social networks and other third-party organizations.

METHODS

Narrative Review

Narrative reviews are fast becoming the most common form of literature review across multiple disciplines (22). Though the literature is summarized in a way that is not explicitly systematic, narrative reviews nevertheless provide a comprehensive synthesis of up-to-date evidence for researchers, designers and policy makers working in the field of digital health (22–24). The synthesis of qualitative and quantitative research is critical to ensuring that patient experiences, needs and preferences are understood and taken into consideration when designing and

TABLE 1 | Inclusion and exclusion criteria for selecting peer-reviewed articles.

Inclusion criteria	Exclusion criteria
• Addresses any of the selected LTHCs (HIV, diabetes (types 1, 2 and unspecified), mental health, sexual health, cancer, cardiovascular disease or dementia); and • Includes a type of communication (with peers, with clinical staff or with public health surveillance); and • Includes a form of digital technology (social media, online forums, mobile apps or other digital platforms); and • Addresses the sharing of PGData; and • A barrier to sharing; or • A facilitator to sharing; or • Considers issues surrounding Trust, Identity, Privacy and Security.	• Addresses the sharing of generic health promotion/education/information; or • Focusses on a specific LTHC outside of the selected categories; or • Does not present empirical data.

implementing healthcare technology (24). In conducting this narrative review, a scale for the quality assessment of narrative review articles (SANRA) was consulted in order to ensure that it meets the expected standards for this category of review (22). This narrative review aimed to better understand issues of Trust, Identity, Privacy and Security (TIPS) in those living with LTHCs when using digital technology to share their personal health and lifestyle data. This review also explores the role that stigma plays in sharing this data *via* technology by people with LTHCs.

Inclusion and Exclusion Criteria

This narrative review was conducted by first establishing the inclusion and exclusion criteria for article selection, which was agreed by the whole research team (see **Table 1**). The LTHCs featured in this inclusion criteria were in line with the wider goals of the INTUIT project and based on the findings of previous research that discussed experiences of stigma among those living with HIV (18, 25, 26), other sexually transmitted infections (27, 28), diabetes (29–31) and Mental Health conditions (32–34). Our inclusion criteria also sought to capture those LTHCs considered most prevalent and impactful on society (cancer, cardiovascular disease and dementia) (35).

The Initial Search

The inclusion and exclusion criteria were then applied to an initial search exercise conducted in December 2019. This initial search was conducted by one member of the research team and involved a search of the available published literature using the following databases: CINAHL, PsychInfo, Web of Knowledge and by referring to the reference lists of relevant articles. An iterative searching strategy was developed as the language and terminology pertaining to PGData became more familiar to the researcher. Within current health literature, there are multiple variations of terms that are used to describe PGData, including "personal health information," "personal health data," "patient-authored information," "patient-generated information," "protected health information," whereas other literature may simply refer to the data as "medical information". Combinations of words and strings representing the sharing of PGData were applied to the selected databases with Boolean operators "AND" and "OR" to broaden the search. This initial search exercise yielded 2,479 results.

Refining the Search

One member of the research team collected the initial articles from the various sources. Duplicates were removed. An iterative process of reading the titles and excluding search results whose titles indicated that they did not satisfy any of the inclusion criteria or contained a relevant feature of the exclusion criteria (see **Table 1**). The abstracts and texts of search results whose titles passed this initial inspection were then reviewed by three members of the research team to determine their relevancy in accordance with the full inclusion and exclusion criteria, thus progressively refining the scope of the initial search.

Article Selection

Three members of the research team independently reviewed the list of potentially relevant articles against the inclusion and exclusion criteria. A meeting was held to compare lists and agree which to take forward. Any articles where one member of the team had identified them for inclusion were discussed and a decision made by mutual agreement. One member of the research team meticulously reviewed the full text for articles that the research team identified as potentially (though not certainly) relevant to the directives of the review. For example, for articles that addressed various health conditions, the researcher examined the text to ensure that significant attention was given by the candidate article to the sharing of personal health information associated with LTHCs. This member also extracted any relevant articles from the references of the candidate articles. Each time new articles were identified the three first reviewers would meet and discuss their inclusion. The full research team evaluated and discussed the short list of candidate articles with respect to the selection criteria and were given the opportunity to suggest any articles known to them that had been missed. This process resulted in 19 peer-reviewed articles being selected by mutual agreement.

Updating the Search

The search, refinement and selection processes described above were repeated in October 2020 to identify further contributions that had been made to the literature since the initial search. The second search produced a further 102 results, four of which were selected for inclusion in the narrative review.

Review

The final dataset comprised 23 peer-reviewed articles. The results from the articles were extracted into Microsoft Excel before NVivo 12 was used to thematically analyse the data. The thematic analysis of the selected articles was undertaken by all members of the research team and involved an iterative review of the findings in consideration of their relevance to the two research questions stated above. All members of the research team mutually discussed the results of the selected articles and subsequent thematic analysis in order to synthesize and present the findings below.

FINDINGS

The review of the selected articles finds that issues surrounding Trust, Identity, Privacy and Security clearly present barriers (but in some cases facilitators) to the sharing of PGData across all contexts (i.e., sharing with clinical staff, public health surveillance, researchers, peer communities, friends, social networks and other third-party organizations). Examples of the specific TIPS issues referred to in the literature, along with a brief overview of the selected articles, are presented and discussed below to provide a review of the literature thus far. **Table 2** provides a description of all of the articles included in this review.

From the selected studies, many focus exclusively on specific LTHCs: diabetes (types 1, 2 and unspecified; $n = 4$), HIV ($n = 4$) and mental health ($n = 4$). One study specifically addresses patients who manage multiple chronic conditions (MCC) and the remainder of the studies comprise participants who have a range of different LTHCs ($n = 10$). One study looking at type 1 diabetes reports the perspectives of adolescent participants (12–17 years) (31) and the remaining studies are of adults participants (18–84 years). The majority of the included studies explore the sharing of PGData with healthcare providers and electronic health record management (17, 26, 29, 30, 33, 34, 36, 40, 41, 47), with some including sharing of data with a wider network including public health and researchers (38, 45). Three of the studies look at the implications of sharing PGData online through social networking sites such as Facebook (16, 31, 45). One study looks at Grindr and the sharing of HIV status (44), whilst the other HIV related studies look at health information technology more broadly (25, 26). The following sections discuss the results in relation to the research questions driving the review.

RQ1: DO TIPS CONCERNS PRESENT A BARRIER TO THE SUCCESSFUL SHARING OF PGDATA USING DIGITAL TECHNOLOGY BY PEOPLE LIVING WITH LONG-TERM HEALTH CONDITIONS?

This narrative review finds that multiple TIPS concerns present barriers to the sharing of PGData *via* digital technology by those living with LTHCs. Distrust in the proposed recipient of PGData inhibits sharing *via* technology. Trust is often shaped by patients' previous experiences of sharing and, in a clinical context, can be facilitated by confidence in the healthcare institution or team with whom sharing is proposed. The desire by patients to control and self-manage their digital identity also impacts on patient willingness to share PGData with others. However, the review suggests that the use of pseudonyms can offer a successful strategy for facilitating sharing of PGData online by those living with LTHCs. Privacy and security concerns present clear barriers to sharing PGData *via* technology. Privacy concerns are reported as being the main reason patients may choose not to share PGData in a clinical context, though these concerns mostly relate to the potential for future sharing with external third parties. Anticipated security breaches by patients also present a barrier to the sharing of PGData with others, whereas believing that digital technology has sufficient safeguards in place is a facilitator to sharing PGData *via* technology. A more detailed discussion of individual TIPS concerns is given below.

Trust

Here we address the degree of trust or distrust that is established between an individual and the proposed recipient of their PGData. A quarter of the articles discussed 'trust' in relation to the sharing of PGData (17, 18, 26, 30, 31, 33, 42, 44). In the majority of these papers, trust as a barrier to the sharing of PGData centered on *distrust of the recipient.* When sharing with healthcare providers and clinical staff, distrust can be shaped by previous negative experiences for people living with multiple chronic conditions (17). Distrust is also developed when patients are asked to provide information that they deem to be highly personal and irrelevant to the given context (30). On the other hand, developing and building trust with recipients is considered a facilitator to the sharing of PGData and is supported by familiarity and confidence in the healthcare institution and healthcare team (25, 26). Where, for example, Teixeira et al. (26) report on willingness to share data for patients living with HIV:

> *"Patients reported having a great deal of trust in their HIV care team. Trust in their care team to deliver high-quality medical care and feeling that providers spent enough time with them were each associated with patients' willingness to share PHI [protected health information] with both clinical and nonclinical staff at their primary clinic"* (26)

The majority of the papers examine PGData sharing within a clinical context, focusing on the barriers and/or facilitators to sharing with HCPs *via* digital technology. In this setting, trust is a key issue that makes patients more likely to share PGData with trusted recipients. Kelley et al. (33) report how sharing PGData improved the relationship and trust between patients and their clinicians, with student participants reporting how they used PGData to provide proof that they were doing exactly what they said they were. We know that higher levels of patient trust in HCPs are associated with more beneficial health behaviors, fewer symptoms, and higher quality of life (48). Conversely, a lack of trust in HCPs can prevent patients from sharing some forms of PGData and engaging with HCPs in face-to-face settings (49).

This review indicates that trust remains an important factor in PGData sharing *via* digital technology. Most papers focussing on a clinical setting examine data provided by patients that constitutes personal health information that they have chosen to incorporate into their electronic health records (EHR). In general, these studies indicate patients are happy to share most information with HCPs but less so with non-clinical staff (26). The focus on the EHR as a digital artifact provides common ground for the patient and the HCP. Shared data can underpin improved communication between patients and

TABLE 2 | Included papers overview.

References	Country of origin	Aim/purpose	Long-term health condition (LTHC)	Population	Sharing data with/platform	Key findings
Agaku et al. (36)	USA	This study assessed the perceptions and behaviors of US adults regarding the security of their protected health information (PHI).	Various conditions	$n = 1,452$ adults	Healthcare professionals (HCPs)	This study reported that most US adults are concerned about the security and privacy of their PHI, and that such concerns are associated with an increased likelihood of non-disclosure of sensitive information to HCPs.
Ancker et al. (17)	USA	This study investigated how patients with multiple chronic conditions (MCC) manage their personal health records and information sharing with HCPs. This study also addressed how patients perceive their own role in managing their health information.	MCC	$n = 22$ adults	HCPs	Personal health information management should be recognized as an additional burden that MCC places upon patients. Effective structural solutions for information sharing, whether institutional ones such as care management or technological ones such as electronic health information exchange, are likely not only to improve the quality of information shared but reduce the burden on patients already weighed down by MCC.
Bernaerdt et al. (37)	Belgium	This study investigated the perceptions and attitudes of vulnerable patients regarding sharing medical information with HCPs and third parties *via* a digital platform.	Various conditions	$n = 14$ adults	Digital patient portal for sharing with HCPs and third parties.	Patients expressed concerns about privacy and security risks. Patients were generally unaware of the meaning and value of health data to third parties which resulted in inconsistent views on data sharing. Patients desire granular control over their medical information but believe that this may negatively impact their quality of care. There is a need for more transparency about the potential consequences of sharing data with third parties.
Bussone et al. (18)	UK	This study investigated the TIPS considerations that people living with HIV make when sharing data with their peers for the purpose of guiding the development of trusted digital tools.	HIV	$n = 26$ adults	Digital health communities (sharing with peers)	TIPS concerns are central to those living with HIV when deciding whether or not to share personal health information with others. Platforms that are associated with a familiar HIV-related organization or charity benefit from enhanced trust. Robust privacy and security measures are key to ensuring trust in digital peer sharing platforms.
Caine and Hanania (38)	USA	The aim of this study was to assess patients' desire for granular level privacy; this includes control over which personal health information should be shared, with whom, and for what purpose. The study also addressed whether these preferences vary based on the sensitivity of health information.	Various conditions	$n = 31$ adults	Multiple recipients	Patients expressed a clear desire for control over which health information should be shared and with whom. Patients also expressed differences in sharing preferences for sensitive vs. less-sensitive health data.
Esmaeilzadeh et al. (39)	USA	This study aimed to examine the interplay between different chronic health problems and different types of sharing interfaces in relation to patient willingness to share personal health information with HCPs.	Chronic mental illness and chronic physical illness	$n = 607$ adults	Structure and unstructured interfaces for sharing personal health information with HCPs.	The results described how individuals managing physical illnesses and mental disorders both favor highly structured data entry interfaces for sharing personal data. Mental health patients perceived less psychological risk, and reported lower privacy concerns when using a well-structured data entry interface to record their PHI compared to an unstructured interface.
Fergie et al. (16)	UK	The aim of this qualitative study was to explore how engagement with user-generated content can support people with LTHCs, and to explore the factors that limit users' adoption of these technologies.	Diabetes (type unspecified) and Common Mental Health Disorders (CMHD)	$n = 40$ adults	Social media	This study highlighted the complexities of users' engagement with user-generated content for support in their experience of LTHCs. The findings highlight the range of considerations which influence production and consumption of health content *via* social media, particularly around identity management and integrating health content into everyday online practice.

(Continued)

TABLE 2 | Continued

References	Country of origin	Aim/purpose	Long-term health condition (LTHC)	Population	Sharing data with/platform	Key findings
Fuji et al. (29)	USA	The purpose of this qualitative study was to explore how patients with type 2 diabetes use an Electronic Health Record (EHR) to manage their information for the purpose of self-care.	Type 2 diabetes	$n = 59$ adults	HCPs *via* an EHR	Patients valued being able to store their medical data on one electronic record that was easily accessible. However, most participants did not share their data with HCPs. Patients expect HCPs to have full access to their data without having to personally disclose it. A strong patient-provider relationship is important for the effective adoption of EHRs.
Hartmann et al. (32)	Germany	The aim of this study was to investigate the self-monitoring and self-management of depression as well as to explore the data sharing preferences of potential users of digital platforms.	Depression	$n = 668$ adults	Mobile apps	Individuals with depression want to take control of sensitive data, they do not want to share with everyone - particularly third parties. Individuals are concerned about tracking, particularly when they perceive that being tracked to a specific place could be used against them.
Kelley et al. (33)	USA	The aim of this study was to investigate student perspectives on self-tracking of mental health and how personal data is used to support mental health and wellness management.	Mental health	focus group $n = 14$, survey $n = 297$ students (18–24 years)	Multiple recipients *via* self-tracking technologies	Students were motivated to share data with family and friends as a sense of 'accomplishment' and sharing with peers was motivated by a sense of altruism. Tracking and sharing data with HCPs changed their experience of healthcare visits and improved communication and decision making.
Lafky and Horan (40)	USA	The aim of the study was to better understand the design implications for EHRs for people living with chronic conditions.	Various conditions	$n = 28$ adults	Electronic health record	Individuals are less concerned about the security of health data (compared with financial data). People living with disabilities are less willing to take measures to secure their health information.
Leventhal et al. (41)	USA	The aim of the study was to assess patient preferences for accessing PGData through a digital system, CareWeb.	Various conditions	$n = 105$ adults	HCPs	More than half of all participants wanted to share all of their data with HCPs. Only 5 participants out of 105 did not want anyone to view their data in the EHR.
Maiorana et al. (25)	USA	The aim of the study was to examine how trust (in tech, people and processes) influences the acceptability of data sharing in an HIV related context.	HIV	$n = 549$ adults	HCPs and other stakeholders *via* Health Information Technology (HIT)	People living with HIV are widely accepting of HIT. Increased experience and comfort with digital technology, confidence in security protocols, trust in providers and institutions who use the technology enhance understanding of the benefits to patients.
Murnane et al. (34)	USA	The aim of this study was to better understand how people living with Bipolar Disorder use data in condition management and how this may be facilitated by the use of personal informatics systems.	Mental health (Bipolar Disorder; BD)	$n = 22$ adults	Multiple recipients *via* self-tracking technologies	People with BD believe that sharing data with HCPs is standard and supports doctor-patient communication. Sharing with family and friends is important for recognizing when patients with BD may need intervention and support.
Nurgalieva et al. (42)	Sweden	This study explored patient perspectives on what technical, ethical, security, and privacy challenges need to be considered when designing platforms for sharing medical information.	Various conditions and a subgroup of cancer patients	Survey $n = 2,587$ adults Interviews of cancer patients $n = 15$ adults	A national online platform for accessing personal electronic health information and sharing with multiple recipients.	Few patients chose to share health information through an online platform despite a majority of patients trusting the security of the system. Cancer patients and psychiatric patients were notably hesitant to share online. Different conditions might cause a range of feelings in patients regarding sharing their health information, such as concerns about stigma.

(Continued)

EXTRACTING INSIGHTS FROM DIGITAL PUBLIC HEALTH DATA USING ARTIFICIAL INTELLIGENCE

Topic Editors:
Yu-Dong Zhang, University of Leicester, United Kingdom
Hong Lin, University of Houston–Downtown, United States
João Manuel R. S. Tavares, University of Porto, Portugal
Steven Fernandes, Creighton University, United States

Citation: Zhang, Y.-D., Lin, H., Tavares, J. M. R. S., Fernandes, S., eds. (2022). Extracting Insights From Digital Public Health Data Using Artificial Intelligence. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-873-2

TABLE 2 | Continued

References	Country of origin	Aim/purpose	Long-term health condition (LTHC)	Population	Sharing data with/platform	Key findings
Zhang et al. (46)	China	This study looked at the sharing of personal health information in online health communities for people living with multiple conditions.	Various conditions	$n = 337$ adults	Sharing with peers *via* online health communities	Health information privacy concerns, together with informational and emotional support, significantly influence personal health information disclosure intention. Privacy concerns are negatively influenced by two coping appraisals (i.e., response efficacy and self-efficacy) and positively affected by two threat appraisals (i.e., perceived vulnerability and perceived severity).
Zhu et al. (47)	USA	This study looked at the use of patient-generated data using digital technology in a clinician-patient consultation.	Various conditions	$n = 12$ adult patients $n = 9$ clinicians interviews	Self-tracking technologies and sharing data with HCPs	Patients are motivated to collect and share PGData to foster a better understanding of their health and improve clinician appointments. Clinicians largely ignored data brought to consultations in this study. Some clinicians and patients feel overwhelmed by raw data.

HCPs encouraging a more patient-centered approach although such artifacts also have the potential to disrupt the doctor-patient relationship (50). The few papers that focus more on the sharing of self-tracking data with clinicians (33, 47) contrast the perceived benefits experienced by patients with the more negative or skeptical feelings toward the data expressed by HCPs.

Trust as a barrier to sharing is discussed less often outside of the context of sharing with HCPs. A notable exception is Warner et al. (44). In discussing the importance of mutual self-disclosures in the development of trust, Warner et al. (44) note that the features of mobile apps do not always support trust in their users. Uncertainties over the disclosure of patient-provided health information (i.e., HIV status in the mobile app dating environment, whereby people do not disclose, or report their last sexual health check as a long time ago) can cause distrust of other people living with HIV. A further study which addresses the role of trust outside of a clinical context is provided by Bussone et al. (18). This study explores the concerns of those living with HIV when sharing personal health information with their peers and finds that trust in digital sharing platforms can be enhanced when it is associated with a recognized HIV charity or trusted medical organization. This study also describes how strong privacy and security measures are vital for building trust in such peer-sharing platforms.

Identity

The literature discusses digital identity in terms of concerns regarding identifiers relevant to one's personal data and online presence. The conscious management of digital identity online has an impact on patient willingness to share PGData with online social networking sites such as Facebook (16, 30, 31, 43, 44). People living with diabetes, mental health or HIV expressed a desire to withhold PGData relating to their condition from their wider social network (16, 31, 44):

> "*Many participants reflected on the undesirability of contributing any health-related content to Facebook, since this platform was seen primarily as a space for the conscious construction of a positive identity. As such, the inclusion of references to diabetes or mental health could jeopardise this.*" (16)

This is further supported by Bussone et al. (18) who explore attitudes toward sharing among those living with HIV and find that participants report a strong desire to self-manage certain aspects of their digital identities by sharing individual attributes of identity if anonymised:

> "*They indicated willingness to share digital identity attributes, including gender, age, medical history, health and well-being data, but not details that could reveal their personal identity.*" (18)

An alternative strategy for managing digital identity is discussed by O'Kane et al. (30) who describe how some people living with either type 1 or type 2 diabetes are happy to share their PGData under pseudonyms in specific health related online forums provided they get the support they need in return:

> "*The use of social media seems to be a fine balance between openly sharing sensitive medical information whilst also remaining in control of what is considered private. If you want to talk about the worst thing that you've done to your diabetes, or you are really ignoring it, or you're in a dark place, you can share that information without sharing your name, without alerting your employer to your potential issue or alerting your family even. You can keep those feeling private but share them publicly in a way gets the support without putting you out there like you're waving a flag saying 'I'm diabetic and I want everyone to look at me!' right? – Patient 14*" (30)

The management of digital identity is closely linked to how well patients manage their condition, even when seeking out support. When the perceived management of the condition is considered poor, some patients are less likely to share their data. Among adolescents with type 1 diabetes, Vaala et al. (31) report, "*Those who consider posting health-related information online face a tension between pursuing health-related goals, such as obtaining advice or emotional support, and maintaining a favorable impression as someone who is healthy and competent it seems the balance may shift in favor of the latter among adolescents who are struggling with glycemic control.*" Other studies investigating the sharing behaviors of people living with diabetes (type unspecified) with public health researchers have discovered that patients with better self-reported measures of glycaemic control are more likely to share their data (45).

Warner et al. (44) report on the reflection of HIV disclosure and identity management as some study participants note how they perceive the sharing of a person's negative HIV status and last test date as a way to show off to other users on Grindr, where one participant states, "*I just don't like it. It's like giving yourself a pat on the back for being lucky or" "better" "than other people".*

In terms of sharing PGData with online social networking sites, identity and privacy are key issues. People living with LTHCs want to be able to withhold PGData relating to the condition from their wider social network and to exert control over what data they share and with whom. For people with LTHCs these needs reflect changing patterns of engagement with social networking sites and online support groups (51, 52). Sharing PGData may occur in a temporary or intermittent manner, depending on the nature of condition and the type of PGData shared, which often varies in relation to the stage of the illness or health condition (53). Many people with LTHCs are less likely to share PGData when they are perceived to be managing their condition poorly (45) and blaming and shaming can often be a core experience for people with diabetes on online forums (54).

"Digital personhood" (a term used to discuss recognition of a human being as having status as a person in the electronic realm) can be impacted by illness, resulting in pre-and post-illness personas (55). Managing our identities across different contexts is often difficult when engaging in social interaction online, a term recognized as "context collapse" (56). People with LTHCs may have to work harder at their online communication, making more conscious decisions about what PGData to share and what to withhold, in order to shape or maintain their preferred digital identity or presentation of self (57). Separating out more generic social networking sites such as Facebook from specific, often anonymous, online health support groups is one strategy. Newman et al. (58) show how people with LTHCs manage their PGData sharing between online health communities and Facebook; Facebook is used to present a positive identity of self-control, whilst an online forum, by contrast, affords a space to be more open about expressing personal difficulties.

Privacy and Security

Privacy and security issues refer to concerns raised by patients surrounding the preservation of individual privacy and the ability to provide secure storage of personal data and information. Privacy concerns are discussed as a barrier to the sharing of PGData in the majority of articles. Agaku et al. (36) report that privacy and security concerns are the main reason why some patients withhold their PGData from healthcare professionals. In addition, the authors report concerns about the security of information whilst being "electronically transferred" or 'faxed', as well as 'the perception that a patient had very little say in how their PGData was used' are all associated with significantly higher odds of withholding personal information from a healthcare professional (36). Similarly, Caine and Hanania (38) report that patients express having less choice over what is shared with third-party organizations, e.g., health insurance companies. The request by patients for granular control over sharing of PGData and medical information is common across many articles (29, 30, 36–38, 44) and informed consent is requested to enable the patient to make decisions about who to share their data with (36, 38). Bernaerdt et al. (37) find that this desire for granular control in certain patient groups is often present despite a lack of awareness of the value or meaning of medical data to third parties. This evidence suggests that patients need to be better informed of the consequences and implications of sharing personal health information with third parties.

Torabi and Beznosov (43) note that privacy risk perceptions of people living with LTHCs are context dependent. Many authors also highlight the perceived sensitivity of PGData to the patient, and that how a person feels about their physical and mental health at the time of sharing impacts privacy risk perception (30, 32, 38, 45, 46). One particular study looking at multiple conditions and sharing PGData from Electronic Health Records (EMR) reports,

> "*There was not one potential recipient (e.g., primary care physician) with whom all patients wanted to share all of the information in their EMR with unconditionally. This was the case for both groups of participants: those with highly-sensitive health information in their EMR (21 participants) and those without highly-sensitive information (nine participants).*" (38)

However, some patients expect healthcare professionals to have complete access, despite the sensitivity of data, "*they need to know everything that is going on in your health*" (30).

Hartmann et al. (32) describe how patients may wish to minimize the potential risk of data being used against them by third-party organizations:

> "*Individuals want to keep control of such sensitive data and just do not want to share it with everybody or more precisely with third-party agents from whom negative consequences could arise from, such as German public health insurance, for instance. People are worried about being tracked at places that indicate risk behavior or self-damaging behavior, which could result in financial consequences (e.g., higher insurance rates or loss of treatment reimbursement).*" (32)

Concerns over sharing PGData with HCPs typically focus on the potential for the data to be shared more widely with third-party organizations, and the review indicates that patients are keen to be able to control or limit this wider sharing to protect the privacy of their data.

On social media use for diabetes support, O'Kane et al. (30) report patients' changing perspectives on privacy, where social media use is a delicate balance of sharing openly sensitive medical information whilst also having control over what is considered private, based on how vulnerable they feel:

> "*People may choose to view previously held privacy beliefs as overly cautious and want to reveal more about their previous medical history, but they still have their own individual levels of comfort. Although Patient 13 would write his diabetes blog under his own name and picture as mentioned above, one group interview participant did not feel comfortable with this level of privacy. I think it would be alright to share information about how your, maybe how your blood sugars go… […]but I don't think it is necessary to say your name and your address or anything like that. You can have a blog where everyone has a username or something. And then I think it's really helpful. I don't think you really need to identify yourself. – Group Interview Participant*" (30)

However, sometimes the interest in maintaining dignity and privacy (on any digital platform) can outweigh the interest in health and subsequently results in patients withholding PGData (30).

Privacy and security concerns are shown to be significantly influenced by particular demographics (e.g., age and education level), and characteristics (e.g., self-efficacy) (46), as well as the trajectory of a person's illness and "*other temporally-situated outside influences*" (30). Furthermore, differences between LTHCs may influence the extent to which privacy concerns influence sharing preferences and behaviors. For example, Esmaeilzadeh et al. (39) describe how differences between mental and physical conditions result in differences in sharing propensities:

> "*Individuals with a physical illness favor higher levels of structure mainly due to information quality dimensions (i.e., better understandability, accessibility, and usefulness). However, individuals with mental disorders prefer highly structured interfaces due to lower psychological risks and privacy concerns.*" (39)

Nurgalieva et al. (42) also highlight how different conditions may elicit a range of privacy concerns. They show how cancer patients and psychiatric patients were notably hesitant to share *via* a national digital platform for the sharing of personal health information. This may be explained by certain conditions being more likely to provoke fears surrounding potential stigma or causing family members to worry (42). Further understanding of the influence of both demographic and health condition factors is required so that healthcare organizations may adequately structure their patient platforms to accommodate the differing privacy concerns of patient groups, for example by providing information to patients about how data is going to be used and stored.

Anticipated security breaches present a barrier to the sharing of PGData (30, 36), whilst in contrast, having confidence that digital technology has safeguards in place is a facilitator to sharing of PGData (36). Patients' concerns are justified by factors including their previous experiences of digital technology and security breaches occurring both electronically and using paper health records (30).

Privacy concerns affect sharing PGData in online settings. People with LTHCs have to make judgements about the type and amount of information they share with others, weighing up the contextual integrity of their personal data sharing against potential privacy and security posed by the "silent listeners" on the network, i.e., third-party applications or advertisements (59). Site ownership and funding plays into this directly with peer-sharing resources now being hosted by large pharmaceutical companies, charities, healthcare organizations and individuals. Some data-driven sites such as PatientsLikeMe have been built to support information exchange between patients (11) but their relationship with third-party organizations can cause some users to feel uncomfortable (60). Recent changes to the ownership of such sites may increase concern in this context; for example the acquisition of PatientsLikeMe by the healthcare and insurance company UnitedHealth Group caused some users to express privacy and security concerns regarding their personal data (61).

In comparison to sharing with HCPs or sharing *via* social media, there are relatively few papers that focus on sharing PGData within a third-party context. The papers that do examine this context identify privacy and security as key issues (30, 32, 37) and highlight that some patients may have little understanding of the value of PGData to third parties (37). However, clearly more work is needed to understand whether the TIPS barriers and facilitators play a role within this setting. The key messages in this setting are that people want to be able control the privacy of their data and to have the option of changing their consent preferences with regard to sharing. Patients are also more likely to share with organizations that have the potential to impact their health directly and less likely with organizations further from this premise (i.e., researchers, government or health insurance companies). Although the papers examine patients' attitudes toward sharing PGData with third-party organizations, they do not explore differences in sharing behaviors depending on whether or not PGData is anonymised.

RQ2: WHAT IS THE IMPACT OF STIGMA ON THE SHARING OF PGDATA *VIA* DIGITAL TECHNOLOGY BY THOSE LIVING WITH LTHCS?

Stigma can be both internal (felt stigma or self-stigmatization) or enacted (external or discrimination) experiencing unfair treatment from others (62). Anticipated stigma presents a barrier to the sharing of PGData, across multiple platforms and with various recipients (18, 30, 31, 36, 44). A range of health conditions are associated with significant stigma (63), such as living with HIV (18, 64), mental health problems (65, 66), and chronic pain (67). People living with LTHCs are at risk of losing out on the benefits of sharing data when affected by stigma and are more likely to withhold information. Both internal and enacted stigma impact the way in which patients develop trust with the recipients of PGData.

Internal and enacted stigma can create a barrier to sharing PGData, particularly for people living with HIV. When exploring the use of Grindr to disclose HIV status, Warner et al., (44) report how people living with HIV are sometimes keen to withhold this information due to concerns of social exclusion and loss of sexual opportunity. Although in contrast, the article also describes some comments from Grindr users about how stigma can be used as a motivator for disclosure for some men living with HIV as a way to "reduce their stigma exposure". However, Warner notes,

> "*Stigma around HIV could lead some users to purposefully misreport their HIV status to avoid exposure to stigma. This is reflected in our findings, where users report their desire for HIV disclosure choice. In an environment where all users are expected to disclose, privacy unravelling around non-disclosures may limit this choice. When all said and done, it's forced disclosure that I dislike, or the fact that HIV+ users are expected to self-disclose their status straight away. Why should they? (Paraphrased comment from NW8).*" (44)

The majority of findings relating to stigma are of people living with HIV (18, 25, 26, 44). However, in other conditions, authors note how participants express their concerns over their PGData being used against them by healthcare providers and third-party organizations:

> "*…A woman with a previous psychiatric diagnosis believed her history had been misused by ambulance personnel who "put my name in the computer" and diverted her to psychiatric care instead of the medical emergency care she was seeking. Another individual was concerned about how doctors interpreted the history of sexually transmitted infection in his medical record. One woman was strongly motivated to conceal her diabetes from her insurer because she was concerned the company would raise her premiums.*" (17)

> "*Individuals want to keep control of such sensitive data and just do not want to share it with everybody or more precisely with third-party agents from whom negative consequences could arise from, such as German public health insurance, for instance. People are worried about being tracked at places that indicate risk behavior or self-damaging behavior, which could result in financial consequences (eg, higher insurance rates or loss of treatment reimbursement).*" (32)

Among adolescents with type 1 diabetes, an increase in restrictive sharing settings through social media are considered a factor of anticipated stigma when adolescents have higher than normal blood glucose levels (31, 45). Insights into the sharing preferences of previously explored groups, such as those living with HIV and diabetes, may help to guide the further study of the role that stigma plays in the formation of attitudes and sharing behaviors in those living with other LTHCs.

DISCUSSION

Summary of Findings

Trust, Identity, Privacy and Security (TIPS) concerns can present a barrier to sharing health and lifestyle data when using digital technology to share data in multiple contexts. A quarter of the articles discussed the role of trust in sharing PGData. Privacy as a barrier to sharing was present across most articles and across most settings. Other TIPS concerns were more readily identified as barriers to sharing in certain contexts. Identity management was seen as a barrier to sharing more frequently within the context of social networking sites and the issue of security was a barrier to the sharing of PGData with third parties. The presence of enacted stigma acted as a barrier to sharing PGData across all settings although this was most noticeable in relation to HIV compared to other LTHCs.

The narrative review has shown that TIPS issues are a considerable barrier to the sharing of PGData across all settings. The presence of specific TIPS issues varied by context, such that in certain settings particular barriers were more prominent. However, the literature shows that the majority of research looking at the sharing of PGData has focused on clinical settings with relatively few studies examining attitudes toward sharing with third parties such as public health and research. In clinical settings the key TIPS issue was trust. Distrust in the recipient of the information was highlighted as a key barrier to sharing PGData *via* digital technology.

In social network sharing online, we found that identity and privacy concerns were expressed in relation to the self-management of health and concerns regarding oversharing. These issues were key barriers to sharing but there was a lack of more detailed and nuanced information about the kind of PGData individuals were or were not sharing with respect to these concerns. Whilst the focus of this review paper was on the barriers and facilitators of sharing PGData more broadly rather than types of data *per se*, it was interesting to note that the studies covered a range of PGData. In clinical settings, unsurprisingly the focus was on electronic health records and clinical data, whereas in the social networking settings, the range of PGData was more varied and included more subjective data around mood, sleep and emotions. Despite focussing on stigmatized health conditions, there was relatively little focus on the role stigma played in decisions regarding sharing PGData *via* digital technology. References to stigma were most prevalent in relation to HIV but far less mentioned with respect to other conditions. Understanding the roles of both internal and enacted stigma regarding the sharing of PGData needs further attention. Much of the discussion surrounding stigma related to the unwanted disclosure of sensitive information. Despite a lack of consensus about what should be considered sensitive information, previous literature suggests five categories of sensitive health data: sexually transmitted infections, HIV/AIDS status, sexual health and pregnancy, mental health information, and substance use (41). However, legal definitions of what constitutes sensitive personal data are often very broad in scope; for example, the European Commission categorizes "health-related data" as sensitive personal data (68). Further research may seek to examine how perceptions of information sensitivity among those with various LTHCs affect patient privacy concerns and explore how these concerns may vary across different conditions.

Whilst we have assumed that sharing is a beneficial activity, it is also worth considering that, as part of supporting the management of PGData, we need to think about how people make sense of their data. We cannot always expect people to be able to successfully interpret their data (34), and collecting and monitoring data can be overwhelming for some people leading to negative health consequences (69). Patients may express varying preferences for managing PGData and have different technological abilities relevant to the skills required to actively record, monitor and manage personal health information. Understanding these patient differences may help to avoid burdening people with the "invisible work" of managing personal health information (17, 70). Managing PGData can also add to the increasing demands faced by HCPs due to the time required to analyse and make sense of the data that patients provide. As well as understanding the role of health literacy in relation to managing PGData (17), and the burden placed on both patients and HCPs, we need to know more about the motivations for both collecting and sharing PGData in different contexts to see if TIPS issues vary accordingly. Understanding more about the types of PGData people with LTHCs are happy to share and how the TIPS barriers might differentially apply to these forms of data would be a useful next step. Finally, there is a need for more qualitative studies in this area, especially in relation to TIPS barriers and facilitators to sharing PGData with third-party organizations as the majority of these studies are based on quantitative data.

Whilst our review highlights some of the key TIPS concerns that people living with LTHCs have with respect to sharing their PGData, none of the studies evaluated solutions or interventions to overcome these barriers. A few papers discussed participants' suggestions or desires concerning greater transparency and control over the information. Clearer informed consent to improve the transparency of the sharing process would increase the granular control for participants (30). A growing body of literature, that is beyond the scope of this narrative review, continues to explore technology and policy-based solutions to resolve general concerns about health data to facilitate secure and privacy-preserving sharing (71–73). However, given the specific TIPS concerns that this narrative review highlights with respect to the sharing of PGData by those living with LTHCs, future research may look to investigate how successful those solutions proposed to tackle general concerns about health data are at alleviating the TIPS concerns of those living with LTHCs. Furthermore, though recent research examining dynamic consent models for the sharing of clinical data (blood and tissue samples) in third-party contexts showed promising results in terms of acceptability (74), it remains to be seen how such models would work across more stigmatized health conditions and across more varied PGData types. Although there is still little empirical work in this area, the UK EPSRC funded programme INTUIT is examining TIPS concerns around PGData sharing primarily for people living with HIV but also for those with other stigmatized conditions. The INTUIT project aims to identify TIPS concerns and to design tools that remove the barriers to collecting and sharing PGData in order to improve the health and well-being of stigmatized populations. As part of this project, we are conducting interviews with people living with LTHCs to examine the role of sharing context and health condition in relation to TIPS barriers. This is the first study of its kind to focus specifically on TIPS issues in relation to sharing PGData *via* digital technology across a variety of stigmatized LTHCs and across a range of different sharing contexts.

CONCLUSION

This narrative review has provided a broader perspective on the TIPS challenges faced by people managing LTHCs and has shown that TIPS issues are a considerable barrier to the sharing of PGData *via* technology by those living with LTHCs across all settings (i.e., sharing with clinical staff, public health surveillance, researchers, peer communities, friends, social networks and other third-party organizations). Distrust in the proposed recipient of PGData, the need to manage one's digital identity and broadly held privacy and security concerns present barriers to sharing in a clinical setting but more research is needed to understand other contexts, particularly sharing with third parties. The presence of internal and enacted stigma has also been shown to impede the sharing of PGData across all settings, although most research in this area has centered on those living with HIV. This highlights the need for further research to consider differences between conditions in experiences of stigma, and to consider how these differences interact with the influence that TIPS concerns have over sharing. Whilst the technological sharing of PGData holds great potential benefits for the health, well-being and social outcomes of people managing LTHCs, the TIPS challenges faced by those individuals must be better understood and addressed if interactions with care services, peer support networks, and private organizations are to be optimized.

DATA AVAILABILITY STATEMENT

This paper provides a narrative review of existing literature. All data underlying this review are cited in the references.

AUTHOR CONTRIBUTIONS

The initial concept for the project was founded by LC, ES, and AD. The research questions and search criteria were then developed by LC, ES, and EmS and reviewed by all authors. EmS conducted the initial search and review and RB conducted the final search and review. LC and ES supported EmS in the shortlisting of papers against the criteria. EmS, RB, ES, LC, KL, JG, ST, and AD participated in discussing, revising and editing the manuscript.

FUNDING

This work was funded by the EPSRC, grant number EP/R 033900/1.

ACKNOWLEDGMENTS

This study has been conducted as part of a UK EPSRC funded programme (INTUIT: Interaction Design for Trusted Sharing of Personal Health Data to Live Well with HIV, 2020; EP/R033900/2), examining TIPS concerns around the sharing of self-generated health and lifestyle data primarily among people living with HIV but also for those with other potentially stigmatized conditions.

REFERENCES

1. Rau R, Soroko E, Jasilionis D, Vaupel JW. Continued reductions in mortality at advanced ages. *Popul Dev Rev.* (2008) 34:747–68. doi: 10.1111/j.1728-4457.2008.00249.x
2. Vaupel JW. Biodemography of human ageing. *Nature.* (2010) 464:536–42. doi: 10.1038/nature08984
3. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-care technologies in HCI: trends, tensions, and opportunities. *ACM Trans Comput Hum Interact (TOCHI).* (2015) 22:1–45. doi: 10.1145/2803173
4. Sands DZ, Wald JS. Transforming health care delivery through consumer engagement, health data transparency, and patient-generated health information. *Yearb Med Inform.* (2014) 9:170–6. doi: 10.15265/IY-2014-0017
5. Cohen DJ, Keller SR, Hayes GR, Dorr DA, Ash JS, Sittig DF. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project healthdesign. *JMIR Hum Factors.* (2016) 3:e5919. doi: 10.2196/humanfactors.5919
6. Bussey L, Sillence E, editors. (How) do people negotiate online information into their decision making with healthcare professionals? In: *Proceedings of the 2017 International Conference on Digital Health*. London (2017). doi: 10.1145/3079452.3079495
7. Rodgers S, Chen Q. Internet community group participation: Psychosocial benefits for women with breast cancer. *J Comput Mediated Commun.* (2005) 10:JCMC1047. doi: 10.1111/j.1083-6101.2005.tb00268.x
8. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. *J Biomed Inform*. (2015) 56:265–72. doi: 10.1016/j.jbi.2015.06.003
9. Abernethy AP, Herndon JE, Wheeler JL, Day JM, Hood L, Patwardhan M, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. *J Pain Symptom Manage.* (2009) 37:1027–38. doi: 10.1016/j.jpainsymman.2008.07.011
10. Smith T, Dunn M, Levin K, Tsakraklides S, Mitchell S, van de Poll-Franse L, et al. Cancer survivor perspectives on sharing patient-generated health data with central cancer registries. *Qual Life Res.* (2019) 28:2957–67. doi: 10.1007/s11136-019-02263-0
11. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. *J Med Internet Res.* (2010) 12:e19. doi: 10.2196/jmir.1549
12. Hulsen T. Sharing is caring—data sharing initiatives in healthcare. *Int J Environ Res Public Health.* (2020) 17:3046. doi: 10.3390/ijerph17093046
13. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. *Health Aff.* (2014) 33:1115–22. doi: 10.1377/hlthaff.2014.0147
14. Sillence E, Briggs P. Managing your health online: issues in the selection, curation, and sharing of digital health information. In: *Oxford Handbook of Cyberpsychology.* (2019). Oxford: Oxford University Press.
15. Li I, Dey AK, Forlizzi J, editors. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In: *Proceedings of the 13th international conference on Ubiquitous computing*. Beijing (2011). doi: 10.1145/2030112.2030166
16. Fergie G, Hunt K, Hilton S. Social science & medicine social media as a space for support: young adults' perspectives on producing and consuming user-generated content about diabetes and mental health. *Soc Sci Med.* (2016) 170:46–54. doi: 10.1016/j.socscimed.2016.10.006
17. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. *J Med Internet Res.* (2015) 17:e137. doi: 10.2196/jmir.4381
18. Bussone A, Kasadha B, Stumpf S, Durrant AC, Tariq S, Gibbs J, et al. Trust, identity, privacy, and security considerations for designing a peer data sharing platform between people living with HIV. *Proc ACM Hum-Comput Interact.* (2020) 4:173. doi: 10.1145/3415244
19. *INTUIT: Interaction Design for Trusted Sharing of Personal Health Data to Live Well with HIV* (2018). Available online at: https://intuitproject.org/ (accessed December 01, 2020).
20. Earnshaw VA, Quinn DM, Park CL. Anticipated stigma and quality of life among people living with chronic illnesses. *Chronic Illn.* (2011) 8:79–88. doi: 10.1177/1742395311429393
21. Sheehan L, Corrigan P. Stigma of Disease and Its Impact on Health. In: *The Wiley Encyclopedia of Health Psychology*. (2020). p. 57–65. doi: 10.1002/9781119057840.ch139
22. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev*. (2019) 4:5. doi: 10.1186/s41073-019-0064-8
23. Noble H, Smith J. Reviewing the literature: choosing a review design. *Evid-Based Nurs.* (2018) 21:39–41. doi: 10.1136/eb-2018-102895
24. Ring N, Jepson R, Ritchie K. Methods of synthesizing qualitative research studies for health technology assessment. *Int J Technol Assess Health Care.* (2011) 27:384–90. doi: 10.1017/S0266462311000389
25. Maiorana A, Steward WT, Koester KA, Pearson C, Shade SB, Chakravarty D, et al. Trust, confidentiality, and the acceptability of sharing HIV-related patient data: lessons learned from a mixed methods study about Health Information Exchanges. *Implementation Sci.* (2012) 7:1–14. doi: 10.1186/1748-5908-7-34
26. Teixeira PA, Gordon P, Camhi E, Bakken S, HIV. patients' willingness to share personal health information electronically. *Patient Educ Couns.* (2011) 84:e9–12. doi: 10.1016/j.pec.2010.07.013
27. Hood JE, Friedman AL. Unveiling the hidden epidemic: a review of stigma associated with sexually transmissible infections. *Sex Health.* (2011) 8:159–70. doi: 10.1071/SH10070
28. Lee AS, Cody SL. The stigma of sexually transmitted infections. *Sex Transmit Infect Issue Nurs Clin, E-Book.* (2020) 55:295. doi: 10.1016/j.cnur.2020.05.002
29. Fuji KT, Abbott AA, Galt KA, A qualitative study of how patients with type 2 diabetes use an electronic stand-alone personal health record. *Telemed e-Health.* (2015) 21:296–300. doi: 10.1089/tmj.2014.0084
30. O'Kane AA, Mentis HM, Thereska E, editors. Non-static nature of patient consent: shifting privacy perspectives in health information sharing. In: *Proceedings of the 2013 conference on Computer supported cooperative work*. San Antonio, TX (2013). doi: 10.1145/2441776.2441838
31. Vaala SE, Lee JM, Hood KK, Mulvaney SA. Sharing and helping: predictors of adolescents' willingness to share diabetes personal health information with peers. *J Am Med Inf Assoc.* (2018) 25:135–41. doi: 10.1093/jamia/ocx051
32. Hartmann R, Sander C, Lorenz N, Böttger D, Hegerl U. Utilization of patient-generated data collected through mobile devices: insights from a survey on attitudes toward mobile self-monitoring and self-management apps for depression. *JMIR Mental Health.* (2019) 6:e11671. doi: 10.2196/11671

33. Kelley C, Lee B, Wilcox L, editors. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver, CO (2017). doi: 10.1145/3025453.3025750
34. Murnane EL, Walker TG, Tench B, Voida S, Snyder J. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. In: *Proceedings of the ACM on Human-Computer Interaction*. (2018) 2:1–27. doi: 10.1145/3274396
35. The Kings Fund. *Long-Term Conditions and Multi-Morbidity*. Kings Fund (2015).
36. Agaku IT, Adisa AO, Ayo-Yusuf OA, Connolly GN. Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from healthcare providers. *J Am Med Inf Assoc.* (2014) 21:374–8. doi: 10.1136/amiajnl-2013-002079
37. Bernaerdt J, Moerenhout T, Devisch I. Vulnerable patients' attitudes towards sharing medical data and granular control in patient portal systems: an interview study. *J Eval Clin Pract.* (2021) 27:429–37. doi: 10.1111/jep.13465
38. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. *J Am Med Inf Assoc.* (2012) 20:7–15. doi: 10.1136/amiajnl-2012-001023
39. Esmaeilzadeh P, Mirzaei T, Dharanikota S. The impact of data entry structures on perceptions of individuals with chronic mental disorders and physical diseases towards health information sharing. *Int J Med Inform.* (2020) 141:104157. doi: 10.1016/j.ijmedinf.2020.104157
40. Lafky DB, Horan TA. Personal health records: consumer attitudes toward privacy and security of their personal health information. *Health Inf J.* (2011) 17:63–71. doi: 10.1177/1460458211399403
41. Leventhal JC, Cummins JA, Schwartz PH, Martin DK, Tierney WM. Designing a system for patients controlling providers' access to their electronic health records: organizational and technical challenges. *J Gen Intern Med.* (2015) 30:17–24. doi: 10.1007/s11606-014-3055-y
42. Nurgalieva L, Cajander Å, Moll J, Åhlfeldt R-M, Huvila I, Marchese M. 'I do not share it with others. No, it's for me, it's my care': On sharing of patient accessible electronic health records. *Health Inform J.* (2020) 26:2554–67. doi: 10.1177/1460458220912559
43. Torabi S, Beznosov K, editors. Privacy aspects of health related information sharing in online social networks. In: *2013 {USENIX} Workshop on Health Information Technologies (HealthTech 13).* Washington, DC (2013).
44. Warner M, Gutmann A, Sasse MA, Blandford A. Privacy unraveling around explicit HIV status disclosure fields in the online geosocial hookup app Grindr. *Proc ACM Hum Comp Interact*. (2018) 2:1–22. doi: 10.1145/3274450
45. Weitzman ER, Adida B, Kelemen S, Mandl KD. Sharing data for public health research by members of an international online diabetes social network. *PLoS ONE.* (2011) 6:e19256. doi: 10.1371/journal.pone.0019256
46. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. *Inform Manage.* (2018) 55:482–93. doi: 10.1016/j.im.2017.11.003
47. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: an interview study. *AMIA Ann Symp proc AMIA Symp.* (2017) 2016:1303–12.
48. Birkhäuer J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, et al. Trust in the health care professional and health outcome: a meta-analysis. *PLoS ONE.* (2017) 12:e0170988. doi: 10.1371/journal.pone.0170988
49. Ciechanowski P, Katon WJ. The interpersonal experience of health care through the eyes of patients with diabetes. *Soc Sci Med.* (2006) 63:3067–79. doi: 10.1016/j.socscimed.2006.08.002
50. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why do they need to check me?" patient participation through eHealth and the doctor-patient relationship: qualitative study. *J Med Internet Res.* (2018) 20:e11. doi: 10.2196/jmir.8444
51. Baumer EP, Khovanskaya V, Adams P, Pollak JP, Voida S, Gay G. Designing for engaging experiences in mobile social-health support systems. *IEEE Pervasive Comput.* (2013) 12:32–9. doi: 10.1109/MPRV.2013.47
52. Massimi M, Bender JL, Witteman HO, Ahmed OH, editors. Life transitions and online health communities: reflecting on adoption, use, and disengagement. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. Baltimore, MD (2014). doi: 10.1145/2531602.2531622
53. Eschler J, Dehlawi Z, Pratt W, editors. Self-characterized illness phase and information needs of participants in an online cancer forum. In: *Proceedings of the International AAAI Conference on Web and Social Media* (2015).
54. Basinger ED, Farris M, Delaney AL. Investigating the experience of diabetes stigma in online forums. *South Commun J.* (2020) 85:43–57. doi: 10.1080/1041794X.2019.1655662
55. Kerrigan F, Hart A. Theorising digital personhood: a dramaturgical approach. *J Mark Manage.* (2016) 32:1701–21. doi: 10.1080/0267257X.2016.1260630
56. Boyd DM. Taken out of context: American teen sociality in networked publics. Berkeley : University of California (2008).
57. Van Der Velden M, El Emam K. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. *J Am Med Inform Assoc.* (2013) 20:16–24. doi: 10.1136/amiajnl-2012-000949
58. Newman MW, Lauterbach D, Munson SA, Resnick P, Morris ME, editors. It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. New York, NY (2011). doi: 10.1145/1958824.1958876
59. Stutzman FD, Gross R, Acquisti A. Silent listeners: the evolution of privacy and disclosure on Facebook. *J Priv Confidentiality.* (2013) 4:2. doi: 10.29012/jpc.v4i2.620
60. Angwin J, Stecklow S. Scrapers' dig deep for data on Web. *Wall Str J.* (2010) 12. Available online at: https://www.wsj.com/articles/SB10001424052748703358504575544381288117888 (accessed January 12, 2020).
61. Truong K. PatientsLikeMe acquired by UnitedHealth Group. *MedCity News.* (2019).
62. Gray AJ. Stigma in psychiatry. *J R Soc Med.* (2002) 95:72–6. doi: 10.1177/014107680209500205
63. Corrigan PW. *The Stigma of Disease and Disability: Understanding Causes and Overcoming Injustices*. American Psychological Association (2014). doi: 10.1037/14297-000
64. Turan B, Budhwani H, Fazeli PL, Browning WR, Raper JL, Mugavero MJ, et al. How does stigma affect people living with HIV? The mediating roles of internalized and anticipated HIV stigma in the effects of perceived community stigma on health and psychosocial outcomes. *AIDS Behav.* (2017) 21:283–91. doi: 10.1007/s10461-016-1451-5
65. Bharadwaj P, Pai MM, Suziedelyte A. Mental health stigma. *Econ Lett.* (2017) 159:57–60. doi: 10.1016/j.econlet.2017.06.028
66. Corrigan PW, Rao D. On the self-stigma of mental illness: stages, disclosure, and strategies for change. *Can J Psychiatry*. (2012) 57:464–9. doi: 10.1177/070674371205700804
67. De Ruddere L, Craig KD. Understanding stigma and chronic pain: a-state-of-the-art review. *Pain.* (2016) 157:1607–10. doi: 10.1097/j.pain.0000000000000512
68. European Commission. *What Personal Data is Considered Sensitive?* Available from: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en
69. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. *Social Theory Health.* (2013) 11:256–70. doi: 10.1057/sth.2013.10
70. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information quality challenges of patient-generated data in clinical practice. *Front Publ Health.* (2017) 5:284. doi: 10.3389/fpubh.2017.00284
71. Liang X, Barua M, Lu R, Lin X, Shen XS. HealthShare: achieving secure and privacy-preserving health information sharing through health social networks. *Comput Commun.* (2012) 35:1910–20. doi: 10.1016/j.comcom.2012.01.009

72. Jiang S, Zhu X, Wang L, EPPS Efficient and privacy-preserving personal health information sharing in mobile healthcare social networks. *Sensors.* (2015) 15:22419–38. doi: 10.3390/s150922419
73. Jin H, Luo Y, Li P, Mathew J, A review of secure and privacy-preserving medical data sharing. *IEEE Access.* (2019) 7:61656–69. doi: 10.1109/ACCESS.2019.2916503
74. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. *J Med Internet Res.* (2016) 18:e5011. doi: 10.2196/jmir.5011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fpubh.2021.770111

Evaluating Rumor Debunking Effectiveness During the COVID-19 Pandemic Crisis: Utilizing User Stance in Comments on Sina Weibo

*Xin Wang, Fan Chao and Guang Yu**

School of Management, Harbin Institute of Technology, Harbin, China

Background: The spread of rumors related to COVID-19 on social media has posed substantial challenges to public health governance, and thus exposing rumors and curbing their spread quickly and effectively has become an urgent task. This study aimed to assist in formulating effective strategies to debunk rumors and curb their spread on social media.

Methods: A total of 2,053 original postings and 100,348 comments that replied to the postings of five false rumors related to COVID-19 (dated from January 20, 2020, to June 28, 2020) belonging to three categories, authoritative, social, and political, on Sina Weibo in China were randomly selected. To study the effectiveness of different debunking methods, a new annotation scheme was proposed that divides debunking methods into six categories: denial, further fact-checking, refutation, person response, organization response, and combination methods. Text classifiers using deep learning methods were built to automatically identify four user stances in comments that replied to debunking postings: supporting, denying, querying, and commenting stances. Then, based on stance responses, a debunking effectiveness index (*DEI*) was developed to measure the effectiveness of different debunking methods.

Results: The refutation method with cited evidence has the best debunking effect, whether used alone or in combination with other debunking methods. For the social category of *Car* rumor and political category of *Russia* rumor, using the refutation method alone can achieve the optimal debunking effect. For authoritative rumors, a combination method has the optimal debunking effect, but the most effective combination method requires avoiding the use of a combination of a debunking method where the person or organization defamed by the authoritative rumor responds personally and the refutation method.

Conclusion: The findings provide relevant insights into ways to debunk rumors effectively, support crisis management of false information, and take necessary actions in response to rumors amid public health emergencies.

Keywords: COVID-19, rumor, false information, stance detection, debunking, effectiveness, social media

OPEN ACCESS

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Alessandro Rovetta,
Redeev SRL, Italy
Junxiang Chen,
University of Pittsburgh, United States
Donghui Yang,
Southeast University, China

****Correspondence:***
Guang Yu
yug@hit.edu.cn

Specialty section:
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: *03 September 2021*
Accepted: *05 November 2021*
Published: *30 November 2021*

Citation:
Wang X, Chao F and Yu G (2021)
Evaluating Rumor Debunking
Effectiveness During the COVID-19
Pandemic Crisis: Utilizing User Stance
in Comments on Sina Weibo.
Front. Public Health 9:770111.
doi: 10.3389/fpubh.2021.770111

INTRODUCTION

COVID-19 spread globally, severely threatening the lives and health of people, and significantly affecting the economy, education, and daily life in various countries (1–3). The United Nations assessed that the COVID-19 pandemic has wiped out decades of developmental gains (4). One unique aspect of the COVID-19 pandemic has been the over-abundance of information on COVID-19-related topics in media and the internet, where true and false information intertwine; the result has been a disruption of the social order. The World Health Organization (WHO) has flagged this situation as an infodemic (information + epidemic) (5–7), noting that an infodemic is the excessive amount of both accurate and inaccurate health information that can spread misinformation, disinformation, misinformation, and rumors during a health emergency, which can hamper an effective public health response (4–8). The role of social and traditional media (e.g., news, TV programs, newspapers, and other mass media) in spreading disinformation and misinformation in the COVID-19 infodemic has now been recognized in a substantial number of literature reports (9, 10). In addition, online social media platforms can easily "facilitate rapid information sharing and large-scale information cascades," further exacerbating the issue (11, 12). As Gallotti et al. noted on the verge of the global pandemic emergency, human communication on social media is largely characterized by the production of informational noise and misleading or false information (13).

A paradigm-shifting feature of social media is that any user can produce, access, and disseminate content (14) thereby altering the way people communicate, share, receive, use, and search for both general and health-specific information (15); without any doubt, this has the potential to foster the rapid exchange and diffusion of false information (16). However, studies also demonstrate that social media may be very useful for fighting false information during a public health crisis (15, 17) as these social media platforms can be effective tools for combating and mitigating rumors during public health crises (18). Notably, cognitive psychology research suggests that debunking false information is not an easy task (19). Even clearly corrected false information can continue to influence many users who then intentionally or unintentionally ignore the truth and spread false rumors. Despite considerable efforts to debunk rumors, the desired effect has not been achieved (20). Consequently, debunking false information on social media quickly and effectively and curbing its spread has become an urgent task.

Before commencing this research, brief remarks on the terminology are necessary. Regarding the definition and connotation of fake news, misinformation and disinformation, conspiracy theories, satire, and rumors, scholars have different understandings, which are often used interchangeably in academic research (21, 22). In this study, owing to the fact that it includes within it broader related concepts, we have retained the term "rumor" for our research objectives (23). We usually understand rumors as "unverified and instrumentally relevant information statements in circulation that arise in the context of ambiguity, danger or potential threat, and that function to help people make sense and manage risk" (24). The term is applied to a piece of information whose veracity at the time of dissemination is unclear, and thus it may spread false information in the absence of verifiable information (25, 26). In this study, we examined rumors, as they do not imply a judgment about the sender's intention or the veracity of the presented information (23).

Given the seemingly increasing trend of rumor proliferation, an increasing number of researchers have pursued effective debunking and remedial methods to alleviate the negative influence of rumors (27). A common strategy is to use certain fact-checking services to debunk rumors by providing authoritative statements that discern the truth from falsehood (27, 28). These fact-checking platforms include FactCheck.org (http://fackcheck.org), TruthOrFiction.com (http://truthorfiction.com), and Sina Community Management Centre (https://service.account.weibo.com), etc. These online resources can present the truth to the public and play a key role in distinguishing between true and false information (29). However, most fact-checking processes require considerable human labor and material resources, that is, they are time-consuming and expensive (30, 31). Another common strategy is based on psychological theories to explore the psychological mechanisms of rumor acceptance and how corrections and debunking messages attempt to cope with them. For example, Lewandowsky et al. analyzed the cognitive factors that often render misinformation resistant to correction and recommended ways in which corrections should be designed, structured, and applied to maximize the impact of the corrections and reduce the spread of false information (32). In recent years, researchers have focused more on this topic in the context of digital media (33). For instance, several studies indicated that online rumor spread can be reduced by logic- and humor-based corrections (34), links citing truth-related evidence (35), and corrections from reputable sources such as official government agencies (36) and knowledgeable users (37).

However, compared with unverified reports, debunking messages are not as widely shared and spread by users on social media (23, 38). Therefore, some researchers have begun to study improving the effectiveness of rumor debunking on social media, focusing on studying the effectiveness of debunking methods. Only a few research papers qualitatively studied the effectiveness of rumor debunking methods (39, 40) Some analyzed the effectiveness of different rumor-control strategies or specific rumor-control cases (40); for example, Kimmel et al. investigated the efficacy of rumor-control marketing tactics (41); Paek et al. investigated the most effective rumor response strategies to control food safety risks (42, 43); Yang et al. focused on the effectiveness analysis of mixed rumor-quelling strategies by modeling the rumor-truth competing process (44). Other studies aimed to identify key influencing factors to organize debunking methods; for example, Li et al. explored the relationship between social media rumor refutation effectiveness and its possible affecting factors and provided practical suggestions to help accelerate the rumor refutation process (40).

In addition, some studies have shown that effective management of rumors depends not only on the choice of debunking methods but also on the evaluation and response of people to different ways of debunking (42). Thus, some researchers have also begun to study how people respond after rumors have been debunked on social media (29), such as emotional reactions (45), changes in attitudes, and perceived credibility (27). Decades of well-researched experiments, meta-analyses, questionnaires, and surveys offer guidelines for rumor correction and debunking, for example, Pal et al. used an online survey to reveal that denials could be crafted to effectively debunk rumors by incorporating salient beliefs (46), Walter et al. used a meta-analysis to evaluate the relative impact of social media interventions designed to correct health-related misinformation (47). However, these efforts require control over the manipulation of debunking methods (e.g., randomly assigning people to different methods or controlling for the results of many alternative explanations), resulting in higher human, financial, and time costs for debunking rumors on social media, and thus, do little to improve the real-time effectiveness of rumor debunking. Furthermore, the advent of big data has brought a large, constant, and rapidly growing amount of data generated by numerous sources; however, the large amount of interaction data generated by social media has not been extensively studied (40). To better utilize the rich, objective, user-generated research data on social media platforms, this study considers COVID-19 rumors that are widely spread on Sina Weibo, a popular social media in China, as a research sample. Moreover, this study investigates the classification of debunking methods; and based on stance responses to debunking postings in user comments, this study also establishes a method to measure the effectiveness of different debunking methods. This study seeks to reveal the implementation effects of different debunking methods in order to optimize debunking methods and to provide an objective basis for improving their effectiveness.

Thus, to resolve the objectives, we developed a new annotation scheme that manually categorized six debunking methods (i.e., denial, further fact-checking, refutation, person response, organization response, and combination methods) to classify the debunking methods used for the posts on Sina Weibo. Then, deep learning algorithms were used to build text classification models to automatically detect the stance responses (i.e., supporting, denying, querying, and commenting stances) to debunking postings in user comments. In addition, based on the results of identifying user stances in comments, we developed an index to measure the effectiveness of rumor debunking. Finally, an empirical study was conducted using five rumors related to COVID-19 as case studies to compare the effectiveness of different debunking methods, and countermeasures for effective monitoring and combating rumors in public health crises were proposed based on the results of the study.

METHODS

Study Context and Data Collection

Sina Weibo (http://www.weibo.com), often referred to as China's Twitter (http://www.twitter.com), is one of the most influential social network platforms in China (48). In contrast to other social networks such as Facebook, Instagram, and WeChat, communication on Sina Weibo is almost entirely public. As a popular social media platform with a large user base in China, everyone on social media can communicate, share, receive, use, and search for both general and health-specific information (15). Sina Weibo has not only become one of the primary platforms for releasing and propagating various false information amid the public health crisis (48), but also is a particularly effective tool for combating rumors (18). Therefore, quantitative research on the role of debunking posts on Sina Weibo is relevant to enhancing social media rumor debunking effectiveness (40).

In this study, collaborating with the Zhiwei Data Sharing Platform (http://university.zhiweidata.com, hereinafter referred to as Zhiwei Data), we collected data on COVID-19 rumors through the Business Application Programming Interface (API) of Sina Weibo, as follows:

(a) Collecting original postings (i.e., non-reposted postings, hereinafter referred to as postings) related to COVID-19 rumors. First, we defined keyword combinations related to rumors using multiple logical relationships (e.g., AND and OR). Then we collected rumor data related to the COVID-19 pandemic dating from January 20, 2020, to June 28, 2020 (these rumors had been confirmed as "false" by authoritative statements). Second, to avoid data selection bias, we randomly selected five sensational rumors from three categories: authority, society, and politics. These five rumors are the five fake news stories with the highest number of comments from January 20, 2020, to June 28, 2020, which have the characteristics of the most widespread dissemination and influence on social media platforms in China during the COVID-19 pandemic. The annotation scheme of three categories was developed through an iterative process of rounds of annotation and evaluation with three researchers (two Ph.D. students and one expert from Zhiwei Data who are experienced in rumor research). Similarly, after manual data denoising and cleaning by three annotators, we randomly selected 2,053 postings related to the following five rumors:

1. Yansong Bai dialogued with Nanshan Zhong on "News 1 + 1" (*News*). On January 26, 2020, it was rumored on the internet that CCTV news channel "News 1 + 1" broadcasted a special program on COVID-19 at 21:30, in which host Yansong Bai (a renowned anchor) invited Nanshan Zhong (an authoritative respiratory pathologist) to introduce the effective treatments to prevent the COVID-19 pandemic. The truth was that January 26, 2020, was a Sunday and there was no news show "News 1 + 1."
2. Materials of Jiangsu province aided medical team of Hubei province were detained (*Jiangsu*). On February 9, 2020, some netizens released false rumors that a "medical team from Jiangsu province arrived at Wuhan (the capital of Hubei Province) airport, its supplies were looted, luggage was lost, and local doctors and nurses were transported in trucks," which sparked a heated debate on social media. On February 11, @Jiangsu Province Internet

TABLE 1 | Detailed description of rumors about COVID-19.

Category of rumor	Coding scheme for category of rumor	Postings	Comments
Authority	Those slandered by the rumors are authoritative individuals or organizations in expertise fields related COVID-19, such as Nanshan Zhong (*News* rumor), the Wuhan Institute of Virology (*PatientZero* rumor), respectively.		
News rumor		478	11,207
PatientZero rumor		676	49,763
Society	The rumors about social events, social problems, and social style involving people's daily life about COVID-19, especially reflecting social morality and ethics.		
Jiangsu rumor		114	8,006
Car rumor		527	15,848
Politics	The rumors related to COVID-19 about politics, i.e., the activities of classes, parties, social groups, and individuals in domestic and international relations.		
Russia rumor		258	15,524

The data collection of postings and comments ended on September 12, 2020, 23:59:59.

Reporting Centre's official microblogging website clarified that the news was not true.

3. A postgraduate from the Wuhan Institute of Virology was the "Patient Zero" (*PatientZero*). On February 15, 2020, a false rumor was circulated on social media: Yanling Huang (a postgraduate) from Wuhan Institute of Virology (an authoritative research institute affiliated with the Chinese Academy of Sciences) was "Patient Zero" (the first one to contract the disease and thus the one who started spreading the virus) of COVID-19. The truth was that Yanling Huang graduated from the institute in 2015 with a master's degree, and had been working and living in other provinces since graduation, and has not returned to Wuhan, has not been infected by the COVID-19, and was in good health.
4. Eighty Chinese citizens were abused while being quarantined in Russian Federation (*Russia*). On March 1, 2020, the Chinese Embassy in Russia informed that a rumor claimed that Moscow (the capital of the Russian Federation) police violently enforced the law, mistreated isolated personnel, and took away Chinese citizens for no reason. It was later revealed that the news was not true.
5. A car owner in Hubei province died of COVID-19 (*Car*). On June 22, 2020, a rumor was spread that a car owner, who had volunteered to help people by delivering them vegetables, had died from COVID-19. The truth was that the car owner was still alive, and the rumor monger was detained for three days.

(b) Collecting all comments that replied to 2,053 postings. From the postings obtained as mentioned earlier, we collected comment conversations associated with each posting. To collect comments, we scraped the web page of each posting to retrieve the URL. However, if the URL was missing or invalid, we chose MIDs (a unique ID of each posting on Sina Weibo) to retrieve the comments; consequently, we collected 100,348 comments (either direct replies or nested replies to the postings) that replied to the 2,053 postings.

As the focus of the study is on debunking rumors, all the selected rumors were false (hereinafter false rumor referred to as rumor) and debunked while the COVID-19 pandemic was still spread. A detailed description of rumors about COVID-19 is shown in **Table 1**, including the category of rumors, the coding scheme for each category of rumor, and the number of postings and comments of each rumor. In addition, because we used publicly available data, we only referred to the summarized results and did not derive any sensitive information.

Annotation Scheme for Debunking Methods

Filtering of Debunking Postings

We used manual annotation to filter out postings that debunked rumors. A three-person expert panel of social media researchers (two Ph.D. students and one expert from Zhiwei Data, who are experienced in research on rumor amid public health crisis) labeled the debunking postings out of the 2,053 postings. Referring to Tian et al. (49), the specific steps are as follows:

First, according to the supporting, denying, querying, or commenting (SDQC) stance (annotation rules, see section Stance Classification in Comments That Replied to Debunking Postings), two annotators with a detailed understanding of rumors independently labeled all 2,053 postings.

Second, to eliminate the differences due to human factors, the two members discussed all the annotation results and re-annotated the postings to reach an agreement on the differences.

Third, the third annotator annotated 2,053 postings to calculate inter-rater reliability. Cohen's kappa (κ) for the annotators was 0.921 (95% confidence interval (CI) $[0.856, 0.986]$, $p < 0.001$), indicating a good agreement among them (50, 51).

Finally, from the 2,053 postings, we obtained 1,721 postings of labeled debunking rumors.

Categorization of Debunking Methods

The decision as to what debunking methods should be defined, we mainly considered the psychological mechanisms of rumor acceptance and how corrections attempt to cope with them, such as rumor content is sufficiently important (i.e., personally involving or relevant) to people (52), the need for individuals and organizations to combat rumors, and the effect difference between "refutation" and "denial" in debunking rumors (53). In addition, to categorize the debunking methods as broadly as possible, we also referred to some categorization standards described by previous studies:

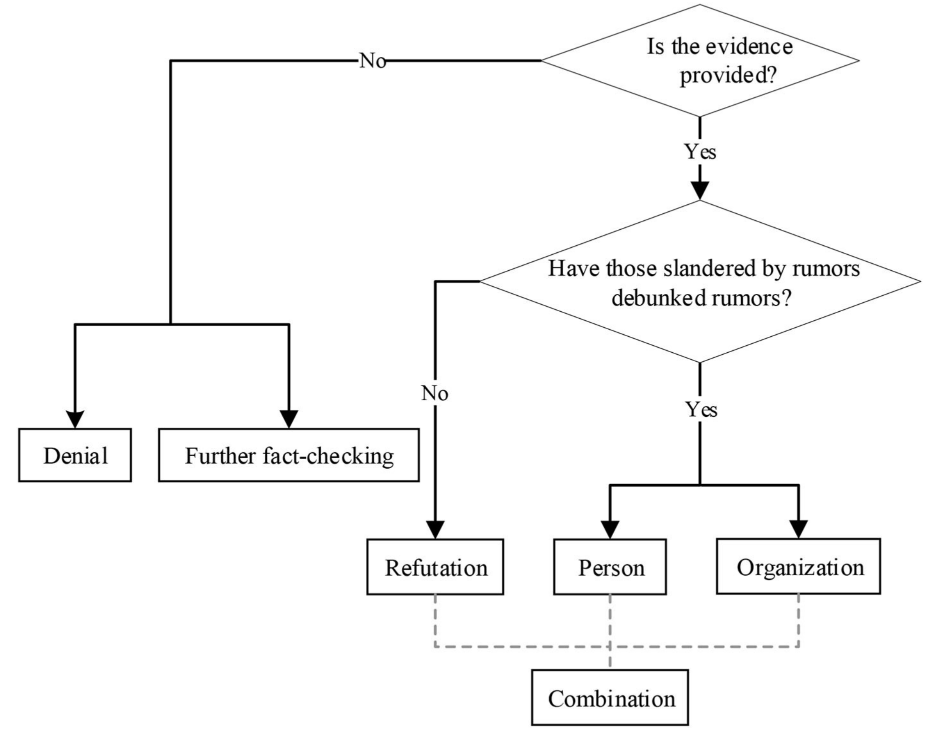


FIGURE 1 | Classification flowchart of debunking methods.

(a) Methods of providing evidence to debunk rumors (38, 54): first-hand experience, URL providing the evidence, quotation by a person or organization, image attachment, quotation by an unverifiable source, reasoning, and without evidence.
(b) Rumor-control strategies: refutation, denial, attacking the attacker (42).
(c) Debunking response strategies: tweet deletion, rumor clarification with a new tweet, and neither deletion nor clarification (29).
(d) Coding scheme for rumor-related messages: rumor messages, debunking messages, uncertainty about rumors, uncertainty about debunking messages, and others (23).

As shown in **Figure 1**, we designed the classification annotation scheme for the debunking methods as follows.

First, we classified the debunking methods into two main categories: debunking methods with uncited evidence and those with cited evidence. The following are two debunking methods with uncited evidence:

(a) **Denial (Den):** A flat counterstatement of a rumor without providing evidence. E.g., in the *Russia* rumor, "80 Chinese citizens were mistreated in isolation in Russia is ***not true***."
(b) **Further fact-checking (Fur):** Although people already know the veracity of debunking postings, they still have questions about some details and wish to obtain more relevant information or evidence. For example, in the *News rumor*, "Yansong Bai did not talk with Nanshan Zhong. However, I would want ***the relevant departments do disclose who the rumor monger is.***"

Second, depending on whether a rumor is debunked by rumors, debunking methods with cited evidence are classified into the following three categories:

(c) **Refutation (Ref):** Those not belonging to those slandered by the rumor provide evidence to debunk rumors. For example, in *the Jiangsu* rumor, the "***Public Security Bureau of Jiangsu Province*** clarified that the news was not true."

Considering those slandered by the rumor (e.g., persons and organizations slandered by the rumor to debunk rumors personally), another two debunking methods are subcategorized as follows:

(d) **Person response (Per):** Individuals slandered by rumors refute rumors. For example, in the *New* rumor, "Refute the rumor! ***Yansong Bai*** exclusive reply: tonight I did not dialog Nanshan Zhong."
(e) **Organization response (Org):** Organizations slandered by rumors refute rumors. For example, in *Patient Zero* rumor, "***Wuhan Institute of Virology***: can guarantee that Wuhan Institute of Virology currently has zero infections."

Third, we defined a combination of debunking methods as follows:

(f) **Combination (Com):** None of the combination methods contain any "debunking methods with uncited evidence"

TABLE 2 | Coding scheme for four stances (Translated into English from Chinese).

Stance	Description	Example
Supporting (S)	Users who commented to debunking postings believe that a rumor is true, i.e., they think debunking postings is false.	In the *News* rumor, "I'm specifically setting my alarm clock to wait to read about the COVID-19 on "News 1 + 1." Take precautions, wish for peace, and defeat this epidemic."
Denying (D)	Users who commented to debunking postings believe that a rumor is false, i.e., they think debunking postings is true.	In the *Russia* rumor, "80 Chinese citizens were mistreated in isolation in Russia is not true."
Querying (Q)	Users who commented to debunking postings while ask for additional evidence in relation to the veracity of a rumor.	In the *PatientZero* rumor, "I'm a little suspicious, can you continue to investigate? The mentors have come out to clarify, can Huang Yanling come out to speak?"
Commenting (C)	Users who commented to debunking postings not express their clearly stance whether they wanted to assess the veracity of a rumor.	In the *PatientZero* rumor, "What exactly does Patient Zero mean? Patients are counted from "1," and there is something wrong with counting from 0 itself, right? I thought in the world of mathematics, 0 is not exact."

(i.e., Den, and Fur). For example, in the *Car* rumor, "***Hubei Provincial Police*** debunked the rumor; at the same time, ***the car owner's company*** solemnly declared: the car owner is healthy, and the rumors on the online social media are not true."

Finally, we defined six debunking methods: Den, Fur, Ref, Per, Org, and Com.

Notably, the annotation scheme of these debunking methods was achieved through an iterative annotation process. Similarly, we also asked the three annotators to label the 1,721 debunking postings using the aforementioned annotation scheme. We also assessed the validity of the annotation scheme using Cohen's kappa ($\kappa = 0.747,\ 95\%\ CI\ [0.676,\ 0.818],\ p\ < 0.001$), and the results indicated that the annotation process had been validated.

Stance Classification in Comments That Replied to Debunking Postings

Classification Scheme for Comments

With the emergence of rumors on social media, people often express different stances and participate in extensive discussions (29), probably providing more opinions and evidence for rumor detection and debunking. Therefore, mining comments and revealing their SDQC [supporting (S), denying (D), querying (Q), or commenting (C)] stances on a false rumor help determine the veracity of rumors (55, 56), thereby realizing early detection and intervention of rumor spread. SDQC stance classification is an effective approach used by researchers to judge opinions and stances (57). Therefore, utilizing the rich information in comments that replied to the postings on Sina Weibo, we employed SDQC stance classification (38) to classify the stance in comments on a debunking posting into one of the four stances. As shown in **Table 2**, the comments of each user were categorized using the following four stances (56).

Stance Classification

To improve the classification efficiency and obtain satisfactory results in the case of massive text data, we designed text classifiers with supervised learning methods to automatically identify and classify the four types of SDQC stances in 100,348 comments.

Similarly, first, we asked two members of our team to label 12,000 comments (randomly selected from 100,348 comments) with the corresponding SDQC stances.

Second, the third member labeled 12,000 comments to validate the inter-annotator agreement, and the annotation process was assessed and validated using Cohen's kappa ($\kappa = 0.876,\ 95\%\ CI\ [0.856,\ 0.896],\ p\ <\ 0.001$).

Third, for the 100,000-level dataset in this research, based on an empirical rule, therefore we had to use a greater percentage of data to develop and test models and used the old way of splitting data to divide the dataset scientifically, that is, the non-redundant 12,000 comments were randomly divided into the independent training set and testing set according to the ratio of 7:3 (58–60). The label distribution is presented in **Table 3**.

Fourth, we trained and compared the text classifiers using deep learning methods. Because the new language representation model, bidirectional encoder representations from transformers (BERT), developed by Google in 2018, is conceptually simple, empirically powerful, and obtains new state-of-the-art results on 11 natural language processing tasks without substantial task-specific architectural modifications, we chose BERT and its related improved two models to train the classifiers (61). These three models are BERT, RoBERTa-wwm-ext (robust optimized BERT approach-whole-word masking-extended data), and RBT3 (three-layer RoBERTa-wwm-ext) (62–64). The performance of the proposed classifiers was evaluated using the macro-F1 score (used in the multi-class stance classification) (65, 66). Due to the prevalent phenomenon of class imbalance in text classification, the most widely used performance measure for multi-class text classification is the F1 score which is defined as the harmonic mean of precision and recall (67). It is known to be more informative and more useful than classification accuracy etc., especially for multi-class imbalance problems (65, 68). In macro-F1, we used each stance j to compute that particular stance's precision P_j as well as recall R_j, and finally computed a simple average of the F1 scores over classes (equal weight to each class) to get macro-F1 (65, 68).

During the fine-tuning process, we compared 72 sets of hyperparameters based on the three models to obtain the best-performing stance classification model. **Table 4** presents the performances of these classifiers under different combinations

TABLE 3 | Label distribution of the training and testing sets.

	Stance			
	Supporting	**Denying**	**Querying**	**Commenting**
Training set	227	2,305	1,126	4,742
Testing set	98	956	472	2,074

TABLE 4 | Results of stance classification.

Model	**Hyperparameters**	**Accuracy**	**Macro-precision**	**Macro-recall**	**Macro-F1**
BERT	(70, 16, 2e−5, 3)	80.33%	67.71%	65.94%	66.51%
RBT3	(140, 32, 5e−5, 3)	80.17%	68.78%	63.08%	64.53%
RoBERTa-wwm-ext	(140, 16, 3e−5, 3)	80.89%	68.76%	67.88%	68.06%

Hyperparameters (x, y, z, w), where x: max_seq_length (70, 140), y: train_batch_size (16, 32), z: learning_rate (2e−5, 3e−5, 5e−5), and w: num_train_epochs (2, 3). The values in parentheses represent the hyperparameters in the fine-tuning process.

of hyperparameters. For brevity, we have only shown the hyperparameter results with the best performance for each model. Detailed results of the stance classification are described in **Supplementary Table 1**. Finally, we chose the RoBERTa-wwm-ext model with hyperparameters (max_seq_length = 140, train_batch_size = 16, learning_rate = 3e−5, num_train_epochs = 3) to predict the SDQC stance for large comment datasets, as it performed satisfactorily with an accuracy of 80.89% and a macro-F1 of 68.06%.

Evaluation of the Effectiveness of Debunking Methods

Users' commenting behavior represents their behavioral decisions after they are exposed to postings and process those postings (69). As we used users who commented on debunking postings, we are approximating what the users may have been exposed to regarding debunking postings. Thus, based on stance responses to debunking postings in user comments, this section details the method established to evaluate the effectiveness of different debunking methods.

First, we established a method to evaluate the denial of rumors by users, considering the following two aspects. On the one hand, we considered the stance gap between clear stances in comments, i.e., the stance gap between rumor denial and rumor support. Only a large stance gap between denying stance and supporting stance can effectively show the advantage of denial of rumors by users, i.e., denying stance beats supporting stance, which can exclude the illusion of high denial generated by a high proportion of denying stances and a high proportion of supporting stances as well. On the other hand, we considered clear stances (i.e., supporting and denying stances), which could show the impact of debunking postings, i.e., the validity and persuasiveness of debunking postings in determining truth or falsity, because not all social media users would deny or support rumors even if they had already been debunked by accurate information (29). Thus, referring to the equation by Zubiaga et al. for a simplified analysis of rumor support and denial, they omitted other stances in comments (i.e., querying and commenting stances), which do not contribute to resolving the veracity of a rumor (38). Based on the above analysis, we used the ratio of "the difference between the denying and supporting stances" to "the sum of the two stances, i.e., omitting other stances that do not contribute to resolving the veracity of a rumor from all stances" to calculate the degree of denial [*denial index (DI)*], which could normalize the denying and supporting stances and make *DI* comparable across rumors and events (29, 38). *DI* was calculated as follows:

$$DI_{ij} = \frac{\#deny_{ij} - \#support_{ij}}{\#deny_{ij} + \#support_{ij}} \tag{1}$$

where $\#deny_{ij}$ and $\#support_{ij}$ denote the number of comments denying and supporting a rumor, respectively, under postings that used debunking method j in rumor i. In this study, $i \in \{1, 2, 3, 4, 5\}$ refers to the *News* rumor, *Jiangsu* rumor, *PatientZero* rumor, *Russia* rumor, and *Car* rumor, respectively, and $j \in \left\{Den, Fur, Ref, Per, Org, Com\right\}$ refers to the debunking method.

Second, to evaluate the effectiveness of different debunking methods accurately, we analyzed the redundancy of users toward different debunking methods; that is, although the users have seen the debunking message, their comments do not contribute toward resolving the veracity of rumors. *The redundancy index (RI)* indicates the extent to which users participate in the process of debunking a rumor. It is defined as "the number of comments with the commenting stance" divided by "the total number of comments." Accordingly, we have the following:

$$RI_{ij} = \frac{\#comment_{ij}}{\#all\_stance_{ij}} \tag{2}$$

where $\#comment_{ij}$ and $\#all\_stance_{ij}$ denote under postings that used the debunking method j in rumor i, the number of comments in which users did not clearly indicate their stance on the veracity of a rumor and the number of all comments, respectively.

Third, we defined the *debunking effectiveness index (DEI)* to measure the effectiveness of debunking rumors.

To comprehensively develop a measurement for the effectiveness of debunking rumors, we considered two aspects. On the one hand, the *DEI* measures the impact of the debunking posting, including two aspects, one aspect is related to the overall number of comments, i.e., examining whether the debunking posting is more influential, such that it receives more comments; Another aspect is related to lower redundancy in comments, such that it receives fewer comments that do not contribute to resolving the veracity of a rumor. Thus, the impact of the debunking posting was measured by *RI*. On the other hand, the *DEI* measures the denial of debunking postings and, similarly, includes two aspects. One measures whether the denying stances get more comments than supporting stances; the other measures for the validity and persuasiveness of debunking postings in determining truth or falsity, i.e., a simplified analysis of rumor support and denial to normalize the denying and supporting stances. Thus, the denial of the debunking posting was measured by *DI*. Generally, in comments in which users respond to a debunking method, if the number of comments with the denying stance is sufficiently higher than that of comments with irrelevant information that does not contribute to the veracity of rumors, then using the debunking method, users can clearly express their denial of fake news without expressing too many redundant or irrelevant comments. Therefore, this way of debunking might be satisfactory in disseminating true information and debunking rumors. Conversely, we considered that *DEI* was directly proportional to *DI* and inversely proportional to *RI*. Given these two values, *DEI* was calculated as follows:

$$DEI_{ij} = \frac{DI_{ij}}{RI_{ij}} \tag{3}$$

where DI_{ij} and RI_{ij} denote the degree of user denial and the redundancy of comments, respectively, when debunking method j is used in rumor i.

Figure 2 shows the technology roadmap of the overall research process. The flowchart on the left side of the dotted line describes the overall research process; the tree diagram on the right side of the dotted line represents the data annotation related to the research process.

Statistical Analysis

Descriptive analysis of the basic information of the rumors was conducted using SPSS for Windows, version 25.0.0 (IBM Corporation). The Pearson chi-square (χ^2) tests were performed to compare the difference in posting distribution of each debunking method across five rumors, as well as the differences in SDQC stance distribution under different debunking methods across five rumors (70). The Kruskal–Wallis tests were applied to evaluate whether there were statistically significant differences in the distribution of *DIs*, *RI*s, and *DEIs* under different debunking methods for each rumor (71). The *post-hoc* test with Bonferroni correction was used for pairwise comparisons when the overall test was statistically significant (72). In addition, we adopted a more general view of the *p*-value as a statistical summary of the compatibility between the observed data and what we would predict or expect to see if we knew the entire statistical model (all the assumptions used to compute the *p*-value) were correct (73, 74). Thus, the *p*-value can be cautiously interpreted as a continuous measure of the compatibility between the data and the entire model used to compute it, ranging from 0 for complete incompatibility to 1 for perfect compatibility, and in this sense may be viewed as measuring the fit of the model to the data (74).

RESULTS

Distribution of Debunking Postings

Figure 3 shows the proportion of postings for each debunking method, indicating the frequency of use of different debunking methods. The result of the Pearson Chi-square test was $\chi^2\,(25) = 958.273$, $p < 0.001$, indicating that the result had a very high degree of statistical significance. Generally, as shown in **Figure 3A**, more than half of the methods were used to debunk rumors through a combination of methods (Com), with a percentage of 61.2%, considerably exceeding the respective proportion of other debunking methods, followed by the refutation method (Ref, 16.2%) and denial method (Den, 12.2%).

Additionally, for each rumor, we analyzed the proportion distribution of postings that used different debunking methods. In **Figure 3B**, we observed the following: the combination method was most commonly used in the *Russia* rumor (Com, 87.3%); the refutation method was most commonly used in *Car* rumor (Ref, 52.3%); the denial method was most commonly used in the *Jiangsu* rumor (Den, 44.7%). The debunking methods that employed organizational (Org) and personal (Per) responses were most used in *the PatientZero* rumor, with percentages of 17.0 and 7.5%, respectively. However, the further fact-checking method (Fur) was used less frequently in all rumors, with a maximum of 1.9% in *PatientZero* rumor.

Stance Classification in Comments That Replied to Debunking Postings

We also investigated the SDQC stance distribution in the comments of users who participated in different debunking methods. The results of our stance classifier are shown in **Figure 4**; they include the number and proportion of comments for each SDQC stance under various debunking methods. For five rumors, Pearson Chi-square tests indicated that the results had a very high degree of statistical significance in the SDQC stance distribution under different debunking methods [**Figure 4A**: $\chi^2\,(30) = 19733.495, p < 0.001$; **Figure 4B**: $\chi^2\,(15) = 299.396, p < 0.001$; **Figure 4C**: $\chi^2\,(9) = 80.294, p < 0.001$; **Figure 4D**: $\chi^2\,(18) = 1649.622, p < 0.001$; **Figure 4E**: $\chi^2\,(9) = 32.584, p < 0.001$; **Figure 4F**: $\chi^2\,(9) = 3335.866, p < 0.001$].

From **Figure 4A**, it is evident that overall, compared with the debunking postings, the supporting (S) stance accounted for a greater proportion of non-debunking methods, with a percentage of 10.4%. This result shows that postings that use different debunking methods can reduce the impact of rumors. However, the denying (D) stance accounted for a higher percentage for the refutation method (47.4%). Additionally, the querying (Q) stance accounted for a higher proportion for the organization debunking method (Org, 27.6%). The commenting (C) stance accounted for the highest proportion for each debunking

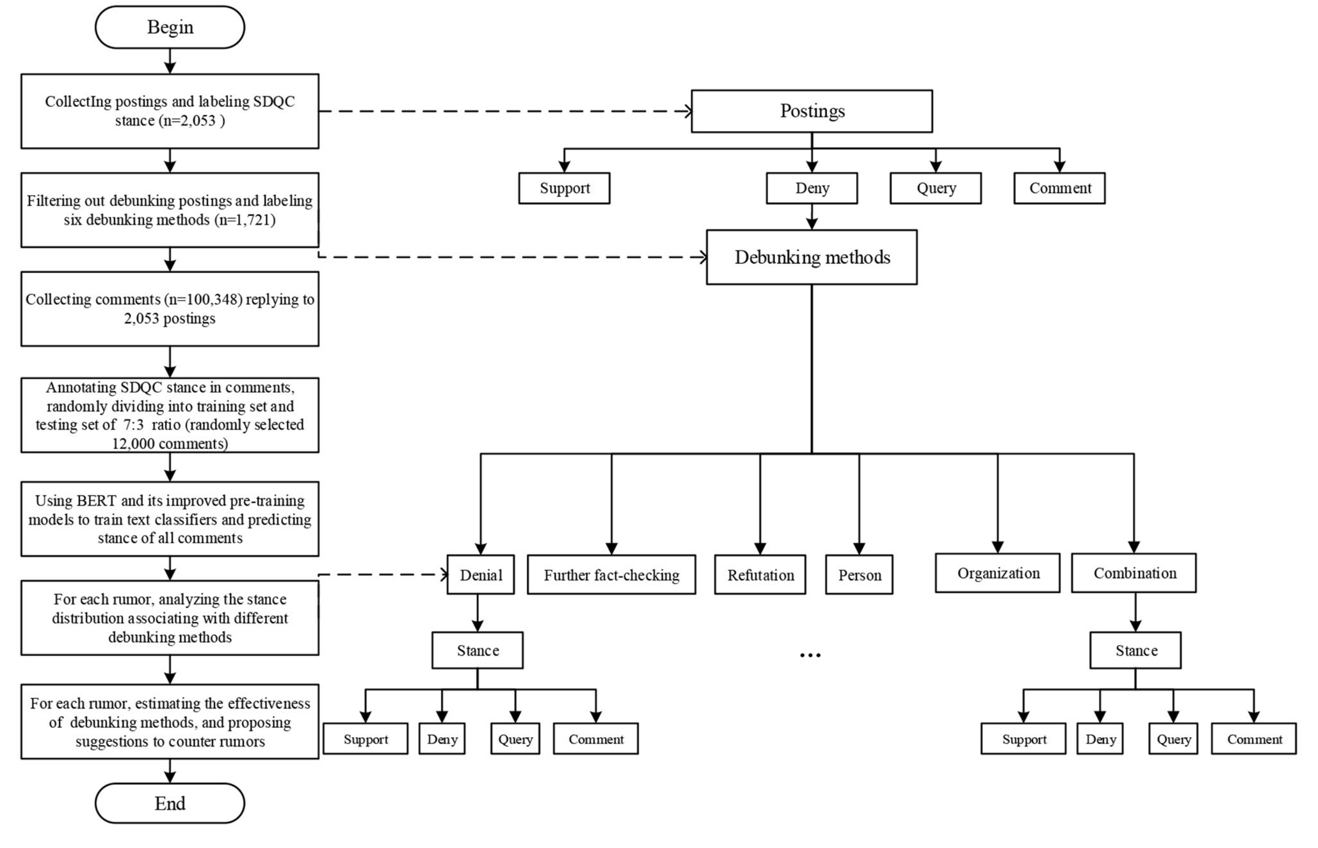


FIGURE 2 | Technology roadmap.

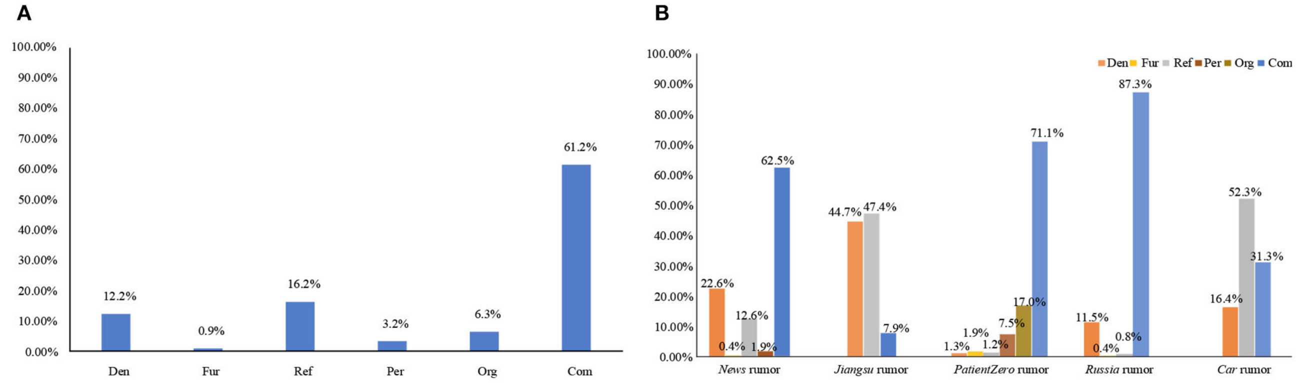


FIGURE 3 | Proportion of postings of the six debunking methods in the **(A)** overall and **(B)** each rumor. Den, Denial; Fur, Further fact-checking; Ref, Refutation; Per, Person response; Org, Organization response; Com, Combination.

method, with a maximum for the further fact-checking method (Fur, 77.9%), indicating that the number of comments that did not contribute to the veracity (i.e., true or false) of a rumor was high on Sina Weibo.

Similarly, as shown in **Figures 4B–F**, we analyzed the stance classification results for each rumor. As shown in **Figure 4F**, the supporting (S) stance accounted for the highest proportion of non-debunking methods in *Car* rumor (23.8%). Similarly, the denying (D) stance had the highest proportion for the refutation method in *Car* rumor (Ref, 58.7%). Simultaneously, for each rumor, the proportion of the querying (Q) stance was lower, although it accounted for a relatively higher proportion of the debunking method for the organizational response in *the PatientZero* rumor (Org, 27.6%) (see **Figure 4D**).

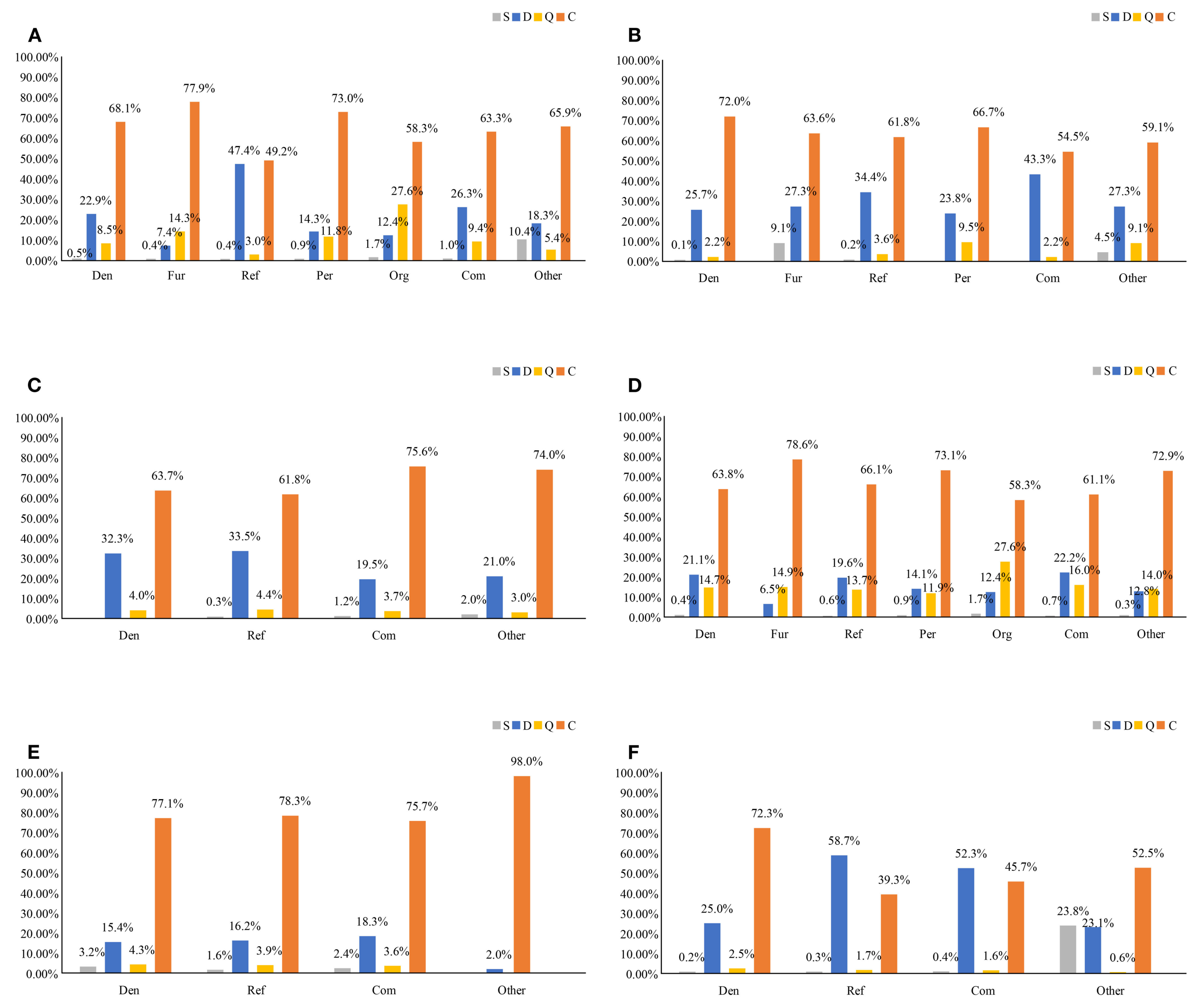


FIGURE 4 | Comparison of the number and proportion of comments for the SDQC stances corresponding to different debunking methods in **(A)** overall, **(B)** *News* rumor, **(C)** *Jiangsu* rumor, **(D)** *PatientZero* rumor, **(E)** *Russia* rumor, and **(F)** *Car* rumor. Other: all non-debunking methods with SDQC stances on rumors.

Finally, the commenting (C) stance had the highest proportion (98.0%) for non-debunking methods in the *Russia* rumor (see **Figure 4E**).

Estimation of the Effectiveness of Debunking Methods

Measurement of Debunking Effectiveness

Through the aforementioned analysis results, we can see the differences among various debunking methods, both overall and in each rumor. As shown in **Table 5**, to analyze the performance of social media users without considering the characteristics of rumors, we describe and analyze the overall performance of denial, redundancy, and debunking effectiveness index (*DI*, *RI*, and *DEI*, respectively). For different debunking methods, Kruskal–Wallis tests indicated that the results had a very high degree of statistical significance in the distribution of *DIs*, *RIs*, and *DEIs* (see **Table 5**). The results of *post-hoc* comparisons *via* Kruskal–Wallis test with Bonferroni correction were also shown in **Table 5**.

For *DEIs*, the *post-hoc* comparisons *via* Kruskal–Wallis test with Bonferroni correction showed that the results had a very high degree of statistical significance between combination method and other five types of debunking methods, and the results had a very high degree of statistical significance between refutation method and other five types of debunking methods (see **Table 5**). Thus, the refutation method has the best debunking effect ($DEI_{Ref} = 2.004$), followed by the combination method ($DEI_{Com} = 1.461$). Obviously, the debunking effects of these two methods considerably exceed those of the other debunking methods. Moreover, compared with using the refutation method alone, the combination method did not achieve a higher debunking effect.

TABLE 5 | Comparison of *DEI* under different debunking methods.

Debunking method[a]	DI[b]	RI	DEI	Pairs (I, J), Adj. Sig.[c, d]
Den	0.959	0.681	1.409	(Ref, Den), $p < 0.001$; (Ref, Fur), $p < 0.001$; (Ref, Per), $p < 0.001$; (Ref, Org), $p < 0.001$; (Com, Den), $p < 0.001$; (Com, Fur), $p < 0.001$; (Com, Ref), $p < 0.001$; (Com, Per), $p < 0.001$; (Com, Org), $p < 0.001$;
Fur	0.905	0.779	1.161	
Ref	0.986	0.492	2.004	
Per	0.881	0.730	1.207	
Org	0.762	0.583	1.307	
Com	0.925	0.633	1.461	
Mean 95% CI	0.901 [0.837, 0.964]	0.654 [0.605, 0.704]	1.430 [1.238, 1.622]	
Median	0.925	0.633	1.461	
Chi-Square (df)	851.081(5)	709.992(5)	782.305(5)	
Sig.	$p < 0.001$	$p < 0.001$	$p < 0.001$	
Kruskal-Wallis				
H	800.471	827.813	748.546	
Sig.	$p < 0.001$	$p < 0.001$	$p < 0.001$	

[a] *Den-Denial; Fur-Further fact-checking; Ref-Refutation; Per-Person response; Org-Organization response; Com-Combination.*

[b] *Most of our rumor data were collected retrospectively after the truth was revealed, which represents the eventual general trend in public opinion as truth-driven, hence explaining the calculated high value of DI in our results.*

[c] *The Adj. Sig. value was the adjusted p-value, which was employed with a Bonferroni-type adjustment of p-value.*

[d] *For brevity, we only listed the main post hoc testing results.*

TABLE 6 | Comparison of *DEI* among five rumors.

Method Rumor	News	Jiangsu	PatientZero	Russia	Car
Den	1.373	1.571	1.506	0.852	1.355
Fur	0.786	/	1.273	/	/
Ref	1.596	1.591	1.417	1.047	2.522
Per	1.500	/	1.198	/	/
Org	/	/	1.307	/	/
Com	1.831	1.167	1.535	1.019	2.153
Mean 95% CI	1.520 [1.181, 1.860]	1.443 [0.849, 2.037]	1.519 [1.240, 1.798]	0.973 [0.711, 1.235]	1.844 [1.209, 2.479]
Median	1.831	1.587	1.535	1.019	2.398
Chi-Square (df)	59.970 (4)	7.500 (2)	25.251 (5)	7.781 (2)	128.749(2)
Sig.	$p < 0.001$	$p = 0.024$	$p < 0.001$	$p = 0.020$	$p < 0.001$
Kruskal-Wallis					
H	177.256	4.160	228.989	48.002	141.033
Sig.	$p < 0.001$	$p = 0.125$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Pairs (I, J), Adj. Sig.[a,b]	(Ref, Den), $p = 0.033$; (Com, Den), $p < 0.001$; (Com, Fur), $p = 0.159$; (Com, Ref), $p < 0.001$; (Com, Org), $p = 0.037$;	/	(Com, Den), $p = 0.008$; (Com, Fur), $p < 0.001$; (Com, Ref), $p = 0.001$; (Com, Per), $p < 0.001$; (Com, Org), $p < 0.001$;	(Com, Den), $p < 0.001$; (Ref, Den), $p = 0.001$; (Ref, Com), $p = 0.283$;	(Den, Com), $p = 0.001$; (Ref, Den), $p < 0.001$; (Ref, Com), $p < 0.001$;

[a] *The Adj. Sig. value was the adjusted p-value, which was employed with a Bonferroni-type adjustment of p-value.*

[b] *For brevity, we only listed the main post-hoc testing results.*

Comparison of Debunking Effectiveness Among Rumors

Because we considered different types of rumors, the same debunking method may have different effects on each rumor. Thus, we used *DEI* to measure and compare the differences in debunking effects among different types of rumors. For each rumor, we also used the Kruskal-Wallis tests to determine whether there were statistically significant differences between

DEIs across different debunking methods (see **Table 6**). The *post-hoc* comparisons *via* Kruskal–Wallis test with Bonferroni correction showed that the results had a very high degree of statistical significance between combination method and other five types of debunking methods for *PatientZero* rumor; for *News* rumor, the results had a very high degree of statistical significance between combination method and other four types of debunking methods (excepting further fact-checking method); for *Russia* rumor, the results had a very high degree of statistical significance between denial method and other two methods (combination and refutation methods); and for *Car* rumor, the results had a very high degree of statistical significance between refutation method and other two types of debunking methods (see **Table 6**). Thus, in the *post-hoc* testing results with a very high degree of statistical significance in **Table 6**, for authority rumors, i.e., *News* rumor and *PatientZero* rumor, the combination method was the most effective (e.g., $DEI_{NewsCom} = 1.831$ and $DEI_{PatientZeroCom} = 1.535$). However, for the social category of *Car* rumor and political category of *Russia* rumor, the refutation method was the most effective (e.g., $DEI_{CarRef} = 2.522$ and $DEI_{RussiaRef} = 1.047$). Furthermore, for *Jiangsu* rumor in the social category, although the refutation method was the most effective ($DEI_{JiangsuRef} = 1.591$), the result had a low degree of statistical significance ($H = 4.160$, $p = 0.125$).

Analysis of Combination Method

The research results show that the combination method obtains the optimal or suboptimal debunking effect. Therefore, we further analyzed the specific distribution of the combination method. First, we considered decomposing the combination method into five corresponding single debunking methods. In **Figure 5**, the nodes represent different debunking methods, the edges represent a combination of the respective nodes, and the edge thickness represents the number of comments obtained using this combination method; the thicker the edge, the greater the number of comments obtained in this combination method, whereas the thinner the edge, the smaller the number of comments obtained. In particular, the weight of an edge represents the *DEI* of the combination method. In addition, for different combination methods, Kruskal–Wallis tests indicated that the results had a very high degree of statistical differences in the distribution of *DEIs* (see **Figure 5**).

For the selected rumors, the combination methods comprise different combinations of three single debunking methods: Ref (a), Per (b), and Org (c). We have used notations such as $Com < a, b >$ or $Com < a, b, c >$ to indicate the combination of different debunking methods. Additionally, as shown in **Figure 5**, the solid-line edge represents the two-tuple $Com < a, b >$, whereas the triple-tuple $Com < a, b, c >$ was formed by the nodes at both ends of the dotted-line edge and their common neighbor node. Meanwhile, the weights represented by each edge are independent, and the weight value on the solid-line edge represents the *DEI* of $Com < a, b >$, whereas that on the dotted-line edge represents the *DEI* of $Com < a, b, c >$.

First, as shown in **Figure 5A**, judging from the popularity of comment users, $Com < Ref, Org >$ as a debunking method received more comments, meaning that using this combination of debunking methods could induce widespread discussion among users. Simultaneously, all the combination methods were combinations of the refutation method and other debunking methods, and *DEIs* of different combinations were relatively different from one another ($H = 347.485$, $p < 0.001$). The *post-hoc* comparisons *via* Kruskal–Wallis test with Bonferroni correction showed that the results had a very high degree of statistical significance between $Com < Ref, Org >$ and other two types of debunking methods ($Com < Ref, Per, Org >$ and $Com < Ref, Per >$) (see **Figure 5**). For $Com < Ref, Org >$ ($DEI_{Com<Ref,Org>} = 1.406$) and $Com < Ref, Per, Org >$ ($DEI_{Com<Ref,Per,Org>} = 1.684$), with the incorporation of one debunking method, DEI_{Com} increased slightly. However, for $Com < Ref, Per >$ ($DEI_{Com<Ref,Per>} = 1.763$) and $Com < Ref, Org >$ ($DEI_{Com<Ref,Org>} = 1.406$), the *DEI* of the combination method varies with different combinations of debunking methods. Therefore, the research results show that incorporating another debunking method is not necessarily conducive to improving the debunking effect; however, the scenario needs to be analyzed depending on specific rumors.

Second, for each rumor, the distribution of each combination method is shown in **Figures 5B–F**. The *post-hoc* comparisons *via* Kruskal–Wallis test with Bonferroni correction showed that the results had a very high degree of statistical significance between $Com < Ref, Org >$ and other two types of debunking methods ($Com < Ref, Per, Org >$ and $Com < Ref, Per >$) for *News* rumor; for *PatientZero* rumor, the results had a very high degree of statistical significance between $Com < Ref, Org >$ and $Com < Ref, Per, Org >$; and for *Car* rumor, the results had a very high degree of statistical significance between $Com < Ref, Org >$ and $Com < Ref, Per >$ (see **Figure 5**). According to the results in **Table 6**, for authority rumors (e.g., *News* rumor and *Patient Zero* rumor), the combination method was the most effective in debunking rumors. In particular, as shown in **Figure 5B**, in the combination method for *News* rumor, $Com < Ref, Org >$ was the most effective debunking method ($DEI_{NewsCom<Ref,Org>} = 1.875$), followed by $Com < Ref, Per >$ ($DEI_{NewsCom<Ref,Per>} = 1.798$, $H = 98.246$, $p < 0.001$). Additionally, $Com < Ref, Per, Org >$ had the weakest debunking effect ($DEI_{NewsCom<Ref,Per,Org>} = 1.714$). However, for *the PatientZero* rumor, as shown in **Figure 5D**, $Com < Ref, Per, Org >$ was the most effective for debunking rumors ($DEI_{PatientZeroCom<Ref,Per,\ Org>} = 1.683$), followed by $Com < Ref, Org >$ ($DEI_{PatientZeroCom<Ref,\ Org>} = 1.520$, $H = 80.030$, $p < 0.001$). Additionally, in *post-hoc* comparisons, for *Patient Zero* rumor, although $Com < Ref, Per >$ was the most effective in debunking rumors ($DEI_{PatientZeroCom<Ref,\ Per>} = 2.250$), the results had a low degree of statistical significance between Com < Ref, Org> and Com < Ref, Per> (*adjusted* $p = 0.068$), and between Com < Ref, Per, Org> and Com < Ref, Per> (*adjusted* $p = 1.000$).

Third, the results in **Table 6** show that for the social category of *Car* rumor and political category of *Russia* rumor, although the refutation method was the most effective debunking way, the combination method still achieved a suboptimal debunking effect. Thus, we still considered the decomposition combination method; the distributions of the corresponding combination

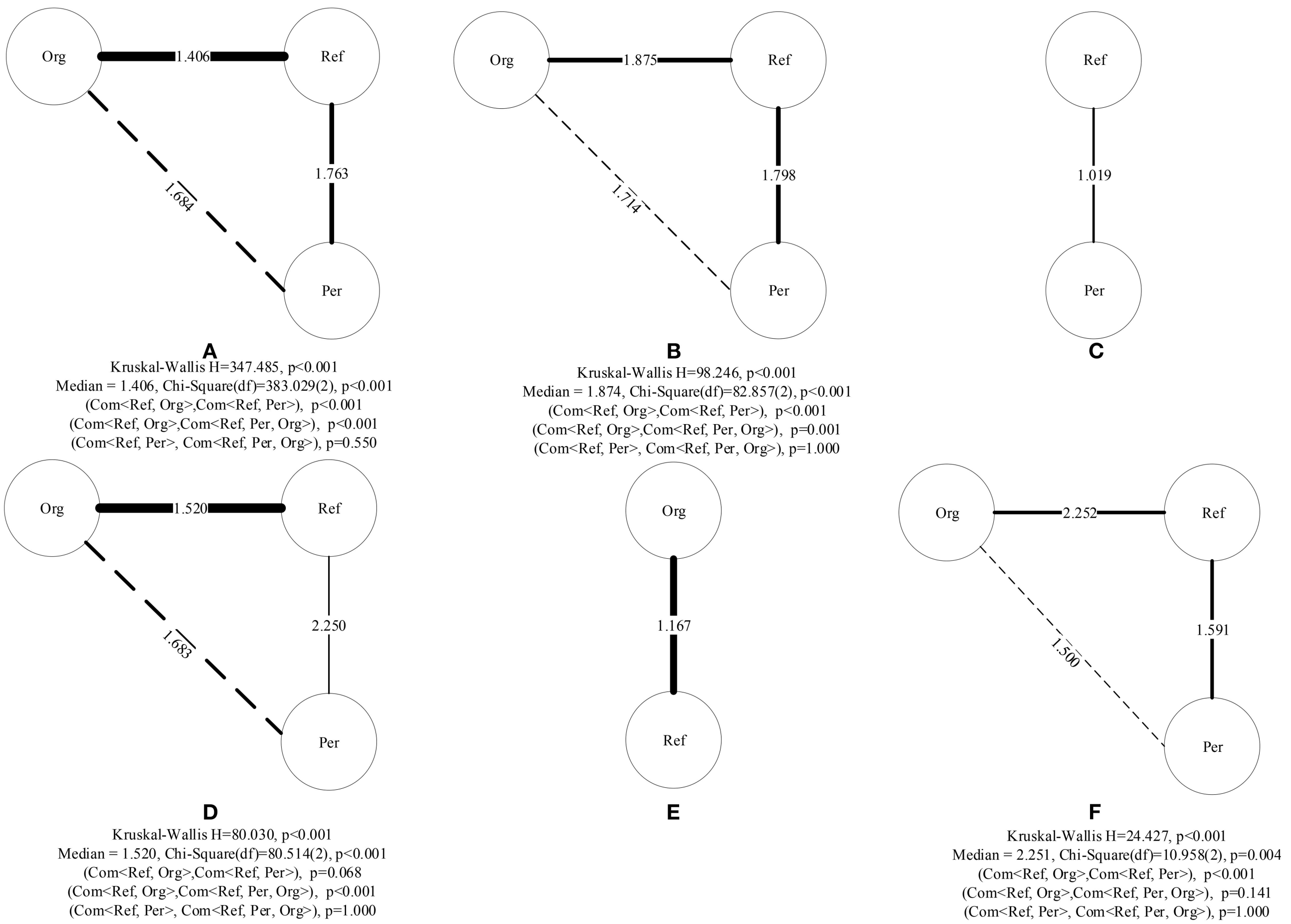


FIGURE 5 | Distribution of debunking methods in a combination method: **(A)** overall, **(B)** *News* rumor, **(C)** *Jiangsu* rumor, **(D)** *PatientZero* rumor, **(E)** *Russia* rumor, and **(F)** *Car* rumor. (a) Although certain combinations of methods are missing in some rumors, our study is based on available data to obtain the best combination of methods to be used for debunking rumors. (b) The *p*-value was the adjusted *p*-value, which was employed with a Bonferroni-type adjustment of *p*-value.

methods are shown in **Figures 5E,F**. Interestingly, for the two rumors, the debunking effectiveness of the combination method did not exceed that obtained using the refutation method alone. Our results indicated that using the refutation method alone could achieve satisfactory rumor debunking effects for these two rumors.

Fourth, from **Table 6** and **Figure 5**, it is evident that the same debunking method differs significantly in effectiveness for different rumors. For example, for the *Car* rumor and *Russia* rumor, the debunking effects of the refutation method are also different (for example, $DEI_{CarRef} = 2.522$ and $DEI_{RussiaRef} = 1.047$). Similarly, the combination method has different debunking effects in *News* rumor ($DEI_{NewsCom<Ref,Org>} = 1.875$, $DEI_{NewsCom<Ref,Per,\ Org>} = 1.714$) and *PatientZero* rumor ($DEI_{PatientZeroCom<Ref,Org>} = 1.520$, $DEI_{PatientZeroCom<Ref,Per,\ Org>} = 1.683$). Generally, in a manner similar to that for the refutation method, the debunking effect for *Car* rumor is better than that for *Russia* rumor. For the same combination method of $Com < Ref, Org >$ and $Com < Ref, Per, Org >$, if an authoritative person is slandered by the rumor, the debunking effect is better than that for the case in which an authoritative organization is slandered by the rumor.

DISCUSSION

This study offers three key findings, based on the research results. First, our analysis results showed that, among the six commonly used debunking methods, the refutation method with cited evidence is the most effective method for debunking rumors. Both the refutation and combination methods can achieve satisfactory results, and each combination method contains the refutation method. Thus, for each rumor, the refutation method with cited evidence plays a greater role in debunking rumors. This finding is consistent with those in previous research on rumor psychology literature, that is, the refutation method provides evidence that indicates why false information should not be believed (24). A convincing explanation is a must for debunking inaccurate information and preventing further propagation thereof (41). Compared with the debunking method without evidence, there are two main reasons for the better debunking effect of the refutation method with cited evidence: (a) the vividness and persuasiveness of messages (75), and (b) an evidence-providing refutation message may be perceived as more lucid and persuasive than a flat denial (42). For the latter reason, combined with the rumor psychology formula ($Rumor = Importance \times Ambiguity, R = I \times A$) proposed by Allport and Postman in 1947 (76, 77), the more important the rumor information or the more obscure the evidence, the easier it is to spread, meaning that a refutation method with detailed debunking evidence can have a positive effect (78). This shows that when debunking rumors, the government and other authorities must provide persuasive evidence related to the incident to weaken the ambiguity of rumors and help people identify rumors.

Second, there are different best-performing debunking methods for different rumors. For *Car* rumor and *Russia* rumor, using refutation alone can most effectively combat rumors. Compared with *Russia* rumor, we found that the same refutation method has better debunking effects on *Car* rumor. Based on the analysis of the characteristics of the rumor, we believe that the effectiveness of debunking *Russia* rumor is lower because of the following two possible reasons. First, the global adoption of the Internet has accelerated the rapid spread of the political category of *Russia* rumor among countries and regions. Because of misunderstandings in translation and policies, it has become highly difficult to expose rumors. Second, because of factors such as international political relations, geography, and political sensitivity of the event, there are only a few subsequent reports tracking debunking information, thereby increasing the uncertainty and ambiguity around the spread of rumors. This shows that when dealing with the social category of *Car* rumor and political category of *Russia* rumor, it is necessary to provide concise and strong evidence to debunk it while requiring more focus on *Russia* rumor. For the authority category of *News* rumor and *PatientZero* rumor, our results show that the combination method is the most effective for debunking rumors. This may be because of the trust in authoritative persons and organizations, and people may expect more information or evidence from these authoritative sources, no matter these authoritative sources spread true or false news online (79). Especially during the COVID-19 pandemic, the public has expressed a high degree of trust and expectations in authoritative persons and organizations (80), thereby also reflecting the urgent need for information about the latest developments in COVID-19 treatment. Therefore, a combination of multiple debunking methods is more effective in combating such rumors. Additionally, comparing different rumors, our results show that a combination of too many debunking methods may not have a satisfactory effect in deterring the spread of rumors. This may be because a combination of too many debunking methods would interfere with the normal information selection and cognitive analysis performed by people, decreasing the attention and recognition ability of people owing to information overload, thereby increasing the chances of people being misled by rumors (81).

Third, for rumors of the same category, the corresponding debunking strategies must be adopted according to the different characteristics of rumors. In this study, we selected *News* rumor and *PatientZero* rumor, which belong to the same category as authority rumors, and analyzed the combination method that had the optimal debunking effect. First, when using $Com < Ref, Org >$ and $Com < Ref, Per, Org >$ to debunk rumors, the debunking effect for the *News* rumor was better than that for the *PatientZero* rumor. This may be because in the *PatientZero* rumor, owing to public dissatisfaction with the delayed treatment of the early COVID-19 epidemic by the Wuhan government and the accumulation of negative emotions, people doubted the abilities of the government or authoritative organizations to deal with the COVID-19 epidemic, thereby weakening the credibility of official institutions. Thus, irrespective of the measures taken, the public will give negative comments, and consequently, the authoritative organizations will fall into the Tacitus trap owing to the loss of credibility (82, 83). Behind the loss of credibility of the government, there is a psychological phenomenon: the public habitually questions the government (84). This kind of habitual questioning is deep thought and is instinctive but persists for a long time; that is, it will continue to exist regardless of the evidence that is later presented. Therefore, to gain the trust of the public and improve their credibility, governments and other authoritative organizations should focus on the construction of public opinion ecology. Second, for the *PatientZero* rumor, using $Com < Ref, Per, Org >$ to debunk rumors, the debunking effect is better than using $Com < Ref, Org >$. This might be because, in the *Patient Zero* rumor, most of the public indicated the tendency of a person (i.e., Yanling Huang) slandered by the rumor to respond, and the combinations using the refutation, organization response, and personal response methods satisfy the public demand, thus achieving a better debunking effect. This suggests that to curb rumors early, governments should focus on the needs of the public, respond on time to their concerns, and provide more support to meet their information needs. Specifically, when an authoritative organization is slandered by a rumor (e.g., the *Patient Zero* rumor), $Com < Ref, Per, Org >$ is the most effective combination method, and when the authoritative person is slandered by the rumor (e.g., the *News* rumor), $Com < Ref, Org >$ is the most effective combination method. Therefore, for authoritative rumors, the choice of the most effective debunking method varies depending on those slandered by the rumor (i.e., person or organization) to debunk rumors. In other words, the most effective combination method requires avoiding the use of a combination of a debunking method where the person or organization defamed by the authoritative rumor responds personally and the refutation method.

Strengths and Limitations

This study is significant on two fronts, both in theory and practice. On the theoretical front, first, this study proposed a new annotation scheme for debunking methods, aiming to propose corresponding debunking response strategies for different rumors, which enriches research related to debunking methods for social media. To the best of our knowledge, this is the first study to establish a comprehensive coding scheme that can be used to categorize debunking methods in debunking postings, particularly on Sina Weibo. Second, this study proposed a novel method to measure the effectiveness of different debunking methods based on stance responses to debunking postings in user comments. To the best of our knowledge, this study is one of the first to develop a measurement for the effectiveness of debunking rumors. Compared with the decades of well-researched experimental, meta-analysis, questionnaire, and survey work done on false rumors and their correction (46, 47), this study uses deep learning methods to develop optimal classifiers to detect user stance in comments and focuses on developing an index to measure the debunking effectiveness based on stance responses to debunking postings in user comments on social media, which is more objective and scientific. On the one hand, the biggest validity of this assessment method is that it can provide relevant organizations with an effective way to utilize the massive amount of objective data from social media to detect users' stances toward rumors or debunking messages in real-time and achieve timely and effective debunking of rumors. On the other hand, this assessment method does not need to control the manipulation of debunking methods in many traditional works (e.g., randomly assigning people to different methods or controlling for many alternative explanations for findings), which makes debunking rumors on social media more cost-effective in terms of labor, money, and time, and increases the effectiveness of debunking. Our findings have enriched the literature on the mechanism of online effective rumor debunking management and intervention from a text mining perspective during the COVID-19 pandemic, which has rarely been researched in the field of information systems and public health.

This study provides three practical implications for news outlet professionals, social platform managers, and Chinese government regulators regarding the use of rebuttals in combating online rumors during the COVID-19 pandemic. First, we would expect media practitioners and news organizations to comply with professional values while debunking rumors, such as publishing well-founded news after fact-checking, quoting trustworthy sources, and providing appropriately detailed evidence. In particular, practitioners should do a detailed verification of the many aspects of information released by authoritative sources. One of the main reasons is when such fake news stories are conveyed by authoritative figures, as in the case of the lab incident in Wuhan while testing an HIV vaccine (79). When this scenario finds the support of authoritative persons or organizations, it becomes popular and quasi-real, even without being supported by data or evidence, feeding off the idea that there is some sort of plot to silence people that are perceived as menacing or challenging the status quo in the scientific community (79). Therefore, regardless of who publishes debunking information, practitioners should do a good job of fact-checking, they can't let their guard down just because it's the word of an authoritative person or organization. Second, social media managers, in the case of information overload, should consciously verify and strictly control information sources and their published information, take up the vital responsibility of providing users with real and high-quality information, and induce a positive public opinion environment. Third, the Chinese government staff must improve the cognitive ability of its citizens to identify rumors and provide support for the early detection and interruption of rumor spreading, such as improving their digital literacy and health literacy (i.e., the ability to obtain, read, understand, and use healthcare information to make appropriate health decisions and follow instructions for treatment) (85). Simultaneously, to enhance the public's "critical thinking ability" and "trust in science," different debunking methods should be adopted for different rumors, especially for political rumors and rumors slandering authoritative persons or organizations. Additionally, the government should enhance the monitoring of the debunking effectiveness of official communication channels on social networking sites, such as the Government Information page on Sina Weibo, consider the ecological construction of public opinion, and improve its credibility to win the approval of public opinion. Government-affiliated accounts should leverage social platforms to combat online rumors, attend to the needs and intentions of the public, pay attention to public feedback, and formulate effective debunking strategies to mitigate the existence of rumors at the earliest. Finally, the government and authoritative organizations should improve their credibility by appropriately encouraging commentary-based interactions of the public on social media platforms and enhancing public trust and intimacy with the government.

However, this study has some limitations. First, the sample size and the type of social media platform surveyed were limited owing to the threat of the rapidly spreading COVID-19 pandemic. This paper focuses on the study of five rumors related to COVID-19 in the Chinese social media environment, the generalization of the results to other countries and cultures is constrained. Thus, in future relevant studies, better results might be achieved using adequate sample size and multiple platforms and making comparisons between platforms, such as Sina Weibo and Twitter. Second, although we attempted to use the optimal classifier for text classification in this study, some errors cannot be avoided in the text analysis process. On the one hand, future research could improve the accuracy of the classifier by improving the algorithm to classify users. On the other hand, in this study, the performance of classifiers was evaluated using the macro-F1 score, which was more intuitive but gave equal weight to precision and recall and was ineffective regarding true negatives. Thus, future research can use other high-quality evaluation methods at the same time. For example, for an unbalanced classification problem, future work will consider the use of Receiver Operating Characteristic (ROC) and Area Under Curve (AUC) to measure the performance of classification models, a weighted arithmetic mean with weight p given to recall and weight $(1-p)$ given to precision to express F-measure (67), and measurement methods in the general multiclass case considering connections between the concepts of informedness, markedness, correlation, and significance as well as their intuitive relationships with recall and precision (86). Third, although we considered two aspects of the effectiveness of debunking, we only used one measure to evaluate the effectiveness of debunking and focused on the post-debunking situation. Thus, on the one hand, future work will consider developing multiple measures that can summarize all aspects of effectiveness, for example, adding an aspect that examines the impact of postings from the perspective of social network structure, such as the depth, breadth, and structural virality of debunking postings, aspects of the characteristics and influence of refuters, and the corresponding aspects of rumor denial and information redundancy. On the other hand, we will consider a better measurement of debunking effectiveness from the perspective of causal effects, such as tracking rumors in real-time, and consider the change in veracity status (such as unverified, true, false) at different points of the life cycle of a rumor, to construct a better measurement of debunking effectiveness which depends on both situations before and after debunking. Fourth, rumors appeal to people because they seem to be able to express or gratify their emotional needs, especially the catharsis of negative emotions related to anger, frustration, hatred, and/or anxiety. Therefore, future research should consider combining the user stance expression, subdividing the problems exposed to negative emotions, and investigating its related influencing factors. Finally, as rumors are often short-lived, there is a temporal dimension that seems to tackle further consideration. This study only portrayed and described the overall phenomenon of rumor debunking and did not consider temporality. Thus, future research will consider analyzing the state changes of various stances that evolve at different periods of the life cycle of a rumor, such as when comments are made or how long rumors or debunking messages persist or remain persuasive, and reveal the underlying reasons behind such changes.

CONCLUSIONS

This study proposed a new annotation scheme to categorize the debunking methods of postings on Sina Weibo. We built text classification models to automatically detect stance responses to debunking postings in user comments. Based on the results of identifying user stance in comments, we proposed a new method to measure the effectiveness of different debunking methods. In addition, we used five rumors related to COVID-19 as cases and compared the effectiveness of different debunking methods. Our main findings are as follows: First, the refutation method is the primary choice among debunking methods and has the best debunking effect, whether it is used alone or in combination with other debunking methods. Second, for the social category of *Car* rumor and political category of *Russia* rumor, using the refutation method alone can achieve the optimal debunking effect. Third, for authoritative rumors, a combination method has the optimal debunking effect, but the most effective combination method requires avoiding using a combination of a debunking method where the person or organization defamed by the authoritative rumor responds personally and the refutation method. Furthermore, we believe this study is significant on two fronts. First, we proposed a new scheme for the classification of debunking methods and formulated a novel method to measure the effectiveness of debunking methods by analyzing the stances of users' comments toward different debunking postings. Second, for each rumor, we revealed debunking strategies that could effectively prevent people from spreading rumors. Our research findings provide relevant insights into ways to effectively debunk rumors amid public health emergencies on social media, aiming to support the crisis management of rumors, with a view to taking necessary actions in response to COVID-19 rumor outbreaks.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XW: conceptualization, methodology, software, visualization, formal analysis, data curation, and writing. FC: data curation, methodology, and formal analysis. GY: conceptualization, writing, and funding acquisition.

FUNDING

This research was funded by the National Natural Science Foundation of China (Grant No. 72074060).

ACKNOWLEDGMENTS

The authors would like to acknowledge the editor's contribution and show appreciation to the reviewers for their helpful comments and recommendations. The authors also would like to acknowledge all the authors cited in the references. The authors would like to thank Zhiwei Data Sharing Platform (http://university.zhiweidata.com/) for our data support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.770111/full#supplementary-material

REFERENCES

1. World Health Organization. *Coronavirus Disease 2019 (COVID-19) Situation Report-51*. Geneva: World Health Organization (2020).
2. Hui HW, Zhou CC, Lu X, Li JR. Spread mechanism and control strategy of social network rumors under the influence of COVID-19. *Nonlinear Dyn.* (2020) 101:1933–49. doi: 10.1007/s11071-020-05842-w
3. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. *Soc Sci Med.* (2020) 263:8. doi: 10.1016/j.socscimed.2020.113356
4. World Health Organization. *Coronavirus Disease 2019 (COVID-19) Situation Report- 45*. Geneva: World Health Organization (2020).
5. World Health Organization. *Coronavirus Disease 2019 (COVID-19) Situation Report-13*. Geneva: World Health Organization (2020).

6. Nazar S, Pieters T. Plandemic revisited: a product of planned disinformation amplifying the COVID-19 "infodemic." *Front Public Health*. (2021) 9:649930. doi: 10.3389/fpubh.2021.649930
7. Banerjee D, Meena K. COVID-19 as an "Infodemic" in public health: critical role of the social media. *Front Public Health.* (2021) 9:231. doi: 10.3389/fpubh.2021.610623
8. Gradoń KT, Hołyst JA, Moy WR, Sienkiewicz J, Suchecki K. Countering misinformation: a multidisciplinary approach. *Big Data Society.* (2021) 8:20539517211013848. doi: 10.1177/20539517211013848
9. Rovetta A, Castaldo L. Influence of mass media on Italian web users during the COVID-19 pandemic: infodemiological analysis. *JMIRx Med.* (2021) 2:e32233. doi: 10.2196/32233
10. Anwar A, Malik M, Raees V, Anwar A. Role of mass media and public health communications in the COVID-19 pandemic. *Cureus.* (2020) 12:e10453. doi: 10.7759/cureus.10453
11. Ali K, Li C, Zaffar MA. Fake news on Facebook: examining the impact of heuristic cues on perceived credibility and sharing intention. *Internet Res.* (2021). doi: 10.1108/INTR-10-2019-044. [Epub ahead of print].
12. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science.* (2018) 359:1146–51. doi: 10.1126/science.aap9559
13. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nature Human Behav.* (2020) 4:1285–93. doi: 10.1038/s41562-020-00994-6
14. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W. Science vs conspiracy: collective narratives in the age of misinformation. *PLoS ONE.* (2015) 10:e0118093. doi: 10.1371/journal.pone.0118093
15. Alvarez-Galvez J, Suarez-Lledo V, Rojas-Garcia A. Determinants of infodemics during disease outbreaks: a systematic review. *Front Public Health.* (2021) 9:236. doi: 10.3389/fpubh.2021.603603
16. Wang YX, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. *Soc Sci Med.* (2019) 240:12. doi: 10.1016/j.socscimed.2019.112552
17. Vos SC, Buckner MM. Social media messages in an emerging health crisis: tweeting bird flu. *J Health Commun.* (2016) 21:301–8. doi: 10.1080/10810730.2015.1064495
18. Househ M. Communicating Ebola through social media and electronic news media outlets: a cross-sectional study. *Health Informatics J.* (2016) 22:470–8. doi: 10.1177/1460458214568037
19. MacFarlane D, Tay LQ, Hurlstone MJ, Ecker UK. Refuting spurious COVID-19 treatment claims reduces demand and misinformation sharing. *J Appl Res Mem Cogn.* (2021) 10:248–58. doi: 10.1016/j.jarmac.2020.12.005
20. Moravec PL, Minas RK, Dennis AR. Fake news on social media: people believe what they want to believe when it makes no sense at all. *Mis Q.* (2019) 43:18–87. doi: 10.25300/misq/2019/15505
21. Wang Q, Yang X, Xi W. Effects of group arguments on rumor belief and transmission in online communities: an information cascade and group polarization perspective. *Inf Manag.* (2018) 55:441–9. doi: 10.1016/j.im.2017.10.004
22. Wang R, He Y, Xu J, Zhang H. Fake news or bad news? Toward an emotion-driven cognitive dissonance model of misinformation diffusion. *Asian J Commun.* (2020) 30:317–42. doi: 10.1080/01292986.2020.1811737
23. Jung A-K, Ross B, Stieglitz S. Caution: rumors ahead—a case study on the debunking of false information on Twitter. *Big Data Society.* (2020) 7:2053951720980127. doi: 10.1177/2053951720980127
24. DiFonzo N, Bordia P. *Rumor Psychology: Social and Organizational Approaches*. Washington, DC: American Psychological Association (2007). doi: 10.1037/11503-000
25. Dunn HB, Allen CA (editors). Rumors, urban legends and internet hoaxes. In: *Proceedings of the Annual Meeting of the Association of Collegiate Marketing Educators*. Dallas, TX (2005).
26. Spiro ES, Fitzhugh S, Sutton J, Pierski N, Greczek M, Butts CT (editors). Rumoring during extreme events: a case study of Deepwater Horizon 2010. In: *Proceedings of the 4th Annual ACM Web Science Conference.* (2012). p. 275–83.doi: 10.1145/2380718.2380754
27. Jang JW, Lee EJ, Shin SY. What debunking of misinformation does and doesn't. *Cyberpsychol Behav Soc Netw.* (2019) 22:423–7. doi: 10.1089/cyber.2018.0608
28. Garrett RK. Troubling consequences of online political rumoring. *Hum Commun Res.* (2011) 37:255–74. doi: 10.1111/j.1468-2958.2010.01401.x
29. Wang BR, Zhuang J. Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. *Natural Hazards.* (2018) 93:1145–62. doi: 10.1007/s11069-018-3344-6
30. Zhang XC, Ghorbani AA. An overview of online fake news: Characterization, detection, and discussion. *Inf Process Manag.* (2020) 57:26. doi: 10.1016/j.ipm.2019.03.004
31. Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media: a data mining perspective. *Acm Sigkdd Explorations Newsletter.* (2017) 19:22–36. doi: 10.1145/3137597.3137600
32. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. *Psychol Sci Public Interest.* (2012) 13:106–31. doi: 10.1177/1529100612451018
33. Malhotra P. A relationship-centered and culturally informed approach to studying misinformation on COVID-19. *Soc Med Soc.* (2020) 6:4. doi: 10.1177/2056305120948224
34. Vraga EK, Kim SC, Cook J. Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *J Broadcast Electron Media.* (2019) 63:393–414. doi: 10.1080/08838151.2019.1653102
35. Vraga EK, Bode L. I do not believe you: how providing a source corrects health misperceptions across social media platforms. *Info Commun Soc.* (2018) 21:1337–53. doi: 10.1080/1369118x.2017.1313883
36. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. *Sci Commun.* (2017) 39:621–45. doi: 10.1177/1075547017731776
37. Alexander DE. Social media in disaster risk reduction and crisis management. *Sci Eng Ethics.* (2014) 20:717–33. doi: 10.1007/s11948-013-9502-z
38. Zubiaga A, Liakata M, Procter R, Hoi GWS, Tolmie P. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE.* (2016) 11:29. doi: 10.1371/journal.pone.0150989
39. Ozturk P, Li H, Sakamoto Y (editors). Combating rumor spread on social media: the effectiveness of refutation and warning. In: *2015 48th Hawaii International Conference on System Sciences.* Kauai (2015). p. 2406–14. doi: 10.1109/HICSS.2015.288
40. Li Z, Zhang Q, Du X, Ma Y, Wang S. Social media rumor refutation effectiveness: evaluation, modelling and enhancement. *Inf Process Manag.* (2021) 58:102420. doi: 10.1016/j.ipm.2020.102420
41. Kimmel AJ, Audrain-Pontevia A-F. Analysis of commercial rumors from the perspective of marketing managers: rumor prevalence, effects, and control tactics. *J Market Commun.* (2010) 16:239–53. doi: 10.1080/13527260902884433
42. Paek HJ, Hove T. Mediating and moderating roles of trust in government in effective risk rumor management: a test case of radiation-contaminated seafood in South Korea. *Risk Anal.* (2019) 39:2653–67. doi: 10.1111/risa.13377
43. Paek H-J, Hove T. Effective strategies for responding to rumors about risks: the case of radiation-contaminated food in South Korea. *Public Relat Rev.* (2019) 45:101762. doi: 10.1016/j.pubrev.2019.02.006
44. Yang L-X, Zhang T, Yang X, Wu Y, Tang YY. Effectiveness analysis of a mixed rumor-quelling strategy. *J Franklin Inst.* (2018) 355:8079–105. doi: 10.1016/j.jfranklin.2018.07.040
45. Lee J, Kanthawala S, Britt BC, Deavours DF, Ott-Fulmore T. Prevalence of anger, engaged in sadness: engagement in misinformation, correction, and emotional tweets during mass shootings. *Online Inf Rev.* (2021). doi: 10.1108/OIR-03-2021-0121. [Epub ahead of print].
46. Pal A, Chua AY, Goh DH-L. Debunking rumors on social media: the use of denials. *Comput Human Behav.* (2019) 96:110–22. doi: 10.1016/j.chb.2019.02.022
47. Walter N, Brooks JJ, Saucier CJ, Suresh S. Evaluating the impact of attempts to correct health misinformation on social media: a meta-analysis. *Health Commun.* (2020) 1–9. doi: 10.1080/10410236.2020.1794553
48. Rodríguez CP, Carballido BV, Redondo-Sama G, Guo M, Ramis M, Flecha R. False news around COVID-19 circulated less on Sina Weibo than on Twitter. How to overcome false information? *Int Multidisc J Social Sci.* (2020) 9:107–28. doi: 10.17583/rimcis.2020.5386
49. Tian X, Yu G, He F. An analysis of sleep complaints on Sina Weibo. *Comput Human Behav.* (2016) 62:230–5. doi: 10.1016/j.chb.2016.04.014
50. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* (1960) 20:37–46. doi: 10.1177/001316446002000104

51. Reichenheim ME. Confidence intervals for the kappa statistic. *Stata J.* (2004) 4:421–8. doi: 10.1177/1536867X0400400404
52. Rosnow RL. Inside rumor: a personal journey. *Am Psychol.* (1991) 46:484. doi: 10.1037/0003-066X.46.5.484
53. Bordia P, DiFonzo N, Haines R, Chaseling E. Rumors denials as persuasive messages: Effects of personal relevance, source, and message characteristics 1. *J Appl Soc Psychol.* (2005) 35:1301–31. doi: 10.1111/j.1559-1816.2005.tb02172.x
54. Zubiaga A, Liakata M, Procter R, Bontcheva K, Tolmie P, Acm. Crowdsourcing the annotation of rumourous conversations in social media. In: *Proceedings of 2015 International Conference on World Wide Web.* (2015). p. 347–53. doi: 10.1145/2740908.2743052
55. Dungs S, Aker A, Fuhr N, Bontcheva K, editors. Can rumour stance alone predict veracity. In: *International Conference on Computational Linguistics.* (2018). p. 3360–70.
56. Derczynski L, Bontcheva K, Liakata M, Procter R, Hoi GWS, Zubiaga A (editors). SemEval-2017 task 8: rumoureval: determining rumour veracity and support for rumours. Meeting of the association for computational linguistics. *arXiv [Preprint]* arXiv:1704.05972. (2017)
57. Zubiaga A, Kochkina E, Liakata M, Procter R, Lukasik M, Bontcheva K, et al. Discourse-aware rumour stance classification in social media using sequential classifiers. *Inf Process Manag.* (2018) 54:273–90. doi: 10.1016/j.ipm.2017.11.009
58. Zhou S, Ng ST, Lee SH, Xu FJ, Yang Y, A. domain knowledge incorporated text mining approach for capturing user needs on BIM applications. *Eng Constr Architect Manag.* (2019) 27:458–82. doi: 10.1108/ECAM-02-2019-0097
59. Wang P, Pei X, Yin X-P, Ren J-L, Wang Y, Ma L-Y, et al. Radiomics models based on enhanced computed tomography to distinguish clear cell from non-clear cell renal cell carcinomas. *Sci Rep.* (2021) 11:1–8. doi: 10.1038/s41598-021-93069-z
60. Lv H, Dao F-Y, Guan Z-X, Yang H, Li Y-W, Lin H. Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method. *Brief Bioinf.* (2021) 22:bbaa255. doi: 10.1093/bib/bbaa255
61. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv* [Preprint]. (2018) arXiv:181004805.
62. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. Roberta: a robustly optimized bert pretraining approach. *arXiv* [Preprint]. (2019) arXiv:190711692.
63. Cui Y, Che W, Liu T, Qin B, Wang S, Hu G. Revisiting pre-trained models for chinese natural language processing. *arXiv* [Preprint]. (2020) arXiv:200413922.
64. Cui Y, Che W, Liu T, Qin B, Yang Z, Wang S, et al. Pre-training with whole word masking for chinese bert. *arXiv* [Preprint]. (2019) arXiv:190608101.
65. Zhang D, Wang J, Zhao X, editors. Estimating the uncertainty of average F1 scores. In: *Proceedings of the 2015 International Conference on The Theory of Information Retrieval.* New York, NY (2015). doi: 10.1145/2808194.2809488
66. Asch VV. *Macro-and Micro-Averaged Evaluation Measures.* Antwerp: Univercity of Antwerp. (2013).
67. Hand D, Christen P. A note on using the F-measure for evaluating record linkage algorithms. *StCom.* (2018) 28:539–47. doi: 10.1007/s11222-017-9746-6
68. Yang Y, Liu X, editors. A re-examination of text categorization methods. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* Berkeley, CA (1999). p. 42–9.
69. Sharpe D. Chi-square test is statistically significant: Now what? *Pract Assess Res Eval.* (2015) 20:8. doi: 10.7275/tbfa-x148
70. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc.* (1952) 47:583–621. doi: 10.1080/01621459.1952.10483441
71. Katz BM, McSweeney M. A multivariate Kruskal-Wallis test with *post hoc* procedures. *Multivariate Behav Res.* (1980) 15:281–97. doi: 10.1207/s15327906mbr1503_4
72. Amrhein V, Korner-Nievergelt F, Roth T. The earth is flat (p> 005): significance thresholds and the crisis of unreplicable research. *PeerJ.* (2017) 5:e3544. doi: 10.7717/peerj.3544
73. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol.* (2016) 31:337–50. doi: 10.1007/s10654-016-0149-3
74. Iyer ES, Debevec K. Origin of rumor and tone of message in rumor quelling strategies. *Psychol Market.* (1991) 8:161–75.
75. Allport GW, Postman L. An analysis of rumor. *Public Opin Q.* (1946) 10:501–17. doi: 10.1093/poq/10.4.501
76. Allport GW, Postman L. *The Psychology of Rumor*. Washington, DC: *Henry Holt* (1947).
77. Chan MPS, Jones CR, Jamieson KH, Albarracin D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci.* (2017) 28:1531–46. doi: 10.1177/0956797617714579
78. Moscadelli A, Albora G, Biamonte MA, Giorgetti D, Innocenzio M, Paoli S, et al. Fake news and covid-19 in Italy: results of a quantitative observational study. *Int J Environ Res Public Health.* (2020) 17:5850. doi: 10.3390/ijerph17165850
79. Falcone R, Sapienza A. How COVID-19 changed the information needs of Italian citizens. *Int J Environ Res Public Health.* (2020) 17:19. doi: 10.3390/ijerph17196988
80. Rubin VL. Disinformation and misinformation triangle A conceptual model for "fake news" epidemic, causal factors and interventions. *J Doc.* (2019) 75:1013–34. doi: 10.1108/jd-12-2018-0209
81. Giddens A. Risk and responsibility. *Mod L Rev.* (1999) 62:1. doi: 10.1111/1468-2230.00188
82. Horn E. Logics of political secrecy. *Theory Cult Soc.* (2011) 28:103–22. doi: 10.1177/0263276411424583
83. Chen S, Yang J. Cultivation of government's credibility in internet age and analyzation of tacitus effect. In: *Proceedings of 2014 International Conference on Public Administration. Vol. 10.* (2014). p. 347–54.
84. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *Eur J Public Health.* (2015) 25:1053–8. doi: 10.1093/eurpub/ckv043
85. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* [Preprint]. (2020) arXiv:201016061.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Public Health

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fpubh.2021.789569

Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM

Ashwini Kodipalli [1,2]* **and** ***Susheela Devi*** [1,2]

[1] Department of Computer Science and Automation, Indian Institute of Science, Bengaluru, India, [2] Department of Artificial Intelligence and Data Science, Global Academy of Technology, Bengaluru, India

Polycystic ovarian syndrome (PCOS) is a hormonal disorder found in women of reproductive age. There are different methods used for the detection of PCOS, but these methods limitedly support the integration of PCOS and mental health issues. To address these issues, in this paper we present an automated early detection and prediction model which can accurately estimate the likelihood of having PCOS and associated mental health issues. In real-life applications, we often see that people are prompted to answer in linguistic terminologies to express their well-being in response to questions asked by the clinician. To model the inherent linguistic nature of the mapping between symptoms and diagnosis of PCOS a fuzzy approach is used. Therefore, in the present study, the Fuzzy Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method is evaluated for its performance. Using the local yet specific dataset collected on a spectrum of women, the Fuzzy TOPSIS is compared with the widely used support vector machines (SVM) algorithm. Both the methods are evaluated on the same dataset. An accuracy of 98.20% using the Fuzzy TOPSIS method and 94.01% using SVM was obtained. Along with the improvement in the performance and methodological contribution, the early detection and treatment of PCOS and mental health issues can together aid in taking preventive measures in advance. The psychological well-being of the women was also objectively evaluated and can be brought into the PCOS treatment protocol.

Keywords: support vector machines, fuzzy TOPSIS, fuzzy AHP, polycystic ovarian syndrome, mental health issues, machine learning, classifiers, fuzzy logic

OPEN ACCESS

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Dinesh Ramegowda,
Samsung, India
Chandan Hegde,
Surana Educational Institutes, India
Duddela Sai Prashanth,
Sahyadri College of Engineering and Management, India

***Correspondence:**
Ashwini Kodipalli
kashwini@iisc.ac.in

Specialty section:
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: *05 October 2021*
Accepted: *25 October 2021*
Published: *30 November 2021*

Citation:
Kodipalli A and Devi S (2021)
Prediction of PCOS and Mental Health
Using Fuzzy Inference and SVM.
Front. Public Health 9:789569.
doi: 10.3389/fpubh.2021.789569

INTRODUCTION

Polycystic ovarian syndrome (PCOS) is a hormonal disorder usually found in young adult women and in women of reproductive age. It is a common and pervasive hormonal disorder with multiple phenotypes having different presentations (1). PCOS was first discovered by Stein and Leventhal (2). According to WHO, around 6–26% of women are affected by PCOS. It includes failure to ovulate and infertility, along with mental health complications. PCOS has several effects ranging from acne and obesity to irregular menstruation and infertility, which might lead to imparity in the quality of life (3–5). PCOS can be detected using the Rotterdam Consensus, where the person under study should meet at least two of these conditions: (a) absence of ovulation, (b) symptoms of hyperandrogenism, and (c) ovaries affected with polycysts (6). PCOS causes several effects such as amenorrhea, obesity, type 2 diabetes, metabolic impairment, and cardiovascular disorders. As per WHO, the prevalence of PCOS is estimated to be between 6 and 26% (7). In India, the prevalence of PCOS ranges between 9.13 and 36% (8). PCOS is one of the major factors for infertility and emotional problems. Approximately up to one-third of females with PCOS have fertility issues. Women with PCOS have more anxiety and depression than women in general. The percentage of depression and anxiety in women with PCOS is 28–39% and 11–25%, respectively (9, 10). International research shows that PCOS has a significant effect on both physical and emotional well-being and, hence, affects the quality of life (11).

Recent research has shown that PCOS can be detected using an ultrasound scan where the doctor counts the size and number of follicles in the ovaries. PCOS can also be detected using biochemical examinations such as blood tests and hormonal examinations, yet it is an expensive investigation. It can also be predicted using clinical parameters such as menstrual cycle length and body mass index (BMI), however, it involves clinical acumen and the subjectivity of the clinician, as well as the intrusion of privacy of the women; also, the process is not automated and the emotional wellness of the women is not considered.

In the literature, various methods are proposed to detect PCOS. They are broadly categorized into two groups.

1. The studies where the association between PCOS and mental health is investigated using various statistical tools such as χ^2 and Fisher's exact test. Here, however, the diagnosis of PCOS and its accuracy are not considered.
2. The studies where PCOS is detected using machine learning algorithms and image processing techniques, but psychological and emotional wellness is not considered.

We were motivated by the above limitations of existing works; thus, the current research is designed to include both clinical and psychological parameters to detect and diagnose the likelihood of having PCOS and its associated mental health problems. Since the clinical and psychological data pertaining to PCOS varies from woman to woman and involves inherent uncertainty, imprecision, and ambiguity, the traditional machine learning algorithms will not model the system under consideration. The fuzzy algorithms such as the fuzzy technique for order of preference by similarity to ideal solution (TOPSIS) and fuzzy analytical hierarchal process (AHP) are evaluated for their efficacy and robustness against traditional algorithms such as support vector machines (SVM), K nearest neighbor, and decision tree (D-tree).

The organization of the paper is as follows. Related work describes the related work, the drawbacks of the existing methods, and the problem statement. Proposed method provides the proposed methodology and collection of data. Results explains the results and Discussion includes the comparative study of the results through discussion. Finally, Conclusion concludes the paper.

RELATED WORK

Some of the studies which proved the relationship between PCOS and mental health and the classification algorithms used for the prediction of PCOS are reviewed and presented below:

As per the systematic and exhaustive survey conducted by Himelein et al. (12), it was indicated that PCOS is also associated with psychological/mental health issues such as body dissatisfaction, depression, anxiety, and eating disorders, and thus, reduces the quality of life. Although in his research he did not describe the implicit relationship between PCOS and mental health, he suggested strongly that the effective treatment of PCOS can stabilize mental health issues.

Conte et al. (13) conducted a systematic review on mental health and physical exercise in women with PCOS and reported that introducing physical activity can improve the quality of life and psychological well-being of women. Kerchner et al. (14) established the relationship between PCOS and mental health and identified the risks and predictors for depression in women with PCOS. In his research, a total of 60 women with PCOS were considered and were given Primary Care Evaluation of Mental Disorders Patient Health questionnaires (PRIME-MD PHQ) (15), the Beck Depression Inventory-II (BDI) (16), and the Beck Anxiety Inventory (BAI) (17) to detect mental health disorders. The data were analyzed using χ^2 and Fisher's exact test to evaluate the categorical variables. It was found that 40% (24 out of 60) of women are suffering from depression and 16.6% suffer from mood disorders which shows that at least 56.6% of women with PCOS have mental health issues.

Wolf et al. (18) conducted a systematic review based on geographical location and ethnicity. The study found that the prevalence of PCOS was between 6 and 9% across the United States, United Kingdom, Spain, Greece, Australia, Asia, and Mexico based on the National Institutes of Health (NIH) diagnostic criteria. The study also suggested that there is no racial or ethnic influence on the prevalence of PCOS. Naz et al. (19), demonstrated the prevalence of PCOS in the adult population. In his meta-analysis, 12 published articles were considered and Egger and Begg's tests were used to check the publication bias. Using 477 participants and the STATA software, it is proved that PCOS in the adult population was 11.04% based on the Rotterdam criteria, 3.39% based on the NIH criteria, and 8.03% based on the Androgen excess and PCOS society.

Banting et al. (20), in their research along with the association of mental health and PCOS, also investigated physical inactivity, motivators, and support providers for women with and without PCOS. Using a local database and computational statistics, he concluded that a woman with PCOS requires more support compared with women without PCOS. Similarly, Berni et al. (21), investigated Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD) and proved that the incidence of ASD and ADHD is higher in children of mothers with PCOS. In a similar development, Chandhari et al. (22) investigated the association of PCOS with psychiatric morbidity and its impact on quality of life. Using binary logistic regression, he could associate PCOS with psychiatric morbidity.

Some studies investigated the detection of PCOS using computational and machine learning algorithms. In a recent development, Sachdeva et al. (23) compared the phenotypes in PCOS by considering their metabolic, clinical, and hormonal profiles, and their differential responses to clomiphene. Using descriptive statistics, the authors concluded that patients with full-blown PCOS (phenotype A) are at a higher risk of metabolic and cardiovascular disorders compared with others, thus concluding that phenotype division can help patients in their accurate diagnoses. In a recent computational study by Denny et al. (24), an early detection and prediction model for PCOS was proposed and it was concluded that the random forest classifier provided an accuracy of 89.02%. He considered 541 women with 23 attributes that consist of both clinical and metabolic parameters. Similarly, another work by Vikas et al. (25) considered 119 women with 18 attributes consisting of clinical parameters that were used to evaluate PCOS and obtained an accuracy of 97.65% using the Naïve Bayes classifier. The study conducted by Anuradha and Priyanka (26) considered 84 women and 13 attributes to detect PCOS and obtained an accuracy of 94% using artificial neural networks. The study by Deshpande and Wakankar (27) considered imagining parameters such as follicles along with the biochemical and clinical parameters such as hormonal levels and BMI for the detection of PCOS. Support vector machine was used for the classification and obtained an accuracy of 95%.

Mishra and Prakash (28) found that, for computing human cognitive capabilities, fuzzy logic is the most effective computing method. The categories of the medical data are clearly explained in this paper. Further, Ansari et al. (29), explained in their research how Fuzzy TOPSIS can be used extensively for trustworthy health care development software.

But none of the studies have considered mental health in their analysis. In all the studies, the data was obtained by interviewing the women and from the laboratory. As the characteristics of the data are uncertain and since the likelihood of having PCOS depends on multiple parameters, fuzzy algorithms are more appropriate than crisp algorithms.

The following lacunae were found from the literature survey:

1. The studies where the association between PCOS and mental health is investigated did not evaluate the performance, efficacy, and robustness of PCOS detection.
2. Among the studies where the PCOS detection is carried out using machine learning algorithms, the psychological well-being parameters are not considered.
3. Integrated studies on PCOS detection and the analysis of its associated mental health were not present in the literature.
4. The existing studies employ statistical tools. However, there are no studies that could implement soft computing approaches like fuzzy sets and multi-criteria decision analysis systems.
5. An effective and early diagnosis system that can screen many people in a time-bound and cost-effective manner does not exist.
6. There are limited computer-assisted tools that help women in the screening of PCOS and its associated mental health privately and comfortably. It only requires the woman who is to be tested to take the questionnaire and get the diagnosis immediately.

In the current study, both physical and psychological parameters are considered for developing an automated PCOS and its associated mental health detection. The clinical data collected from the diverse population is non-deterministic, irregular, imprecise, and uncertain, making soft computing approaches such as fuzzy TOPSIS and fuzzy AHP better candidate methods than traditional methods.

PROPOSED METHOD

The methodology proposed in this study is described below:

Study Design

The research is mainly focused on developing an early prediction model. The study is designed to include young adult women of age <25. Women taking treatment for acute illness, pregnant women, and women under substance abuse are excluded from this study. The study protocol was approved by the ethical committee of the institution where the study is carried out. The written consent of the participants (subjects) was collected before involving them in the study. The study involved the participation of two expert gynecologists and one psychologist who designed the questionnaires. The questionnaire included 11 questions related to physical aspects, 10 questions related to anxiety and depression using the K10 tool, 5 questions on social phobia, and 5 questions on body image dissatisfaction. **Table 1** provides a brief overview of this questionnaire. The expert panels comprising of two gynecologists and one psychologist interviewed the subjects and filled the questionnaire form. Based on the answers for different questions, the expert panels have mapped the subjects into one of the four categories, viz, A1: subject having both PCOS and mental health issues; A2: subject having only PCOS problem; A3: subject having only mental health issues; and A4: subject being normal. The expert panel filled the questions by interviewing the subjects and also mapped every subject to one of the four predefined categories.

Data Collection

In total, 660 subjects were chosen for the study where 624 have agreed to participate and others have declined to participate due to various reasons. The expert panel completed the questions by interviewing the subjects and also mapped every subject to one of the four pre-defined categories. This data is stored in a comma-separated values (CSV) file and it is used and processed for subsequent analysis by algorithms. Data visualization and data analysis, using matplotlib, were performed to gain more insight into the data. After analysis, the results of the regularity in the menstrual cycles of the subjects are shown in the form of a graph in **Figure 1**. The pie chart of the answers of the subjects regarding the length of their menstrual cycle is shown in **Figure 2**.

TABLE 1 | Brief overview of the questionnaire.

Variables	Criteria
Physical aspects	
C11	Regularity of periods
C12	Length of the menstrual cycle
C13	Duration of the flow
C14	Number of pads used per day
C15	During cycle, tendency to grow dark, coarse hair on chest and chin
C16	Weight gain
C17	Eating junk food
C18	Meal times and eating pattern
C19	Sleep Schedule/Sleep pattern
C110	Family history of diabetes
C111	Family history of hypertension
Anxiety and depression (K10 Tool)	
C21	Feel tired for no good reason
C22	Feel nervous
C23	Feel so nervous that nothing could calm you down
C24	Feel hopeless
C25	Feel restless or fidgety
C26	Feel so restless that you could not sit still
C27	Feel depressed
C28	Feel everything was an effort
C29	Feel so sad that nothing could cheer you up
C210	Feel worthless
Social phobia	
C31	Intense and persistent fear that others might evaluate you
C32	Fear of being humiliated in social situations
C33	Feeling extremely self-conscious
C34	Fear that others will notice blushing/sweating
C35	Try hard to avoid social situation/interaction
Body image dissatisfaction	
C41	Spend a lot of time worrying about their appearance
C42	Experience dissatisfaction with their appearance
C43	Avoid wearing certain cloths because they may look fat
C44	Compares their appearance with others and feel low
C45	Dissatisfaction and self-consciousness of appearance interferes with social activities and interactions

Methodology

As noted in the literature survey in the previous section, the majority of research in this area is carried out using traditional methods like the T-square test, Chi-Square test, etc. The current implementation is an approach to apply machine learning and fuzzy systems to the data and perform a comparative study of the two. **Figure 3** depicts the block diagram of the proposed method.

A woman who has agreed to participate in the study is called a subject under study. The data collection for this study includes both physical assessments (menstrual cycle, regularity of the cycle, length of the cycle, duration of the cycle, recent weight gain, hair loss, family history of having diabetes and hypertension, and eating and sleeping habits) and psychological assessments (anxiety, depression, body image dissatisfaction, and social phobia). Based upon the answers given by the subjects to the 31 questions, the panel of doctors decides to put them into one of the four categories. Once the data is collected, it is stored in a CSV file. The data pertaining to each subject corresponds to a row and their answers for each question are noted in the corresponding columns. Thus, the dataset has 31 columns i.e., one column for every question and one last column for the category they fall into. This file is then used to implement classifier algorithms in the machine learning and TOPSIS method in the Fuzzy systems.

In the machine learning module, the acquired data in the CSV file undergoes data processing, which includes checking for missing and erroneous values. As the responses chosen by the subject are categorical, the values are replaced with integers for the representation and processing by machine learning algorithms. Eighty percent of the data is used for training and the remaining 20% of the data is used as test data. After analyzing the training data, it is observed that the data is not balanced as few classes had fewer numbers of samples and the remaining classes had more numbers of samples. Therefore, the data is balanced using the K nearest neighbor algorithm. Some samples are chosen and their K nearest neighbors are found and the mean values of these samples are added to the training data. The first 31 columns are considered as the features and the last column is the target variable. A classifier is used to predict the category of the new instances. The algorithms used here are the D-Tree classifier, K-Nearest Neighbor (KNN) classifier, and the SVM as they are some of the most commonly used classifiers in the domain. The algorithms use the training data to learn. The prediction is performed on the test data. The models are evaluated by checking their accuracy scores. An automated system is also a part of the implementation which is used to take a new instance as input by entering options in the set of questions. This is fed to the classifier to predict the appropriate class of the new instance.

Similarly, the same CSV file is used to implement the TOPSIS method. For the implementation using the fuzzy TOPSIS method, the data in the original form is considered without converting the response (for example, less than a day, more than 7 days, 5–7 days, quite irregular, fairly irregular, sometimes, most of the times, etc.) into numerical form, where the response given by the subjects are uncertain or vague.

Figure 4 depicts the computational model to detect PCOS using physical parameters. As stated in the study design, in total 11 criteria from C11 to C111 are used as input in the computational model involving machine learning algorithms and multi-criteria decision systems such as fuzzy AHP followed by the fuzzy TOPSIS. The output of the computational model is discrete, viz, normal, moderate, and high PCOS problems. Similarly, **Figure 5** shows the detection of mental health disorders using anxiety and depression, social phobia, and body dissatisfaction aspects. As depicted in **Table 1** of the study design, C21–C45 are the criteria variables given as input to the computational models. In the same lines, the output is normal, moderate, and high mental health issues. The outcome of both physical aspects and mental health disorder are combined using fuzzy rules to provide final inference as abnormal having both PCOS and mental health,

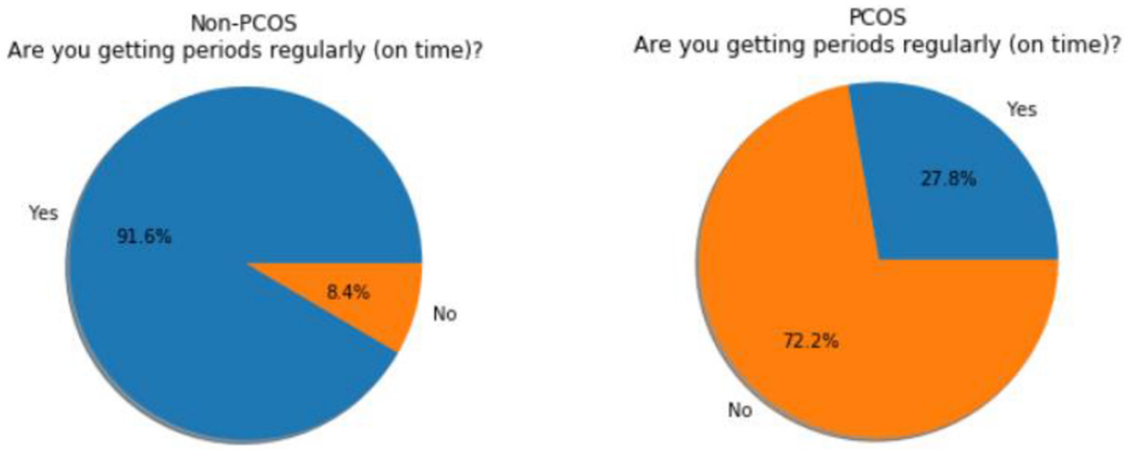


FIGURE 1 | Non- polycystic ovarian syndrome (PCOS) subjects vs. PCOS subjects (regularity in the cycle).

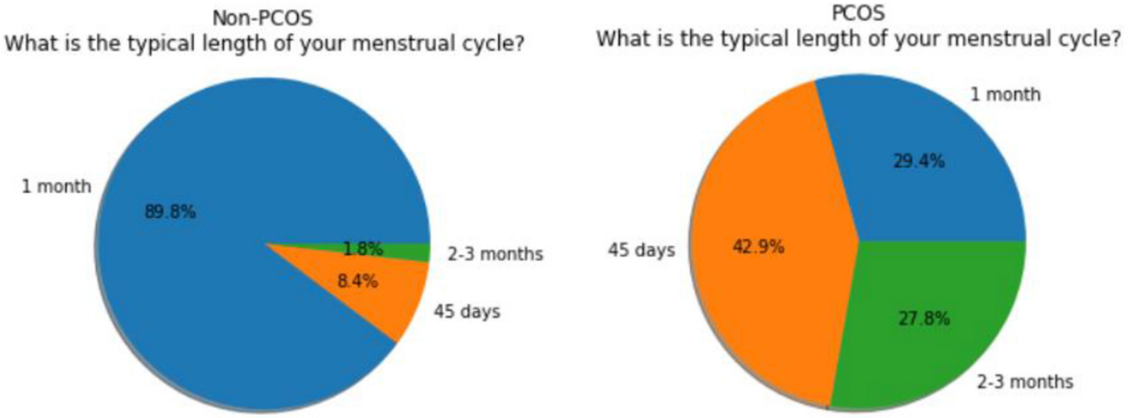


FIGURE 2 | Non-PCOS subjects vs. PCOS subjects (typical length of menstrual cycle).

abnormal having only PCOS, Abnormal having only mental health issues, or Normal.

Fuzzy AHP

In the proposed PCOS detection system, in order to quantify the weights in Fuzzy TOPSIS, Fuzzy AHP (30, 31) is used, which is depicted in **Figure 6**. The systematic steps are as follows:

Step 1: A criteria pair-wise comparison matrix M_{ij} is constructed. In total, 31 criteria are considered for the detection of PCOS and its associated mental health. M_{ij} is constructed by comparing the criteria in pairs to decide which among each pair is preferred or has a greater amount of quantitative property, or whether or not two of the criteria are identical. For this research work, the M_{ij} is constructed with the linguistic variables and its associated triangular fuzzy numbers as shown in **Table 2**.

Step 2: Construct the fuzzified geometric mean value. It is calculated as follows,

$$M_{ij} = (x_1 \times x_2 \times x_3 \times \ldots x_n)^{\frac{1}{n}} \times \left(y_1 \times y_2 \times y_3 \times \ldots y_n\right)^{\frac{1}{n}} \times (z_1 \times z_2 \times z_3 \times \ldots z_n)^{\frac{1}{n}} \quad (1)$$

where $x_1 \ldots x_n$ are the first elements in the fuzzy triangular numbers, $y_1 \ldots y_n$ are the second elements in the fuzzy triangular numbers, similarly, $z_1 \ldots z_n$are the third elements in the fuzzy triangular numbers, and n is the number of criteria, i.e., $n = 31$.

Step 3: Calculate the fuzzy weights ($\tilde{w}$) using the formula,

$$w_i = r_i \times (r_1 + r_2 + r_3 \ldots r_n)^{-1} \quad (2)$$

where r_1 is the addition of the first column of the fuzzified geometric mean value, similarly, r_2 and r_n is the addition

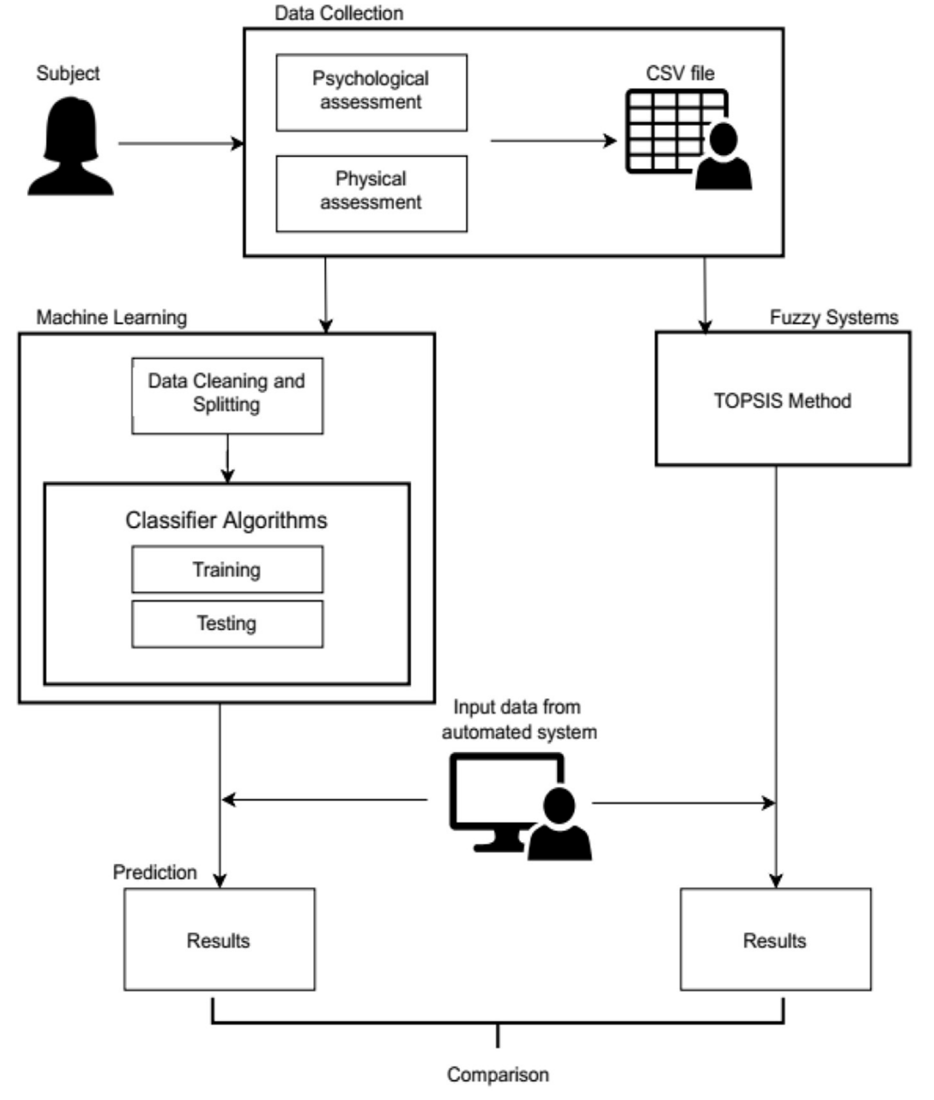


FIGURE 3 | Block diagram of the proposed method.

of second and nth columns of the fuzzified geometric mean value, respectively.

Step 4: Any defuzzification method can be used to calculate the defuzzified weights (w_i). For the proposed work, the Center of Area (COA) method is used.

$$COA = \frac{(a+b+c)}{3} \tag{3}$$

Where a, b, and c are the 3 sides of the triangle formed by the triangular membership function.

Step 5: From the weights (w_i), calculate normalized weights (w_j),

$$w_j = \frac{w_i}{\sum_{i=1}^{n} c_i} \tag{4}$$

These normalized weights are used to assign weights for the TOPSIS method.

Fuzzy TOPSIS

The Fuzzy TOPSIS method (32) is used which is described as follows:

Step 1: Initially the committee of decision-makers is formed. For this research work, the decision-makers are expert doctors in the gynecology department and the psychology department. The fuzzy rating of each decision-maker D_k $where(k = 1, 2, 3)$ can be expressed as a triangular fuzzy number $\tilde{p}_k$ where $(k = 1, 2, 3)$ with membership function $\mu \times \tilde{p}_k(x)$.

Step 2: Identify the key criteria and sub-criteria for evaluation.

Step 3: Select the suitable linguistic variables for evaluating the parameters such as criteria and alternatives. **Tables 3**, **4** show

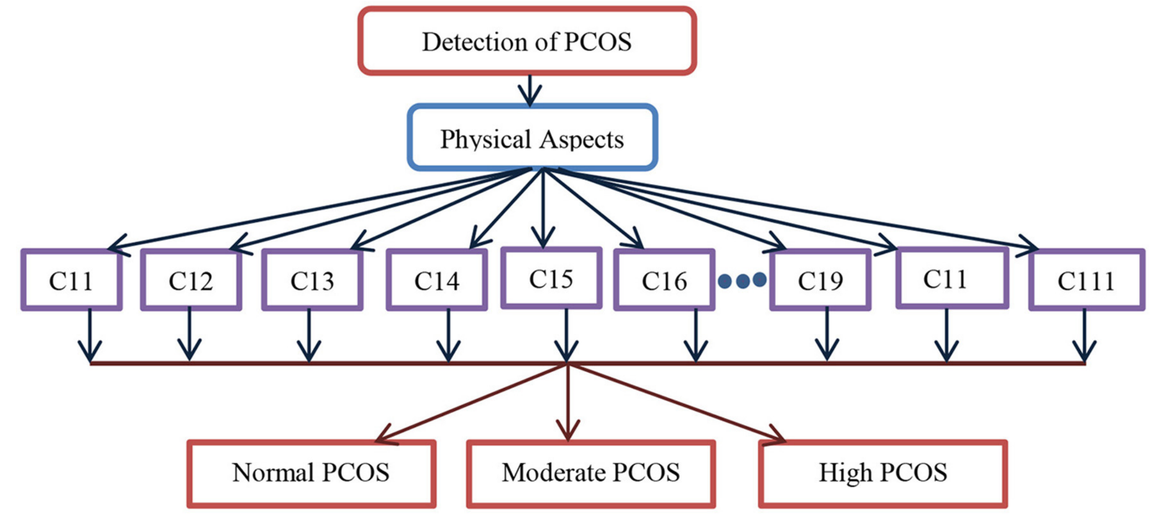


FIGURE 4 | The computational model for the detection of PCOS.

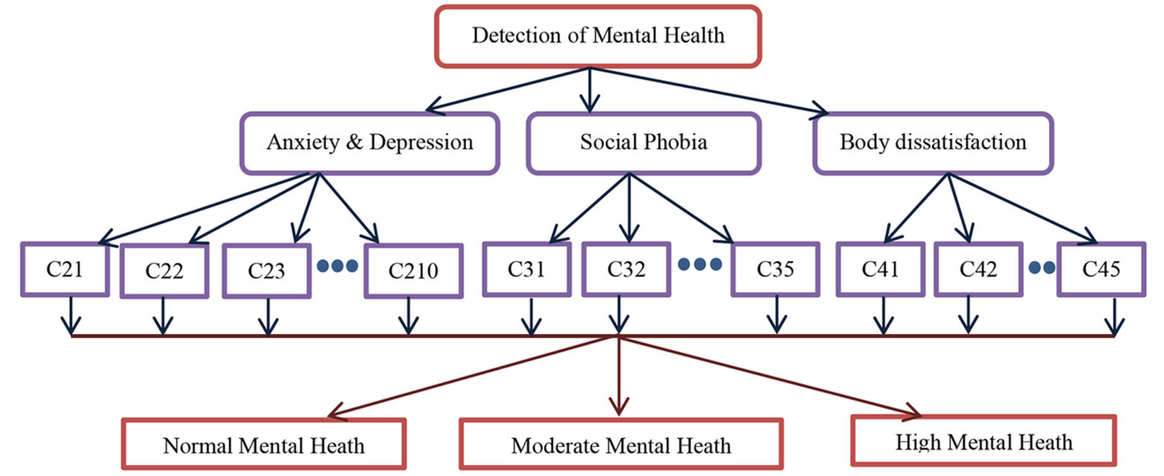


FIGURE 5 | The computational model for the detection of mental health.

the linguistic variables and their associated fuzzy numbers for each criterion and alternatives.

Step 4: Quantify the weights for each criterion based on the values obtained from the AHP method. Using triangular fuzzy numbers, the decision-makers ratings are described as follows, $\tilde{p}_k = (m_k, n_k, o_k)$, where $k = (1, 2, 3)$, then the total fuzzy rating can be determined as $\tilde{p} = (m, n, o)$ where,

$$m = \min_k \{m_k\}, \quad n = \frac{1}{k}\sum_{k=1}^{k} n_k, o = \max_k \{o_k\} \tag{5}$$

If the fuzzy ratings and the weights of kth decision-makers are

$$\tilde{y}_{ijk} = \left(m_{ijk}, n_{ijk}, o_{ijk}\right) \text{ and } \tilde{w}_{ijk} = \left(w_{jk1}, w_{jk2}, w_{jk3}\right) \tag{6}$$

where i = 1, 2, 3 … m and j = 1, 2, 3,……n., then the total fuzzy rating $(\tilde{y}_{ij})$ of alternative will be

$$\tilde{y}_{ij} = \left(m_{ij}, n_{ij}, o_{ij}\right)$$

where

$$m_{ij} = \min_k \left\{m_{ijk}\right\}, n_{ij} = \frac{1}{k}\sum_{k=1}^{k} n_{ijk}, o_{ij} = \max_k \left\{o_{ijk}\right\} \tag{7}$$

$\tilde{w}_{ij}$ is the aggregated fuzzy weights of each criterion and is given by $\tilde{w}_j = \left(w_{j1}, w_{j2}, w_{j3}\right)$ where

$$w_{j1} = \min_k \left\{w_{jk1}\right\}, w_{j2} = \frac{1}{k}\sum_{k=1}^{k} w_{jk2}, w_{j3} = \max_k \left\{w_{jk3}\right\} \tag{8}$$

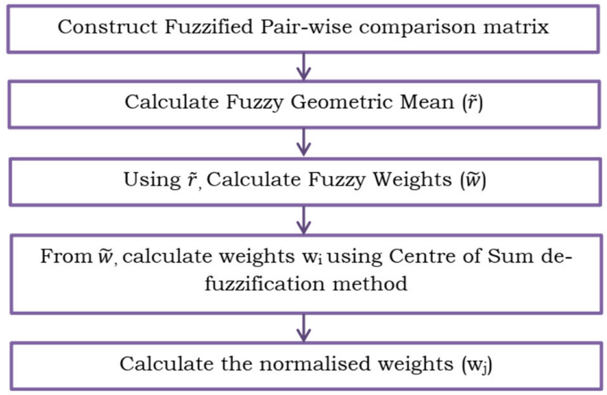


FIGURE 6 | Fuzzy AHP.

Step 5: The fuzzy decision matrix is constructed as

$$\tilde{D} = \begin{matrix} \tilde{y}_{11} & \tilde{y}_{12} & \tilde{y}_{1n} \\ \tilde{y}_{21} & \tilde{y}_{22} & \tilde{y}_{2n} \\ \tilde{y}_{m1} & \tilde{y}_{m2} & \tilde{y}_{mn} \end{matrix}$$

$\tilde{w} = \left[\tilde{w}_1, \tilde{w}_2, \ldots \tilde{w}_n\right]$. Here, $\tilde{y}_{ij}$ and $\tilde{w}_j$ can be approximated by the positive triangular fuzzy numbers.

Step 6: Using linear scale transformation, the values of the criteria are transformed from a criteria scale into a comparable scale as

$$\begin{aligned} \tilde{A} &= \left[\tilde{a}_{ij}\right]_{m\times n} \text{ where} \\ \tilde{a}_{ij} &= \left(\frac{m_{ij}}{c_j^*}, \frac{n_{ij}}{c_j^*}, \frac{o_{ij}}{c_j^*}\right), c_j^* = \max_i\left(c_{ij}\right) \end{aligned} \tag{9}$$

Step 7: The normalized weighted decision matrix is calculated. The weighted normalized decision matrix ($\tilde{v}$) is calculated as

$$\begin{aligned} \tilde{v} &= \left[\tilde{a}_{ij}\right]_{m\times n} \\ \text{where, } \tilde{a}_{ij} &= \tilde{a}_{ij} \times \tilde{w}_j \end{aligned} \tag{10}$$

Step 8: Then the fuzzy positive ideal solution (FPIS,B^*) and fuzzy negative ideal solution (FNIS, B^-) are calculated as

$$\begin{aligned} B^* &= \left(\tilde{a}_1^*, \tilde{a}_2^*, \ldots \tilde{a}_n^*\right) \\ B^- &= \left(\tilde{a}_1^-, \tilde{a}_2^-, \ldots \tilde{a}_n^-\right) \\ \text{where } \tilde{a}_j^* &= \max_i\left\{v_{ij3}\right\} \\ \text{where } \tilde{a}_j^- &= \min_i\left\{v_{ij1}\right\} \end{aligned} \tag{11}$$

Step 9: The distance from the alternative to the FPIS and FNIS as

$$d_i^* = \sum_{j=1}^{n}\left(\tilde{a}_{ij}, \tilde{a}_j^*\right) \times d_v \text{ where } i = 1, 2, \ldots m \tag{12}$$

TABLE 2 | Linguistic variables and their fuzzy numbers.

Sl. No	Linguistic variables	Fuzzy triangular numbers
1.	Equal important	(1, 1, 1)–1
2.	Moderate important	(2, 3, 4)–3
3.	Strong important	(4, 5, 6)–5
4.	Very strong important	(6, 7, 8)–7
5.	Extreme important	(9, 9, 9)–9
6.	Intermediate	2–(1, 2, 3), 4–(3, 4, 5), 6–(5, 6, 7), 8–(7, 8, 9)

TABLE 3 | Linguistic variables for the importance weight of each criterion.

Sl. No	Linguistic variables	Triangular fuzzy numbers
1.	Very Less Relevant (VLR)	(0, 0, 2)
2.	Less Relevant (LR)	(2, 3, 4)
3.	Relevant (R)	(4, 5, 6)
4.	High Relevant (HR)	(6, 7, 8)
5.	Very High Relevant (VHR)	(8, 9, 10)

TABLE 4 | Linguistic variables for alternatives.

Sl. No	Linguistic variables	Triangular fuzzy numbers
1.	Very Low (VL)	(0, 0, 0.2)
2.	Low (L)	(0.20, 0.3, 0.4)
3.	Medium (M)	(0.4, 0.5, 0.6)
4.	High (H)	(0.6, 0.7, 0.8)
5.	Very High (VH)	(0.8, 0.9, 1)

$$d_i^- = \sum_{j=1}^{n}\left(\tilde{a}_{ij}, \tilde{a}_j^-\right) \times d_v \text{ where } i = 1, 2, m \tag{13}$$

where d_v is the distance between 2 fuzzy numbers.

Step 10: The closeness coefficient (CC_i) is calculated as

$$CC_i = \frac{d_i^-}{d_i^* + d_i^-} \tag{14}$$

Step 11: The physical and mental health are obtained in the decreasing order of the closeness coefficient values. The entire flowchart of the TOPSIS method is shown in **Figure 7**.

RESULTS

The proposed methodology using Fuzzy AHP and Fuzzy TOPSIS is implemented in python (Python Software Foundation. Delaware, United States). For quantifying the weights in fuzzy TOPSIS, fuzzy AHP is used. The normalized weights from the fuzzy AHP are shown in **Tables 5A–C** for a test example. As seen from **Table 5A**, it is observed that the weights C11, C12, C13, C14, and C16 have higher values compared with C15, C17, C18, C19, C110, and C111. Higher values for the weights indicate the relatively higher importance of those criteria. From **Table 5B**, it can be seen that C21, C22, C23, C25, and C27 have higher importance since the weights have higher values. In the same lines, as observed from **Table 5C**, C41 and C42 have higher weights and hence, these criteria have more importance compared with other criteria. Hence it is validated that, C11, C12, C13, C14, C16, C21, C22, C23, C25, C27, C41, and C42 are more important criteria in comparison with other criteria in making decision as given by fuzzy AHP. In the same way, the clinicians were also asked to provide the relative importance of the 31 criteria based on their clinical acumen. Their expert opinion also exactly matched with the outcome of the fuzzy AHP. Hence these weights are used for assigning the values in the fuzzy TOPSIS.

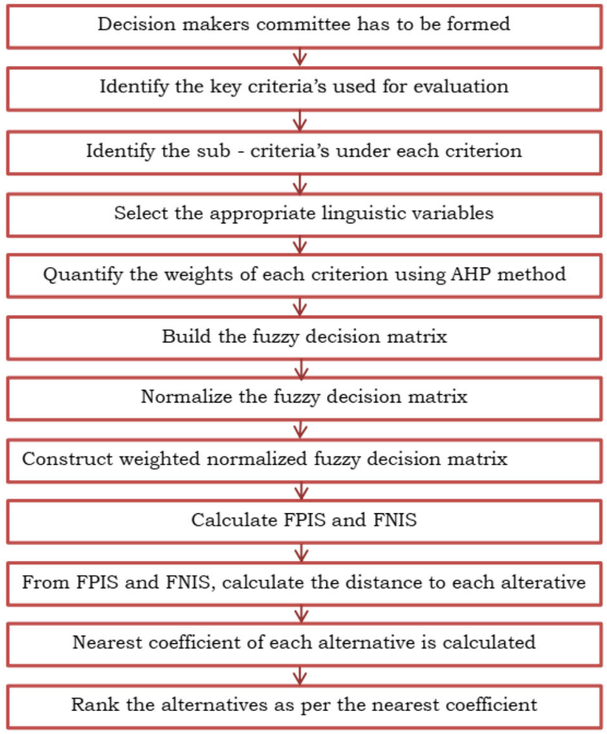


FIGURE 7 | The flow of the Fuzzy TOPSIS method.

From equations 12 and 13, the distance between FPIS and FNIS for different criteria is computed and is given in **Tables 6–8**. The values obtained in these tables are used to compute the closeness coefficients. **Table 9A** provides the closeness coefficients between FPIS and FNIS for PCOS. A higher closeness coefficient value indicates a high probability of having PCOS. It can be seen that CC_1 is higher compared with CC_3 and CC_2. Hence, it is observed that P1 (high PCOS) > P3 (moderate PCOS) > P2 (normal PCOS), therefore it is inferred that the possible condition is high PCOS for this particular test example. Similarly, **Table 9B**, shows the closeness coefficients between FPIS and FNIS for mental health. It is observed that M1 (high mental health) > M3 (moderate mental health) > M2 (normal mental health), and therefore the inference is high mental health issues for the same test example.

Based on the individual (both physical and psychological) closeness coefficients, the fuzzy rules are framed in consultation with the clinicians. The rules are given in **Table 10**.

Similarly, machine learning algorithms, D-tree, KNN, and SVM algorithms are used to predict PCOS and its associated mental health issues independently and the final inference is drawn as per the rules provided in **Table 9**. In this process, the entire dataset is divided into a training set (80%) and testing set (20%).

The confusion matrix due to the classification using SVM and fuzzy TOPSIS is provided in **Table 11** respectively. It can be noted that the Fuzzy TOPSIS method provides the highest performance compared with all other classifiers.

TABLE 5A | Weights for the criteria from C11 to C111.

C11	C12	C13	C14	C15	C16	C17	C18	C19	C110	C111
0.074	0.074	0.074	0.074	0.038	0.068	0.036	0.036	0.036	0.023	0.023

TABLE 5B | Weights for the criteria from C21 to C210.

C21	C22	C23	C24	C25	C26	C27	C28	C29	C210
0.045	0.045	0.045	0.023	0.029	0.04	0.021	0.012	0.012	0.012

TABLE 5C | Weights for the criteria from C31 to C35 and C41 to C45.

C31	C32	C33	C34	C35	C41	C42	C43	C44	C45
0.017	0.017	0.017	0.017	0.017	0.019	0.019	0.014	0.014	0.006

TABLE 6 | Distance between P^* and P_i (i = 1, 2, 3), P^- and P_i (i = 1, 2, 3) for the criteria from C11 to C110.

d(P$_i$, P*)	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	**C17**	**C18**	**C19**	**C110**	**C111**
$d(P, P^*)$	2.35	2.35	3.34	2.35	2.95	4.24	2.95	2.3	2.3	3.1	3.1
$d(P_2, P^*)$	9.38	9.38	8.54	9.38	6.6	8.27	6.6	5.4	5.4	4.08	4.08
$d(P_3, P^*)$	4.8	4.8	6.43	4.97	4.08	5.46	4.08	3.38	3.38	2.42	2.42
$d(P_i, P^-)$	C11	C12	C13	C14	C15	C16	C17	C18	C19	C110	C111
$d(P_1, P)$	8.3	8.3	7.72	8.3	5.94	7.17	5.94	4.91	4.91	2.72	2.72
$d(P_2, P^-)$	1.15	1.15	2.37	1.15	2.02	2.5	2.02	1.57	1.57	1.73	1.73
$d(P_3, P^-)$	5.96	5.96	4.22	5.79	4.57	5.55	4.57	4.34	4.34	3.6	3.6

TABLE 7 | Distance between M^* and M_i (i = 1, 2, 3), M^- and M_i (i = 1, 2, 3) for the criteria from C21 to C210.

d(M$_i$, M)	**C21**	**C22**	**C23**	**C24**	**C25**	**C26**	**C27**	**C28**	**C29**	**C210**
$d(M_1, M^*)$	2.35	3.33	3.33	2.09	3.34	3.33	2.84	2.09	3.24	3.07
$d(M_2, M)$	9.38	9.38	9.38	7.5	8.54	9.38	6.84	7.5	7.5	7.5
$d(M_3, M^*)$	4.8	5.3	5.3	3.95	4.97	5.3	4.08	3.95	4.59	4.46
$d(M_i, M^-)$	C21	C22	C23	C24	C25	C26	C27	C28	C29	C210
$d(M_1, M^-)$	8.3	7.73	7.73	6.5	7.72	7.73	6.07	6.5	5.78	5.96
$d(M_2, M^-)$	1.15	1.15	1.15	0.92	2.37	1.15	1.89	0.92	0.92	0.92
$d(M_3, M^-)$	5.96	5.7	5.7	4.7	5.79	5.7	4.57	4.7	4.33	4.45

TABLE 8 | Distance between M^* and M_i (i = 1, 2, 3), M^- and M_i (i = 1, 2, 3) for the criteria from C31 to C35 and C41. C45.

d(M$_i$, M*)	**C31**	**C32**	**C33**	**C34**	**C35**	**C41**	**C42**	**C43**	**C44**	**C45**
$d(M_1, M^*)$	2.84	2.84	2.84	2.84	2.84	2.84	3.49	3.22	2.6	2.82
$d(M_2, M^*)$	6.84	6.84	6.84	6.84	6.84	8.54	8.54	6.3	4.75	1.89
$d(M_3, M^*)$	3.95	3.95	3.95	3.95	3.95	4.8	4.8	3.38	2.71	2.76
$d(M_i, M^-)$	C31	C32	C33	C34	C35	C41	C42	C43	C44	C45
$d(M_1, M^-)$	6.07	6.07	6.07	6.07	6.07	7.55	7.55	5.88	4.31	1.23
$d(M_2, M^-)$	1.89	1.89	1.89	1.89	1.89	2.37	2.37	2.52	1.88	2.4
$d(M_3, M^-)$	4.7	4.7	4.7	4.7	4.7	5.96	5.96	5.71	4.19	1.26

TABLE 9A | Closeness coefficient for PCOS for the alternatives *P1*, *P2*, and *P3*.

CC$_i$ for PCOS	**Closeness coefficient**	**Alternative**
CC1	*0.68*	*P1*
CC2	*0.53*	*P3*
CC3	*0.2*	*P2*

TABLE 9B | Closeness coefficient for mental health for the alternatives *M1*, *M2*, and *M3*.

CC$_i$ for mental health	**Closeness coefficient**	**Alternative**
CC1	*0.68*	*M1*
CC2	*0.54*	*M3*
CC3	*0.19*	*M2*

TABLE 10 | Fuzzy inference rules.

Sl. No	**Fuzzy rule base for inference**
1.	If P1 & M1 then A1 (having high PCOS and high Mental health)
2.	If P1 & M2 then A2 (having high PCOS and normal Mental health)
3.	If P1 & M3 then A1 (having high PCOS and moderate Mental health)
4.	If P2 & M1 then A3 (having normal PCOS and high Mental health)
5.	If P2 & M2 then A4 (having normal PCOS and normal Mental health)
6.	If P2 & M3 then A3 (having normal PCOS and moderate Mental health)
7.	If P3 & M1 then A1 (having moderate PCOS and high Mental health)
8.	If P3 & M2 then A2 (having moderate PCOS and normal Mental health)
9.	If P3 & M3 then A1 (having moderate PCOS and moderate Mental health)

The number of classes defines the size of the confusion matrix. In the current study, there are four classes in the data and hence, the confusion matrix is 4 × 4. The true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values are evaluated as follows.

TP of current class = Principal diagonal values of current row/column.

TABLE 11 | Confusion matrix during support vector machines (SVM) classification (left) and fuzzy technique for order of preference by similarity to ideal solution (TOPSIS) (right).

	Predicted class		Predicted class
Actual class	A1 A2 A3 A4	Actual class	A1 A2 A3 A4
	A1[[37 0 0 0]		A1[[37 0 0 0]
	A2 [0 24 0 0]		A2 [0 24 0 0]
	A3 [1 0 38 4]		A3 [0 0 41 2]
	A4 [0 3 2 58]]		A4 [0 0 1 62]]

TABLE 12 | Components of confusion matrix for SVM.

Class	A1	A2	A3	A4
True positive	37	24	38	58
False positive	0	0	5	5
False negative	1	3	2	4

TABLE 13 | Components of confusion matrix for fuzzy TOPSIS.

Class	A1	A2	A3	A4
True positive	37	24	41	62
False positive	0	0	2	1
False negative	0	0	1	2

FP of current class = (Sum of current row values–TP of current class).

FN of current class = (Sum of current column values–TP of current class).

TN of current class = Sum of the matrix–TP of current class–Sum of current row values–Sum of current column values.

Tables 12, **13** present the TP, FP, and FN for SVM and the fuzzy TOPSIS classifier, respectively.

It can be seen from **Table 12** that there are 10 FP and 10 FN during SVM classification, whereas during the fuzzy TOPSIS method, as indicated in **Table 13**, there are only three FP and three FN. It is noted that fuzzy TOPSIS provides less FN as compared with other classifiers. The presence of FN is crucial. It means that the woman actually with PCOS are diagnosed as not having PCOS by the automated system. It also indicates that the particular woman actually with PCOS goes undiagnosed, which is a fatal error. It can be inferred that Fuzzy TOPSIS provides less FN and hence, it is superior compared with all other classifiers. On the other hand, fuzzy TOPSIS has provided less FP compared with other classifiers. False-positive means a woman does not actually have PCOS but the automated system diagnoses her with PCOS. This causes a lot of undue panic and anxiety to the patients. Hence the fuzzy TOPSIS is superior in terms of less FP as well.

The accuracy of classification using KNN, D-tree, SVM, and fuzzy TOPSIS is shown in **Figure 8**. It can be noted that the SVM provides an accuracy of 94.01%, and D-tree and KNN provide 88.02 and 87.42%, respectively. It is observed that SVM outperforms the D-tree and KNN. The fuzzy TOPSIS provides an accuracy of 98.20%, outperforming all other classifiers.

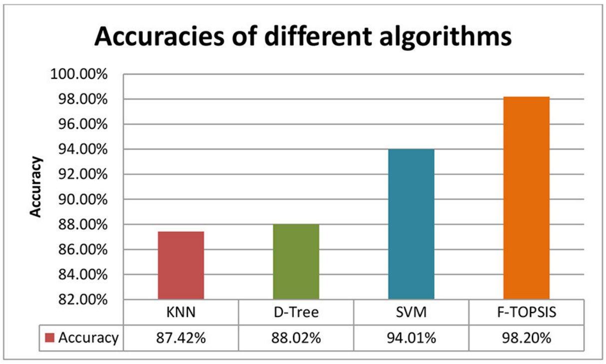


FIGURE 8 | Model evaluation: accuracies of different classifier.

TABLE 14 | The proportion of women having PCOS and mental health issues in the testing data.

#Testing data	PCOS		Mental health	
	P1 (High possibility)	P3 (Moderate)	M1 (High Possibility)	M3 (Moderate)
167	52	4	83	1

It is also interesting to assess the mental health issues of a woman who was diagnosed with PCOS. **Table 14** provides the proportion of women with PCOS in the testing data, as well as the proportion of women with mental health issues in the test data. It can be seen that the fuzzy TOPSIS provided that 52 women have a high possibility of PCOS and four women have a moderate possibility of PCOS. So in total, 56 women have PCOS.

From this data, the mental wellness indicator (I_M) of PCOS patients is defined as,

$$I_M = \frac{n\,\{(P1\ \vee P3)\ \wedge\ (A1)\}}{n\,\{P1 \vee P3\}} \times 100$$

where '*n*' represents the count meeting the condition.

Similarly, the physical wellness indicator (I_P) of a patient suffering from mental illness is defined as,

$$I_P = \frac{n\,\{(M1\ \vee M3)\ \wedge\ (A1)\}}{n\,\{M1 \vee M3\}} \times 100$$

Figure 9 shows the mental wellness indicator (I_M) which resulted in a value of 66.07%, which means the majority of women suffering from PCOS also suffer from mental health issues. This shows that there is a strong association between PCOS and the mental well-being of a woman which needs to be taken care of during the clinical investigation and diagnosis.

Figure 10 depicts the physical wellness indicator (I_P) which has a value of 44%, suggesting that a considerable proportion of women suffering from mental health also have PCOS. This information can be effectively used during clinical evaluation and appropriate diagnosis can be established.

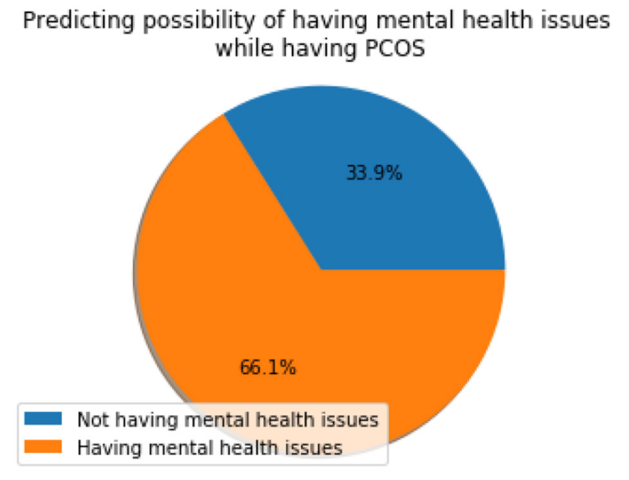


FIGURE 9 | Mental wellness indicator (I_M).

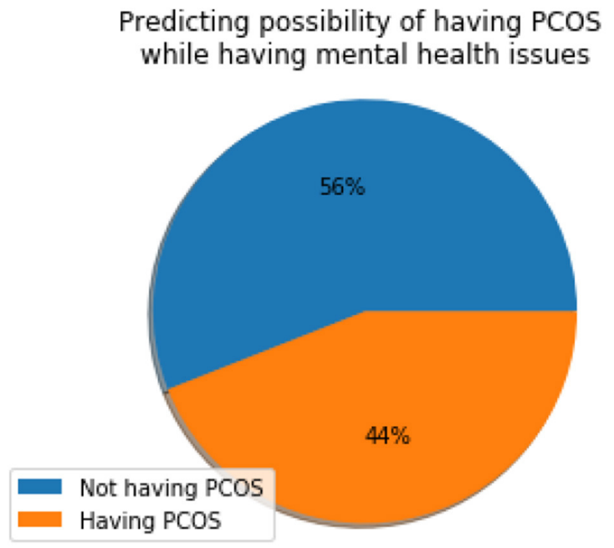


FIGURE 10 | Physical wellness Indicator (I_P).

DISCUSSION

The present study investigates the early prediction of PCOS and its associated mental health issues more objectively by incorporating inherent fuzziness in the decision process. The fuzzy algorithm TOPSIS outperformed the conventional methods. Further, the mental wellness parameters, such as anxiety, depression, social phobia, and body image dissatisfaction, are integrated with clinical parameters to assess the women for the prognosis and diagnosis of PCOS and its associated socio-psychological well-being as 66.07% of women with PCOS have associated mental health issue as per the results of this study.

The computer diagnosis/automation of the early prediction of PCOS can bring a paradigm change in the conventional practices in the management of PCOS. The incorporation of the computer algorithm in the diagnosis can speed up the process and it can help the doctors in delivering the diagnosis quickly and easily. The developed computer diagnosis methodology is faster and more efficient; more women can be screened for PCOS in a short periodof time. Hence this methodology can offer a low-cost and mass screening protocol to detect PCOS. This in turn can help women of low socio-economic conditions by providing a quick, efficient, and low-budget solution.

The regular screening of women for PCOS involves personal interaction, clinical examination, ultra-sound scans, etc. The developed methodology in this study involves a pre-designed questionnaire presented by the computer to the women, which comprises both physical and psychological health questions where the privacy of the women is taken care of. Women can feel free while answering the questions with the computer in their own comfort zone without invading their privacy.

It is interesting to compare our results with previously published research as shown in **Table 15**. Denny et al. (24), in his study, used multiple parameters such as ultrasound result (size of follicles, number of follicles), blood investigations (TSH, AMH, vitamin D3, etc.), and clinical features (cycle length and regularity), and obtained an accuracy of 89.02% using the random forest classifier. Since hormonal tests are involved in the data collection, the process becomes expensive in terms of cost. Further, Vikas et al. (25), in their study, considered parameters such as lifestyle and food intake habits, and psychological parameters like anxiety and depression, and obtained an accuracy of 97.65% using the Naïve Bayes classifier, however, the questionnaire involved binary (yes and no) responses. Using ultrasound scanned images, Anuradha and Priyanka (26) and Deshpande and Wakankar (27) investigated PCOS and obtained an accuracy of 98% using Artificial Neural Networks and 95% using SVM, respectively. Further, Meena et al. (33), using endometrial biopsies, obtained an accuracy of 83.70% using Artificial Neural Networks. However, in these studies, mental health parameters are not considered. Further, Satish et al. (34), using clinical and biochemical parameters such as BMI, pulse rate, hemoglobin, and hormonal tests (which include FSH, Prolactin, Progesterone), obtained an accuracy of 87.72% using the Naïve Bayes classifier, yet even in this study, mental health parameters are not considered. All these above studies involve complex and expensive parameters such as hormonal investigations, blood investigations, and ultrasound scanned images which are difficult to acquire, take a longer time, and are expensive, and hence, women with low socioeconomic status cannot afford them.

Given the limitations of the above studies, this research involves simple regular parameters (physical and mental health), yet an improved performance of 94.01 and 98.20% with SVM and TOPSIS, respectively, are obtained. The present study provides higher performance than the existing standard and widely used methods. It also provides a new paradigm of PCOS diagnosis using computational intelligence by using the fuzzy

TABLE 15 | Comparison of results with the previously published research.

References	Dataset used	No of attributes	Parameters considered	Algorithms used	Performance
Denny et al. (24)	541 women	23	Clinical and metabolic parameters such as no. of follicles, size of follicles, TSH, AMH, Vit D3, cycle length & regularity etc.	6 classification algorithms: CART, SVM, KNN, Logistic regression, Naïve Bayes, Random Forest	Accuracy of 89.02% by Random Forest classifier
Vikas et al. (25)	119 women	18	Life style and food intake habits such as regularity of cycle, anxiety & depression, mental stress	ANN, Naïve Bayes, Decision Tree	Accuracy of 97.65% by Naïve Bayes, 96.27% by ANN, 96.24% by Decision Tree
Anuradha and Priyanka (26)	84 women	13	Acne, irregular periods, sonography, LH & weight	ANN, KNN, & Linear regression	Accuracy of 94% by ANN
Deshpande and Wakankar (27)	20 women	5	clinical (BMI and cycle length), biochemical (FSH and LH levels) and imaging (calculating number of follicles present in the ovary).	SVM	Accuracy of 95% by SVM
Meena et al. (33)	303 women	26	Endometrial biopsies	Decision Tree, Naïve Bayes, SVM, ANN	Accuracy of 76.45% by SVM, 83.70% by ANN, 75.25% by D-Tree, 82.75% by Naïve Bayes
Satish et al. (34)	541 women	41	Clinical & biochemical parameters such as BMI, pulse rate, hemoglobin, hormonal tests include FSH, Prolactin, Progesterone	KNN, SVM, RF, GNB, ANN	Accuracy of 75.45% by KNN, 82.27% by SVM, 85% by RF, 87.72% by Naïve Bayes, 50% by ANN
Proposed work	629 women	33	Physical parameters (length and duration of the cycle etc.) and psychological parameters (anxiety, depression, body dissatisfaction and social phobia)	Fuzzy TOPSIS,SVM, KNN, Decision Tree	Accuracy of 98.20% by Fuzzy TOPSIS, 94.01% by SVM, 88.02% by D-Tree, 87.42% by KNN

TOPSIS method. In addition, the method also provides improved performance when compared with conventional methods. This methodology can also be used for the mass screening of the population by developing countries with large populations and constrained resources. Women can undergo this methodology in their own comfort zone without the intervention of a physician guarding their privacy.

The nature of data comprising of physical and mental health parameters is inherently imprecise, vague, and perceptually subjective. The traditional data analysis methods such as SVM, KNN, and the D-tree can model the crisp data accurately. When the data itself is imprecise, the crisp models cannot address the intricate fuzzy interrelations in the data. Hence the fuzzy set-based models can address the inherent subjectivity in the data and can provide improved performance. The fuzzy TOPSIS can be used to find out the most suitable alternative concerning different selection criteria. Also, the fuzzy distances used in fuzzy TOPSIS provide improved closeness coefficient values compared with other fuzzy measures. Therefore, the selected fuzzy TOPSIS has a higher potential in representing the physical phenomena of the scenario.

In addition to the above contributions, this research work also demonstrates the importance of the mental well-being of a woman during PCOS. The result indicates that the majority of women suffering from PCOS also undergo mental health issues. The outcome of this study recommends having mental health evaluations during the diagnosis of PCOS.

CONCLUSION

The strength of the proposed automated system involves the inclusion of PCOS and mental health questionnaires which have been validated by expert gynecologists and psychiatrists. Considering anxiety, depression, social phobia, and body image dissatisfaction analysis along with physical aspects adds new knowledge to the literature. In this study, the multi-criteria decision analysis system is developed for the early prediction of PCOS problems and mental health before the progression of the disease which can reduce morbidity. The proposed fuzzy TOPSIS-based automated system provides a high performance of 98.20% in real-time scenarios. The inherent uncertainty and imprecision involved in the data and decision-making process are well-modeled by Fuzzy TOPSIS in comparison with other methods which is evident with the high performance obtained. In addition to methodological contributions and improved performance, the study also indicates that the majority of women suffering from PCOS also have associated mental health issues. The present study proved that 66.07% of women with PCOS have associated mental health issues. This strong association implies that there is a strong need for improved protocols and efficient stress and mental health balancing guidelines in the management of PCOS.

The developed system in this study can help physicians in their clinical practice by acting as an adjunct tool. The proposed method is cost-effective and hence, women with low socio-economic conditions can also avail the screening. As the screening is automated and does not involve the intervention of clinicians, women can use the tool in their own comfort level with privacy. The use of the automated algorithm integrated with minimal intervention from physicians can speed up the screening process, making the screening available for more women in a given time and can bring a change in the conventional practices.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Ekramzadeh M, Hajivandi L, Noroozi M, Mostafavi F. Psychological experiences of adolescent girls with polycystic ovary syndrome: a qualitative study. *Iran J Nurs Midwifery Res*. (2020) 25:341–7. doi: 10.4103/ijnmr.IJNMR_276_19
2. Naqvi SMAS, Bhattarai JB, Hui L, Wang XW. Polycystic ovarian syndrome and female infertility. *Yangtze Med*. (2020) 4:11–27. doi: 10.4236/ym.2020.41002
3. Franks S. Polycystic ovary syndrome. *N Engl J Med*. (1995) 333:853–61. doi: 10.1056/NEJM199509283331307
4. Kiddy DS, Sharp PS, White DM, Scanlon MF, Mason HD, Bray CS, et al. Differences in clinical and endocrine features between obese and non-obese subjects with polycystic ovary syndrome: an analysis of 263 consecutive cases. *Clin Endocrinol*. (1990) 32:213–20. doi: 10.1111/j.1365-2265.1990.tb00857.x
5. Lim SS, Clifton PM, Noakes M, Norman RJ. Obesity management in women with polycystic ovary syndrome. *Women Health*. (2007) 3:73–86. doi: 10.2217/17455057.3.1.73
6. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. (2004) 81:19–25. doi: 10.1016/j.fertnstert.2003.10.004
7. Lauritsen MP, Bentzen JG, Pinborg A, Loft A, Forman JL, Thuesen LL, et al. The prevalence of polycystic ovary syndrome in a normal population according to the Rotterdam criteria versus revised criteria including anti-mullerian hormone. *Hum Reprod*. (2014) 29:791–801. doi: 10.1093/humrep/det469
8. Nidhi R, Padmalatha V, Nagarathna R, Amritanshu R. Prevalence of polycystic ovarian syndrome in Indian adolescents. *J Pediatr Adolesc Gynecol*. (2011) 24:223–7. doi: 10.1016/j.jpag.2011.03.002
9. Hussain A, Chandel RK, Ganie MA, Dar MA, Rather YH, Wani ZA, et al. Prevalence of psychiatric disorders in patients with a diagnosis of polycystic ovary syndrome in Kashmir. *Indian J Psychol Med*. (2015) 37:66–70. doi: 10.4103/0253-7176.150822
10. Upadhyaya SK, Sharma A, Agrawal A. Prevalence of anxiety and depression in polycystic ovarian syndrome. *Int J Med Sci Public Health*. (2016) 5:681–3. doi: 10.5455/ijmsph.2016.30072015110
11. Fauser BC, Tarlatzis BC, Rebar RW, Legro RS, Balen AH, Lobo R, et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril*. (2012) 97:28–38. doi: 10.1016/j.fertnstert.2011.09.024
12. Himelein MJ, Thatcher SS. Polycystic ovary syndrome and mental health: a review. *Obstet Gynecol Surv*. (2006) 61:723–32. doi: 10.1097/01.ogx.0000243772.33357.84
13. Conte F, Banting L, Teede HJ, Stepto NK. Mental health and physical activity in women with polycystic ovary syndrome: a brief review. *Sports Med*. (2015) 45:497–504. doi: 10.1007/s40279-014-0291-6
14. Kerchner A, Lester W, Stuart SP, Dokras A. Risk of depression and other mental health disorders in women with polycystic ovary syndrome: a longitudinal study. *Fertil Steril*. (2009) 91:207–12. doi: 10.1016/j.fertnstert.2007.11.022
15. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA*. (1999) 282:1737–44. doi: 10.1001/jama.282.18.1737
16. Beck AT,Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. (1961) 6:561–71. doi: 10.1001/archpsyc.1961.01710120031004
17. Beck Anxiety Inventory. *ACP PIER: The Physicians' Information and Education Resource Evidence Ratings and Criteria*. Philadelphia, PA: American College of Physicians (2007).
18. Wolf WM, Wattick RA, Kinkade ON, Olfert MD. Geographical prevalence of polycystic ovary syndrome as determined by region and race/ethnicity. *Int J Environ Res Public Health*. (2018) 15:2589–602. doi: 10.3390/ijerph15112589
19. Naz MSG, Tehrani FR, Majd HA, Ahmadi F, Ozgoli G, Fakari FR, et al. The prevalence of polycystic ovary syndrome in adolescents: a systematic review and meta-analysis. *Int J Reprod Biomed*. (2019) 17:533–42. doi: 10.18502/ijrm.v17i8.4818
20. Banting LK, Gibson-Helm M, Polman R, Teede HJ, Stepto NK. Physical activity and mental health in women with polycystic ovary syndrome. *BMC Women's Health*. (2014) 14:51–9. doi: 10.1186/1472-6874-14-51
21. Berni TR, Morgan CL, Berni ER, Rees DA. Polycystic ovary syndrome is associated with adverse mental health and neurodevelopmental outcomes. *J Clin Endocrinol Metab*. (2018) 103:2116–25. doi: 10.1210/jc.2017-02667
22. Chaudhari AP, Mazumdar K, Mehta PD. Anxiety, depression, and quality of life in women with polycystic ovarian syndrome. *Indian J Psychol Med*. (2018) 40:239–46. doi: 10.4103/IJPSYM.IJPSYM_561_17
23. Sachdeva G, Gainder S, Suri V, Sachdeva N, Chopra S. Comparison of the different PCOS phenotypes based on clinical metabolic, and hormonal profile, and their response to clomiphene. *Indian J Endocrinol Metab*. (2020) 23:326–31. doi: 10.4103/ijem.IJEM_30_19
24. Denny A, Raj A, Ashok A, Ram CM, George R. i-HOPE: detection and prediction system for Polycystic Ovary Syndrome (PCOS) using machine learning techniques. In *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. Kochi (2019).
25. Vikas B, Anuhya BS, Chilla M, Sarangi S. A critical study of Polycystic Ovarian Syndrome (PCOS) classification techniques. *Int J Comput Eng Manag*. (2018) 4:1–7.
26. Anuradha DT, Priyanka RL. Genetic clustering for polycystic ovary syndrome detection in women of reproductive age. *Int J Eng Adv Technol*. (2020) 3:1359–61. doi: 10.35940/ijeat.C5457.029320
27. Deshpande SS, Wakankar A. Automated detection of polycystic ovarian syndrome using follicle recognition. In *IEEE International Conference on Advanced Communications, Control and Computing Technologies*. Ramanathapuram (2014).
28. Mishra S, Prakash M. Study of fuzzy logic in medical data analytics. *Int J Pure Appl Math*. (2018) 119:16321–42.
29. Ansari MTJ, Al-Zahrani FA, Pandey D, Agrawal A. A fuzzy TOPSIS based analysis toward selection of effective security requirements engineering approach for trustworthy healthcare software development. *BMC Med Inform Decis Mak*. (2020) 20:236. doi: 10.1186/s12911-020-01209-8
30. Buckley JJ. Fuzzy hierarchical analysis. *Fuzzy Sets Syst*. (1985) 17:233–47. doi: 10.1016/0165-0114(85)90090-9
31. Carlsson C, Full'er R. Fuzzy multiple criteria decision making: Recent developments. *Fuzzy Sets Syst*. (1996) 78:139–53. doi: 10.1016/0165-0114(95)00165-4

32. Chen CT. Extension of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst.* (2000) 114:1–9. doi: 10.1016/S0165-0114(97)00377-1
33. Meena K, Manimekalai M, Rethinavalli S. A novel framework for filtering the PCOS attributes using data mining techniques. *Int J Eng Res Technol*. (2015) 4:702–6.
34. Satish C, Nandipati R, Chew X, Khaw KW. Polycystic Ovarian Syndrome (PCOS) classification and featur selection by machine learning techniques. *Appl Math Comput Intelligence*. (2020) 9:65–74.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Public Health

OPINION
published: 10 December 2021
doi: 10.3389/fpubh.2021.779328

Chronic Pain Treatment and Digital Health Era-An Opinion

V. Rejula, J. Anitha, R. V. Belfin and J. Dinesh Peter*

Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

Keywords: chronic pain, health innovation, artificial intelligence, eHealth, digital health, virtual care, digital biomarkers, quality of life

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Anand Deva Durai C,
King Khalid University, Saudi Arabia
I-Hsien Ting,
National University of Kaohsiung,
Taiwan

****Correspondence:***
R. V. Belfin
researchbelfin@gmail.com

Specialty section:
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: *18 September 2021*
Accepted: *22 November 2021*
Published: *10 December 2021*

Citation:
Rejula V, Anitha J, Belfin RV and Peter JD (2021) Chronic Pain Treatment and Digital Health Era-An Opinion.
Front. Public Health 9:779328.
doi: 10.3389/fpubh.2021.779328

1. INTRODUCTION

Pain is acknowledged as the body alarm system that reminds us of the environmental threat, muscle damage or the presence of some disease condition (1, 2). The pain is directly related to negative emotions (3). The stimulation of primary sensory neurons and specialized transduction machinery in their peripheral terminals is thought to be the physiological cause of pain. There are two sorts of pain based on the continuous pain period, namely acute and chronic (4).

Chronic pain is defined as pain that lasts for more than 3 months (5) and it is an significant health issue. Low back pain is one of the disease that wreaks more havoc on people's lives (6–8). Other leading causes and conditions of chronic pain are Rheumatoid arthritis (9–11), Shoulder Pain (8), Headache disorders (12, 13), Cancer (14, 15), Fibromyalgia (16, 17), Cervical and Thoracic Pain (18).

Depression (19, 20), anxiety (21), sleep problems (22), fatigue/lack of energy (23), and neurocognitive abnormalities (24) are all common comorbidities of chronic pain. These comorbidities reduce the patient's quality of life, create lost workdays, and make it difficult to maintain a healthy social life on their own (25, 26).

The relevance of chronic pain research along with the techniques and problems of diagnosing chronic pain, various treatment approaches, and current digital patient interaction tools has been discussed in this article. The summary of chronic pain treatment and digital patient engagement methods has been given in **Figure 1**. Finally, the benefits of digital health approaches during the COVID 19 pandemic and post-pandemic era are presented.

2. PAIN IDENTIFICATION METHODS

There are two main aspects to pain: intensity (magnitude) and unpleasantness (effect) (27). In terms of characteristics such as duration, frequency, location, and severity, reporting the pain is essential (28). The measurement of pain can be divided into three broad classifications. "Self-report" (29) is the primary way for pain measurement. There are various scales such as Verbal Rating Scales (VRS) (30), Numerical rating scales (NRS) (31), Visual Analog Scales (VAS) (32), Smiley- based Wong-Baker Scale (WBS) (33), Faces Pain Scales-Revised (FPS-R) (34) are used to get the self-reports from the chronic pain patients (35–38). The second way of measuring pain is the "Observe behavior and infer" method (39). This method has various scales like Neonatal Infant Pain Scale (NIPS) (40), Crying Requires Increased Vital Signs Expression Sleeplessness (CRIES) (41), Face, Legs, Activity, Cry, Consolability (FLACC) (41) for Infants and Toddlers, then Pain Assessment Checklist for Seniors with Limited Ability to Communicate (PACSLAC) (42), DOLOPLUS2 (43), Pain Assessment in Advanced Dementia Scale (PAINAD) (44) for elderly with dementia and Behavioral Pain Scale (BPS) (45), Critical Care Pain Observation Tool (CPOT) (46), Nonverbal Pain Scale (NVPS) (47) for ill and unconscious persons. The third classification of measuring pain is "Indirect Physiology." These methods use modalities like electroencephalogram (EEG)

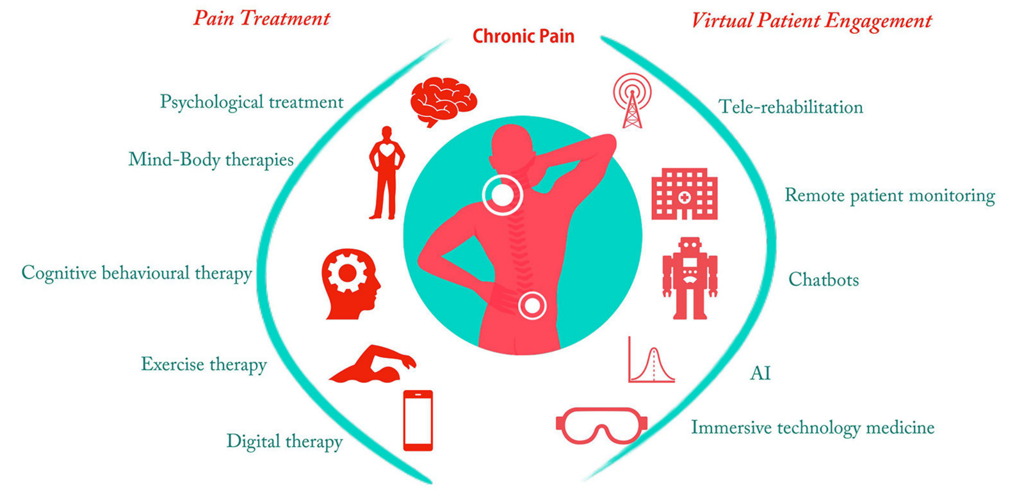


FIGURE 1 | Chronic pain treatment and digital patient engagement methods.

(48, 49), Magnetoencephalography (MEG) (49, 50), Positron emission tomography (PET) (7) and functional MRI (fMRI) (2, 51, 52) to measure pain using the bio-markers. Having a variety of pain measurement methods does not mean that measuring pain is simple. There are many challenges in measuring pain accurately when we use self-reports and observe behavior and infer. The following section provides the challenges in measuring pain.

2.1. Challenges of Measuring Pain

The "self-report" and "Observe behavior and infer" report changes depending on mood, environment, and cognition. Even after the patients have been trained, the sensitivity, robustness, and reliability of the measurements obtained by self-report are quite poor. A baby, a demented old person, or an anesthetized or coma patient cannot self-report pain. One issue with observer rating is that not everyone exhibits pain reactions. A range of behavioral and physiological changes can be used to quantify responses to noxious stimuli. Lack of signals during noxious stimulation does not always imply the absence of pain. It's critical to have measurements that can be translated into pain treatment. Hence, there is a need for a bio-marker (35).

2.2. Digital Bio-Markers

A bio-marker is a feature that may be tested and assessed quantitatively as an indicator of a normal biological process, a pathological process, or the pharmacological (and non-pharmacological) response to a therapeutic treatment (53). Food and Drug Administration -National Institutes of Health (FDA-NIH) (54) working group classifies bio-markers into four types according to the development of drugs and biologic:

- **Diagnostic bio-markers** -To identify and validate the presence of pain.
- **Prognostic bio-markers** -To recognize the likelihood of a clinical event, disease recurrence, or progression in patients with disease of interest.
- **Predictive bio-markers** -To identify individuals who are more likely to have the bio-marker than those who do not.
- **Pharmacodynamic bio-markers** -To show that a biological response occurs in an individual exposed to a medical product.

3. CHRONIC PAIN TREATMENT

Chronic pain is treated in a variety of ways by healthcare practitioners. The treatment plan is determined by the type of pain, the cause (if known), and other factors that differ from person to person. Medication, lifestyle modifications, and treatments are all used in the best treatment programmes.

3.1. Psychological Treatment

Education is an essential element of psychological pain treatment (55) since it teaches patients how to handle a problematic condition. Talk therapy, Relaxation training, Stress management and pain coping skills training are the most frequent psychological treatments. You can obtain the assistance and counselling of a psychiatrist or psychologist through talk therapy. Cognitive therapy helps to alter the facts, ideas, and attitudes to change the amount of pain you feel. Relaxation training teaches you how to achieve a physiological state of profound relaxation that help you relieve pain. You may discover how your ideas impact your stress level and how to build a healthy response to difficult situations via stress management. You may learn how to adjust your life to the pain and have fun again through pain coping skills training (56).

3.2. Mind-Body Therapies

Several researchers on the use of mind-body therapy to treat persistent lower back pain have found that it is effective in reducing pain intensity, tension, and anxiety while also increasing aerobic ability. Mind-body treatments may encourage whole-body healing and include a fitness component to increase the brain's connection with physical function. They are often considered part of supplementary alternative or integrative medicine, which provides movement and energy-based therapies. Patients with persistent lower back pain benefit from integrating complementary alternative mind-body treatments in their therapy, according to a variety of medical and healthcare professionals (e.g., primary care, spine surgeons, pain management). Holistic care is sometimes coupled with traditional medical therapy (57).

3.3. Cognitive Behavioral Therapy

Cognitive Behavioral Therapy (CBT) combines psychotherapy (talk therapy) with behavior therapy to help individuals better comprehend life's difficulties, retrain their attitudes, ideas, and perceptions, and develop problem-solving skills. A CBT therapist guides the person through goal-oriented therapy to learn how to manage stress, anxiety, depression, and sleep issues (58).

3.4. Exercise Therapy

When you're dealing with chronic pain or an injury, exercise therapy is the best option. It can help you feel and move better by making you stronger. Physical therapists are skilled in both treating and preventing pain. They will search for areas of weakness or stiffness that might be contributing to the tension in the painful regions. They'll also prescribe activities to help you move better and relieve discomfort in certain places (59).

3.5. Digital Therapy

The use of software, mobile applications, sensors, and other digital approaches as therapeutic interventions to address various medical problems is known as digital therapy. Digital therapies refer to all of these devices and methods together. In essence, digital treatment is not dissimilar to traditional face-to-face therapeutic procedures. The cognitive behavioral therapy technique is used in the majority of digital therapy practices. It focuses on the notion of providing frequent counseling to patients to improve their behavioral and lifestyle habits (60).

4. CHRONIC PAIN VIRTUAL PATIENT ENGAGEMENT

Many people believe that digital communication channels promote obsessive, compulsive social media illness in individuals. They are currently the effective means of contact between patients and medical professionals. We had no idea they'd become such a robust virtual tool for patient involvement. In these challenging days of the COVID-19, digital platforms assist in reaching out to patients regardless of location and save millions of lives.

4.1. Telerehabilitation

Telerehabilitation for pain treatment makes use of communication technologies to overcome geographical limitations. Access to such technology was indispensable during the COVID'19, and it was especially beneficial for individuals with chronic pain problems and unable to travel. Evaluating and treating such illnesses need a whole-health strategy that personalizes treatment choices and uses a biopsychosocial method to offer care. The care goals are the same as they are in a face-to-face encounter between a patient and a clinician. With proper consideration for staging a pain examination, a systematic approach to the physical visit, and the use of established clinical measures, telerehabilitation can be successfully applied in pain treatment (61).

4.2. Remote Patient Monitoring

RPM (remote patient monitoring) is a critical aspect of medical treatment in the future. When the objective of patient treatment is to enhance the quality of life, doctors must track how patients are doing between the appointments. This method of monitoring patients is especially crucial when dealing with chronic diseases like pain. The technologies like cloud computing, smartphone apps, sensors and wearables paved the way to improve this field (62). These software's collect, track, evaluate, and manage pain related data to improve the quality of life of a pain patient (54).

4.3. Chatbots

Today, algorithms-driven communications done through intelligent devices aim to treat hurts like substance use disorders, combat traumas, chronic pain and a worldwide pandemic. Conversational AI bots (chatbots) is a way to bridge the gap between the people and the knowledge they need and assisting them in completing tasks more quickly, all while utilizing existing tools. Conversational AI's future lies in incorporating it more into the care process across the whole care journey, whether it's assisting with you for appointments or just assisting you in living your best life (63, 64).

4.4. AI

The use of artificial intelligence in healthcare is growing, particularly in diagnosis and therapy management. AI applications in healthcare have recently made enormous impacts throughout medical services, igniting a debate about whether AI physicians would eventually displace human doctors (65). Experts, on the other hand, feel that human physicians will not be replaced by robots anytime soon. Artificial intelligence in healthcare, on the other hand, can assist doctors in making better clinical decisions or potentially replace human judgement in certain sectors of the industry.

Recent research used artificial intelligence, or machine learning algorithms, to physiological information from individuals with chronic pain, including respiration rate, oxygen levels, pulse rate, body temperature, blood pressure, and so on. The investigators' technique outperformed baseline models in determining subjective pain levels and differentiating between pain changes and irregular pain fluctuations (66).

4.5. Immersive Technology Medicine

A virtual reality (VR)-based digital therapy for persons with chronic pain has shown promising results in a clinical study (67), helping patients control their fear of movement, which can limit their activity and slow healing. People with chronic pain can utilize a virtual reality headset to lead them through a sequence of cognitive behavioral therapy (CBT) workouts to give them confidence in their fear of movement, commonly known as kinesiophobia (68).

5. DISCUSSION

The techniques for identifying, treating, and managing chronic pain had been discussed in the preceding section. People nowadays utilize the internet and other online resources to learn about the symptoms, treatments, and comorbidities of any ailments they may be suffering. For any disease, taking the medication without first visiting a doctor is not recommended. However, the utilization of digital health data can aid medical professionals, and scientists make better informed and effective decisions.

As mentioned in previous sections, chronic pain can be self-managed by adequate patient education, lifestyle therapies such as exercise, and weight loss activities. There are various hurdles to sufficient treatment in certain underdeveloped countries, such as geography, a lack of physicians, expense, inconvenience, impairment, or, more recently, COVID'19 limitations. However, given the widespread use of computers, intelligent applications, and the world wide web, digital health methods provide the possibility of delivering lifestyle therapies to individuals with chronic pain remotely and assisting them in self-management.

As people go about their regular lives, cell phones can collect digital traces. The behavioral characteristics of a person may be derived from this raw digital data. This digital data gives us the number of unique goals visited. For example, the average level of ambient noise detected during the night, and the sleep quality data offer information on the behavior patterns of a particular individual in a real-world scenario. Combining smartphone and wearable sensors may enable passive and remote real-time monitoring of changing symptoms by capturing behavioral and physiological changes during pain. Applying machine learning classification to smartphone data and other sensor data can help us classify chronic pain patients from the controls.

In healthcare, AI has a variety of effects. In most cases, AI in healthcare uses a web data set to give doctors and specialists access to a large number of diagnostic tools. Because doctors are extremely skilled in their area and keep up with current research, AI technology in healthcare creates a much faster outcome that can be matched with their clinical expertise.

The possibilities for AI applications in healthcare are vast, ranging from emergency rooms to general care to home care. Artificial intelligence can be used in healthcare to automate patient evaluation and minimise assessor bias. It can assess patient risk, analyse sickness (for example, by decoding ECG findings and X-ray images), choose the best therapy based on a patient's clinical history and clinical trial results, track disease and detect early warning signals of worsening.

A biomarker is a biological indicator of an ailment or disease's presence. Biomarkers assist doctors in assessing and diagnosing patients as well as monitoring their therapeutic efficacy. Body temperature for fever, blood pressure for hypertension, and cholesterol levels for heart problems are examples of biomarkers. Doctors can provide target treatment by identifying and using pain biomarkers in conjunction with digital health technology. There are some drawbacks to using biomarkers for pain. Some pain biomarkers (35, 52, 69–72) are stronger predictors in males, whereas others are better predictors of pain in women concerning gene expression. CNTN1 has been related to chronic pain in women, whereas LY9 (lymphocyte antigen 9) and MFAP3 have been connected to PTSD in males. We will have higher precision when we tailor the health care application by gender since men and women have distinct biomarkers for pain in terms of gene expression. Biomarkers for pain can be identified in brain signal markers also.

Currently, 81 percent of the world's population owns a smartphone, which opens up new possibilities for remote health monitoring. The fast growth of computing technologies like mobile apps, cloud computing, blockchain, big data analytics, AI, AR & VR helps build more intelligent applications in this domain. Clinical trials and guidelines often recommend pharmacotherapy, psychotherapy, integrative therapies, and invasive procedures as part of a personalized multi-modal, multidisciplinary treatment strategy.

Incorporating the doctors with customized multi-modal treatment strategies, historical health data, advanced software technologies, and biomarkers can help the pain patients' improve their lives. The availability of huge data to train prediction algorithms, which assist (rather than replace) human doctors, stimulate curiosity-based thinking, enable cooperation, and reduce routine chores, thereby enhancing patient care, will push the application of AI in healthcare. Digital health solutions that are data-driven have the potential to change the health care domain. Suppose these technologies could be sustainably given at scale, they could provide equal access to expert-level treatment to everyone, everywhere, decreasing the worldwide health and wellness gap.

AUTHOR CONTRIBUTIONS

VR, JA, and RVB devised the work, the main conceptual ideas and the proof outline. VR and JA worked out almost all of the technical details. VR, RVB, and JDP worked on the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Davis KD, Flor H, Greely HT, Iannetti GD, MacKey S, Ploner M, et al. Brain imaging tests for chronic pain: medical, legal and ethical issues and recommendations. *Nat Rev Neurol*. (2017) 13:624–38. doi: 10.1038/nrneurol.2017.122
2. Santana AN, Cifre I, de Santana CN, Montoya P. Using deep learning and resting-state fMRI to classify chronic pain conditions. *Front Neurosci*. (2019) 13:1313. doi: 10.3389/fnins.2019.01313
3. Wiech K, Tracey I. The influence of negative emotions on pain: behavioral effects and neural mechanisms. *Neuroimage*. (2009) 47:987–94. doi: 10.1016/j.neuroimage.2009.05.059

4. Wang S, Du J, Xi D, Shao F, Qiu M, Shao X, et al. Role of GABAAR in the transition from acute to chronic pain and the analgesic effect of electroacupuncture on hyperalgesic priming model rats. *Front Neurosci*. (2021) 15:727. doi: 10.3389/fnins.2021.691455
5. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. *J Med Internet Res*. (2020) 22:1–14. doi: 10.2196/18250
6. Ung H, Brown JE, Johnson KA, Younger J, Hush J, Mackey S. Multivariate classification of structural MRI data detects chronic low back pain. *Cerebral Cortex*. (2014) 24:1037–44. doi: 10.1093/cercor/bhs378
7. Shen W, Tu Y, Gollub RL, Ortiz A, Napadow V, Yu S, et al. Visual network alterations in brain functional connectivity in chronic low back pain: a resting state functional connectivity and machine learning study. *Neuroimage Clin*. (2019) 22:101775. doi: 10.1016/j.nicl.2019.101775
8. Hakala PT, Rimpelä AH, Saarni LA, Salminen JJ. Frequent computer-related activities increase the risk of neck-shoulder and low back pain in adolescents. *Eur J Public Health*. (2006) 16:536–41. doi: 10.1093/eurpub/ckl025
9. Abe N, Fujieda Y, Asou K, Karino K, Kono M, Narita H. Resting-State functional connectivity of pain processing brain region associated with therapeutic response to biologics in rheumatoid arthritis and spondyloarthritis. (2021) p. 3–5.
10. Bennell K. Digital health approaches for the lifestyle management of osteoarthritis. *Osteoarthr Cartil*. (2021) 29:S8. doi: 10.1016/j.joca.2021.02.020
11. Barroso J, Wakaizumi K, Reis AM, Baliki M, Schnitzer TJ, Galhardo V, et al. Reorganization of functional brain network architecture in chronic osteoarthritis pain. *Hum Brain Mapp*. (2021) 42:1206–22. doi: 10.1002/hbm.25287
12. Kim S, Bae DW, Park SG, Park JW. The impact of Pain-related emotions on migraine. *Sci Rep*. (2021) 11:1–9. doi: 10.1038/s41598-020-80094-7
13. Santana L, Liu C. Experience of botulinum toxin a injections for chronic migraine headaches in a pediatric chronic pain clinic. *J Pediatr Pharmacol Ther*. (2021) 26:151–6. doi: 10.5863/1551-6776-26.2.151
14. Adam R, Bond CM, Burton CD, de Bruin M, Murchie P. Can-Pain-a digital intervention to optimise cancer pain control in the community: development and feasibility testing. *Support Care Cancer*. (2021) 29:759–69. doi: 10.1007/s00520-020-05510-0
15. Mercadante S, Portenoy RK. Understanding the chameleonic breakthrough cancer pain. *Drugs*. (2021) 81:411–8. doi: 10.1007/s40265-021-01466-5
16. Gilron I, Robb S, Tu D, Holden R, Towheed T, Ziegler D, et al. Double-blind, randomized, placebo-controlled crossover trial of alpha-lipoic acid for the treatment of fibromyalgia pain: The IMPALA trial. *Pain*. (2021) 162:561–8. doi: 10.1097/j.pain.0000000000002028
17. Goebel A, Krock E, Gentry C, Israel MR, Jurczak A, Urbina CM, et al. Passive transfer of fibromyalgia symptoms from patients to mice. *J Clin Invest*. (2021) 131:e144201. doi: 10.1172/JCI144201
18. Price M, Ravanpay A, Daniels C. Recognition of prodromal cervical spondylotic myelopathy presenting in a US Veteran referred to chiropractic for acute thoracic pain: a case report. *J Bodyw Mov Ther*. (2021) 28:13–7. doi: 10.1016/j.jbmt.2021.07.041
19. Ayzenberg I, Richter D, Henke E, Asseyer S, Paul F, Trebst C, et al. Pain, depression, and quality of life in neuromyelitis optica spectrum disorder: a cross-sectional study of 166 AQP4 antibody-seropositive patients. *Neurol Neuroimmunol Neuroinflammat*. (2021) 8:1–13. doi: 10.1212/NXI.0000000000000985
20. Kremer M, Becker LJ, Barrot M, Yalcin I. How to study anxiety and depression in rodent models of chronic pain? *Eur J Neurosci*. (2021) 53:236–70. doi: 10.1111/ejn.14686
21. Li XH, Matsuura T, Xue M, Chen QY, Liu RH, Lu JS, et al. Oxytocin in the anterior cingulate cortex attenuates neuropathic pain and emotional anxiety by inhibiting presynaptic long-term potentiation. *Cell Rep*. (2021) 36:109411. doi: 10.1016/j.celrep.2021.109411
22. Sorrentino F, Petito A, Angioni S, D'Antonio F, Severo M, Solazzo MC, et al. Impact of anxiety levels on the perception of pain in patients undergoing office hysteroscopy. *Arch Gynecol Obstet*. (2021) 303:999–07. doi: 10.1007/s00404-020-05885-9
23. Eccles JA, Davies KA. The challenges of chronic pain and fatigue. *Clin Med*. (2021) 21:19. doi: 10.7861/clinmed.2020-1009
24. Longoria JN, Pugh NL, Gordeuk V, Hsu LL, Treadwell M, King AA, et al. Patient-reported neurocognitive symptoms influence instrumental activities of daily living in sickle cell disease. *Am J Hematol*. (2021) 96:1396–406. doi: 10.1002/ajh.26315
25. Tomlinson RM, Bax KC, Ashok D, McMurtry CM. Health-related quality of life in youth with abdominal pain: an examination of optimism and pain self-efficacy. *J Psychosom Res*. (2021) 147:110531. doi: 10.1016/j.jpsychores.2021.110531
26. Aviram J, Lewitus GM, Vysotski Y, Yellin B, Berman P, Shapira A, et al. Prolonged medical cannabis treatment is associated with quality of life improvement and reduction of analgesic medication consumption in chronic pain patients. *Front Pharmacol*. (2021) 12:1199. doi: 10.3389/fphar.2021.613805
27. Nahman-Averbuch H, Hughes C, Hoeppli ME, White K, Peugh J, Leon E, et al. Communication of pain intensity and unpleasantness through magnitude ratings: influence of scale type, but not sex of the participant. *medRxiv*. (2021) doi: 10.1101/2021.06.24.21259433
28. Könning A, Rosenthal N, Brown D, Stahlschmidt L, Wager J. Severity of chronic pain in german adolescent school students: a cross-sectional study. *Clin J Pain*. (2021) 37:118–25. doi: 10.1097/AJP.0000000000000898
29. Ali SM, Lau WJ, McBeth J, Dixon WG, van der Veer SN. Digital manikins to self-report pain on a smartphone: a systematic review of mobile apps. *Eur J Pain*. (2021) 25:327–38. doi: 10.1002/ejp.1688
30. Lee HJ, Cho Y, Joo H, Jeon JY, Jang YE, Kim JT. Comparative study of verbal rating scale and numerical rating scale to assess postoperative pain intensity in the post anesthesia care unit: a prospective observational cohort study. *Medicine*. (2021) 100:e24314. doi: 10.1097/MD.0000000000024314
31. Shrestha D, Shrestha R, Grotle M, Nygaard ØP, Solberg TK. Validation of the nepali versions of the neck disability index and the numerical rating scale for neck pain. *Spine*. (2021) 46:E325. doi: 10.1097/BRS.0000000000003810
32. Crellin DJ, Harrison D, Santamaria N, Huque H, Babl FE. The psychometric properties of the visual analogue scale applied by an observer to assess procedural pain in infants and young children: an observational study. *J Pediatr Nurs*. (2021) 59:89–95. doi: 10.1016/j.pedn.2021.01.018
33. Alizadeh Z, Paymard A, Khalili A, Hejr H. A systematic review of pain assessment method in children. *Ann Trop Med Public Health*. (2017) 10:847–7. doi: 10.4103/ATMPH.ATMPH_155_17
34. Balparda K, Herrera-Chalarca T, Cano-Bustamante M, Gómez-González T, Nicholls-Molina MA. Rasch development and validation of a new faces scale for measuring pain, and its comparison with a gold standard: the Balparda-Herrera Pain Scale. *Pain Manag*. (2021) 11:689–703. doi: 10.2217/pmt-2021-0005
35. Tracey I, Woolf CJ, Andrews NA. Composite pain biomarker signatures for objective assessment and effective treatment. *Neuron*. (2019) 101:783–800. doi: 10.1016/j.neuron.2019.02.019
36. Werner P, Lopez-Martinez D, Walter S, Al-Hamadi A, Gruss S, Picard R. Automatic recognition methods supporting pain assessment: a survey. *IEEE Trans Affect Comput*. (2019) p. 1–22. doi: 10.1109/TAFFC.2019.2946774
37. Xu X, Huang Y. Objective pain assessment: a key for the management of chronic pain. *F1000Research*. (2020) 9:1–7. doi: 10.12688/f1000research.20441.1
38. Louwerse MD, Hehenkamp WJ, van Kesteren PJ, Lissenberg BI, Brölmann HA, Huirne JA. Electronic continuous pain measurement vs verbal rating scale in gynaecology: a prospective cohort study. *Eur J Obstetr Gynecol Reprod Biol*. (2021) 256:263–9. doi: 10.1016/j.ejogrb.2020.11.012
39. Lautenbacher S, Sampson EL, Pähl S, Kunz M. Which facial descriptors do care home nurses use to infer whether a person with dementia is in pain? *Pain Med*. (2017) 18:2105–15. doi: 10.1093/pm/pnw281
40. Desai A, Aucott S, Frank K, Silbert-Flagg J. Comparing N-PASS and NIPS: improving pain measurement in the neonate. *Advances in Neonatal Care*. (2018) 18:260–6. doi: 10.1097/ANC.0000000000000521
41. Beltramini A, Milojevic K, Pateron D. Pain assessment in newborns, infants, and children. *Pediatr Ann*. (2017) 46:e387–95. doi: 10.3928/19382359-20170921-03
42. Ruest M, Bourque M, Laroche S, Harvey MP, Martel M, Bergeron-Vézina K, et al. Can we quickly and thoroughly assess pain with the PACSLAC-II? A convergent validity study in long-term care residents suffering from dementia. *Pain Manag Nurs*. (2017) 18:410–7. doi: 10.1016/j.pmn.2017.05.009

43. Rostad HM, Utne I, Grov EK, Puts M, Halvorsrud L. Measurement properties, feasibility and clinical utility of the Doloplus-2 pain scale in older adults with cognitive impairment: a systematic review. *BMC Geriatr*. (2017) 17:257. doi: 10.1186/s12877-017-0643-9
44. Fry M, Elliott R. Pragmatic evaluation of an observational pain assessment scale in the emergency department: the pain assessment in advanced dementia (PAINAD) scale. *Australas Emerg Care*. (2018) 21:131–6. doi: 10.1016/j.auec.2018.09.001
45. Rijkenberg S, Stilma W, Bosman RJ, van der Meer NJ, van der Voort PH. Pain measurement in mechanically ventilated patients after cardiac surgery: comparison of the behavioral pain scale (bps) and the critical-care pain observation tool (CPOT). *J Cardiothorac Vasc Anesth*. (2017) 31:1227–34. doi: 10.1053/j.jvca.2017.03.013
46. Dale CM, Prendergast V, Gélinas C, Rose L. Validation of the critical-care pain observation tool (CPOT) for the detection of oral-pharyngeal pain in critically ill adults. *J Crit Care*. (2018) 48:334–8. doi: 10.1016/j.jcrc.2018.09.024
47. Delwel S, Perez RS, Maier AB, Hertogh CM, de Vet HC, Lobbezoo F, et al. Psychometric evaluation of the orofacial pain scale for non-verbal individuals as a screening tool for orofacial pain in people with dementia. *Gerodontology*. (2018) 35:200–13. doi: 10.1111/ger.12339
48. Afrasiabi S, Boostani R, Masnadi-Shirazi MA, Nezam T. An EEG based hierarchical classification strategy to differentiate five intensities of pain. *Expert Syst Appl*. (2021) 180:115010. doi: 10.1016/j.eswa.2021.115010
49. Chang W, Wang H, Yan G, Lu Z, Liu C, Hua C. EEG based functional connectivity analysis of human pain empathy towards humans and robots. *Neuropsychologia*. (2021) 151:107695. doi: 10.1016/j.neuropsychologia.2020.107695
50. Witjes B, Baillet S, Roy M, Oostenveld R, Huygen FJ, de Vos CC. Magnetoencephalography reveals increased slow-to-fast alpha power ratios in patients with chronic pain. *Pain reports*. (2021) 6:e928. doi: 10.1097/PR9.0000000000000928
51. Tétreault P, Mansour A, Vachon-Presseau E, Schnitzer TJ, Apkarian AV, Baliki MN. Brain connectivity predicts placebo response across chronic pain clinical trials. *PLoS Biol*. (2016) 14:e1002570. doi: 10.1371/journal.pbio.1002570
52. Reddan MC, Wager TD. Modeling pain using fMRI: from regions to biomarkers. *Neurosci Bull*. (2018) 34:208–15. doi: 10.1007/s12264-017-0150-1
53. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. *Nat Rev Neurol*. (2020) 16:381–400. doi: 10.1038/s41582-020-0362-2
54. Pathak YJ, Greenleaf W, Verhagen Metman L, Kubben P, Sarma S, Pepin B, et al. Digital health integration with neuromodulation therapies: the future of patient-centric innovation in neuromodulation. *Front Digit Health*. (2021) 3:618959. doi: 10.3389/fdgth.2021.618959
55. de C Williams AC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database Syst Rev*. (2020) 11:CD007407. doi: 10.1002/14651858.CD007407.pub4
56. Darnall BD. *Psychological Treatment for Patients With Chronic Pain*. Washington, USA: American Psychological Association (2019).
57. Martel MO, Bruneau A, Edwards RR. Mind-body approaches targeting the psychological aspects of opioid use problems in patients with chronic pain: evidence and opportunities. *Transl Res*. (2021) 234:114–28. doi: 10.1016/j.trsl.2021.02.013
58. Zambelli Z, Halstead EJ, Fidalgo AR, Dimitriou D. Good sleep quality improves the relationship between pain and depression among individuals with chronic pain. *Front Psychol*. (2021) 12:1555. doi: 10.3389/fpsyg.2021.668930
59. Bandak E, Boesen M, Bliddal H, Daugaard C, Hangaard S, Bartholdy C, et al. The effect of exercise therapy on inflammatory activity assessed by MRI in knee osteoarthritis: secondary outcomes from a randomized controlled trial. *Knee*. (2021) 28:256–65. doi: 10.1016/j.knee.2020.12.022
60. Badawy SM, Abebe KZ, Reichman CA, Checo G, Hamm ME, Stinson J, et al. Comparing the effectiveness of education versus digital cognitive behavioral therapy for adults with sickle cell disease: protocol for the cognitive behavioral therapy and real-time pain management intervention for sickle cell via mobile applications (CaR. *JMIR Res Protoc*. (2021) 10:e29014. doi: 10.2196/29014
61. McGeary D, McGeary C. Telerehabilitation for headache management. *Phys Med Rehabil Clin*. (2021) 32:373–91. doi: 10.1016/j.pmr.2021.01.005
62. Argüello Prada EJ. The internet of things (IoT) in pain assessment and management: an overview. *Inform Med Unlocked*. (2020) 18:100298. doi: 10.1016/j.imu.2020.100298
63. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. *JMIR mHealth uHealth*. (2020) 8:e15806. doi: 10.2196/15806
64. Belfin R, Shobana A, Manilal M, Mathew AA, Babu B. A graph based chatbot for cancer patients. In: *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. Coimbatore: IEEE(2019). p. 717–21.
65. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *npj Digit Med*. (2020) 3:126. doi: 10.1038/s41746-020-00333-z
66. Panaggio MJ, Abrams DM, Yang F, Banerjee T, Shah NR. Can subjective pain be inferred from objective physiological data? Evidence from patients with sickle cell disease. *PLoS Comput Biol*. (2021) 17:e1008542. doi: 10.1371/journal.pcbi.1008542
67. Logan DE, Simons LE, Caruso TJ, Gold JI, Greenleaf W, Griffin A, et al. Leveraging virtual reality and augmented reality to combat chronic pain in youth: position paper from the interdisciplinary network on virtual and augmented technologies for pain management. *J Med Internet Res*. (2021) 23:e25916. doi: 10.2196/25916
68. Haisley KR, Straw OJ, Müller DT, Antiporda MA, Zihni AM, Reavis KM, et al. Feasibility of implementing a virtual reality program as an adjuvant tool for peri-operative pain control; results of a randomized controlled trial in minimally invasive foregut surgery. *Complement Ther Med*. (2020) 49:102356. doi: 10.1016/j.ctim.2020.102356
69. Su Q, Song Y, Zhao R, Liang M. A review on the ongoing quest for a pain signature in the human brain. *Brain Sci Adv*. (2019) 5:274–87. doi: 10.26599/BSA.2019.9050024
70. Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. (2017) 20:365–77. doi: 10.1038/nn.4478
71. van der Miesen MM, Lindquist MA, Wager TD. Neuroimaging-based biomarkers for pain: state of the field and current directions. *PAIN Rep*. (2019) 4:e751. doi: 10.1097/PR9.0000000000000751
72. Lamichhane B, Jayasekera D, Jakes R, Glasser MF, Zhang J, Yang C, et al. Multi-modal biomarkers of low back pain: a machine learning approach. *Neuroimage Clin*. (2021) 29:102530. doi: 10.1016/j.nicl.2020.102530

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Public Health

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fpubh.2021.811730

Acceptability of Traditional Chinese Medicine in Chinese People Based on 10-Year's Real World Study With Mutiple Big Data Mining

Yan Guo[1†], Tengjiao Wang[2,3†], Wei Chen[2,3], Ted J. Kaptchuk[4], Xilian Li[2,3], Xiang Gao[2,3], Jiahui Yao[5], Xudong Tang[1*] and Ziming Xu[1]*

[1] Xiyuan Hospital Affiliated to China Academy of Chinese Medical Sciences, Beijing, China, [2] Big Data Center for Arts and Sciences, Peking University, Beijing, China, [3] Key Laboratory of High Confidence Software Technologies, Ministry of Education and School of Electronics Engineering and Computer Science (EECS), Peking University, Beijing, China, [4] Global Health and Social Medicine, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA, United States, [5] Institute of Social Science Survey, Peking University, Beijing, China

OPEN ACCESS

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Ziquan Zhu,
University of Florida, United States
Mailan Liu,
Hunan University of Chinese
Medicine, China

***Correspondence:**
Wei Chen
pekingchenwei@pku.edu.cn
Xudong Tang
xudong@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 November 2021*
Accepted: *30 November 2021*
Published: *11 January 2022*

Citation:
Guo Y, Wang T, Chen W, Kaptchuk TJ, Li X, Gao X, Yao J, Tang X and Xu Z (2022) Acceptability of Traditional Chinese Medicine in Chinese People Based on 10-Year's Real World Study With Mutiple Big Data Mining. Front. Public Health 9:811730. doi: 10.3389/fpubh.2021.811730

In the past decades, numerous clinical researches have been conducted to illuminate the effects of traditional Chinese medicine for better inheritance and promotion of it, which are mostly clinical trials designed from the doctor's point of view. This large-scale data mining study was conducted from real-world point of view in up to 10 years' big data sets of Traditional Chinese Medicine (TCM) in China, including both medical visits to hospital and cyberspace and contemporaneous social survey data. Finally, some important and interesting findings appear: (1) More Criticisms vs. More Visits. The intensity of criticism increased by 2.33 times over the past 10 years, while the actual number of visits increased by 2.41 times. (2) The people of younger age, highly educated and from economically developed areas have become the primary population for utilizing TCM, which is contrary to common opinions on the characteristics of TCM users. The discovery of this phenomenon indicates that TCM deserves further study on how it treats illness and maintains health.

Keywords: traditional Chinese medicine, TCM, data mining, big data, social review, China

INTRODUCTION

Along with rapid development of modern medicine, the role of traditional Chinese medicine in health care has received worldwide medical attention (1–5). We found some interesting and important contradictory phenomenon during the development of TCM in China that may be summarized as "More Criticisms vs. More Visits": (i) the intensity of criticism increased by 2.33 times over the past 10 years, while the actual number of visits increased by 2.41 times in China. (ii) Surprisingly, young, highly educated people, and living in economically developed areas have become the primary characters of population utilizing TCM. This phenomenon subverts the wide accepted thinking that only older people, those lack of modern medical knowledge, or people in economically underdeveloped areas tend to choose TCM treatment. The above conclusions are verified by different data sources using a variety of machine learning methods as follows, indicating prospective clinical application of TCM in China. In order to thoroughly study this phenomenon, we conducted a series of large-scale data analysis and mining experiments on two types of data sets. The details are as follows: Medical visit data for 10 years in China: National statistical data of TCM industry are obtained from National Bureau of Statistics of the People's Republic of China (NBSPR) from 2004 to 2014. In addition to the macroeconomic data of the NBSPR, Xiyuan Hospital of China Academy of Chinese Medical Sciences (also called "Xiyuan Hospital") contribute a total of 13 million electronic medical records for 10 years for this research. Xiyuan Hospital is a WHO cooperative hospital and one of the most prestigious Chinese medicine hospitals in China. These invaluable medical records cover widespread sources of patients, containing 34 provincial-level administrative units in China and more than 20 countries and regions except China. Cyberspace and social survey data: the public opinion of TCM in cyberspace is obtained from Tianya Forum (the most influential comprehensive network community in China). We extracted a total of 3 billion posts and 20 million users' information involved during over 10 years since 2004. The public opinion of TCM in real world comes from Chinese Family Panel Studies (CFPS), the most influential social survey project in China. Science has reported CFPS in 2010, "60,000 respondents in 25 provinces-making the survey the largest undertaking of its kind in the developing world" (6). In CFPS, we set up a number of questionnaires for the medical treatment of TCM which can comprehensively and objectively reflect the medical viewpoints and crowd characteristics of the Chinese people. In the light of time sequence, we integrate all data into a unified big data platform, which is available for managing and analyzing structured and unstructured data. On this platform, large-scale experiments are conducted on 100-node servers. Following key methods are carried out: (i) topic tracking-active learning method (7) and topical phrases learning method (8) were employed to detect the posts related to TCM topics; (ii) topic mining—we utilize LDA topic mining method (9, 10) to explore the topic categories in posts; (iii) sentiment analysis–opinion aware knowledge graph (11) and sentiment oriented maximum entropy classification method (12) are used to calculate the posters' opinions on TCM; (iv) factor analysis–Lasso (Least Absolute Shrinkage and Selection Operator) Regression (13) and Bayesian network (14) are employed to calculate the dependencies between the growing visits of TCM and the major factors. The above big data platform and machine learning methods can support us to do such objective quantitative analysis: (i) How intense is the criticism of TCM, and the impacts on the actual TCM visiting population? (ii) What are the primary characteristics of people willing to try TCM treatments? (iii) What are the main points of view that people support and oppose TCM?

RESULTS

How Intense Is the Criticism of TCM and the Impacts on the Actual Population Making TCM Visits?

We apply topic tracking to extract all posts related to TCM topic from the 3-billion-Tianya-data set. Then a training set is produced using the sentiment oriented maximum entropy classification method (12) to generate an Opinion-aware Knowledge Graph (OKG) (11). And we apply OKG to determine the opinions of the posts about TCM. Posts are classified into three categories according to their opinions: positive, negative and neutral. The intensity of the posts in cyberspace is illustrated in terms of Opinion Polarity Ratio (OPR). In addition, the contemporaneous trends of the medical visits are calculated through the statistical data from NBSPR during 10 years from 2004 to 2014. From the experimental results, a contradictory phenomenon can be clearly identified: Although the intensity of criticisms (quantified by OPR) increased by 2.33 times over the past decade, the actual number of visits has increased annually, with an increase of 2.41 times in 10 years as shown in **Figure 1**.

What Are the Characteristics of Populations Utilizing TCM Treatments?

It is thought that older people, lack of modern medical knowledge and people in economically underdeveloped areas tend to choose TCM treatments. However, we obtain completely opposite results by data mining. In order to objectively and accurately identify the characteristics of people who adhere to TCM treatments, we explore all the key factors that may affect TCM visits in NBSPR and CFPS (The details of all factors can be referenced to **Supplementary Table S1**). Typical factor analysis methods, Lasso Regression (13) and Bayesian Network (BN) (14), are used to analyze the relationship between these factors and TCM visits. A similar result is obtained from both Lasso Regression and BN: there is a strong correlation between finance, education and the number of TCM visits in a same area (see **Supplementary Figure S2**). **Figure 2A** shows the impact coefficients of various factors on the visits of TCM analyzed by Lasso Regression according to the data from NBSPR (see **Supplementary Figure S1**; **Supplementary Table S2** for detailed results). The factors include Gross Domestic Product (GDP), education, hospital infrastructure, medical consumption, pollution levels, temperature and humidity and so on of all provinces and municipalities. The larger the coefficient, the greater the impact of the factor. GDP and education are the two most influential factors. Moreover, by analyzing the factors and their conditional dependencies via a directed acyclic graph represented by Bayesian Network, developed economy and high education have the strongest correlation on the number of TCM visits (see **Supplementary Figure S2**; **Supplementary Table S3**). **Figures 2C,D** illustrate the influence of different levels of GDP and education on the number of visits, respectively. The people with highly educated and from economically developed areas are more likely to be treated by TCM. In addition to the macro-data analysis of the NBSPR, a detailed analysis of the actual visits data contributed by Xiyuan Hospital are realized. **Figure 2B** depicts the age distributions of Xiyuan Hospital's visit population. It is shown that young people (25–35 years old) with modern science education have become the primary group utilizing TCM treatment! It is surprised to find that the "Little Emperor Group" (15), aged 0–8, has also become a major group utilizing TCM treatment. Thus it can be seen that young parents have high confidence in treating their children with TCM. From the above experimental results, we can draw the conclusion that age, education and economy are the key factors influencing TCM

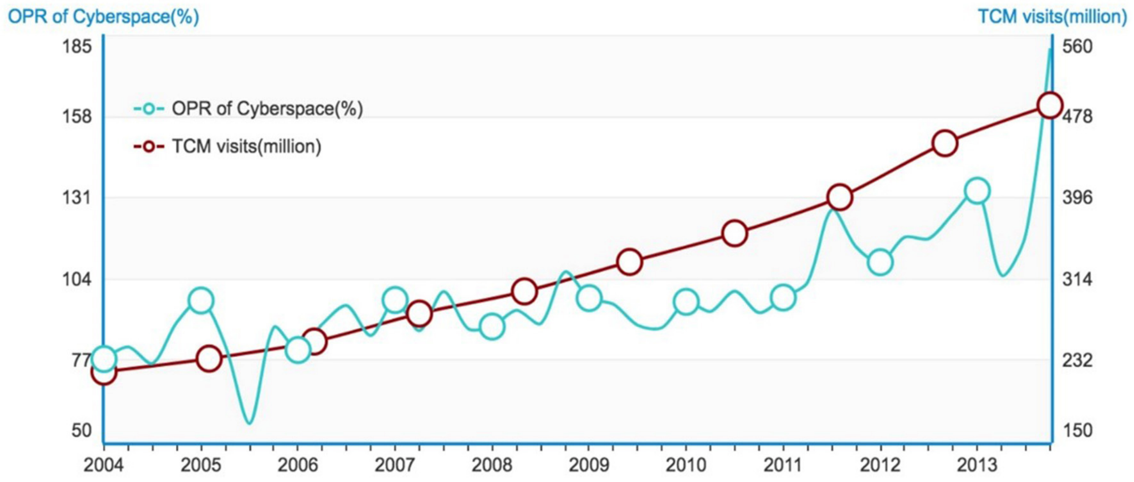


FIGURE 1 | OPR of Cyberspace vs. TCM visits.

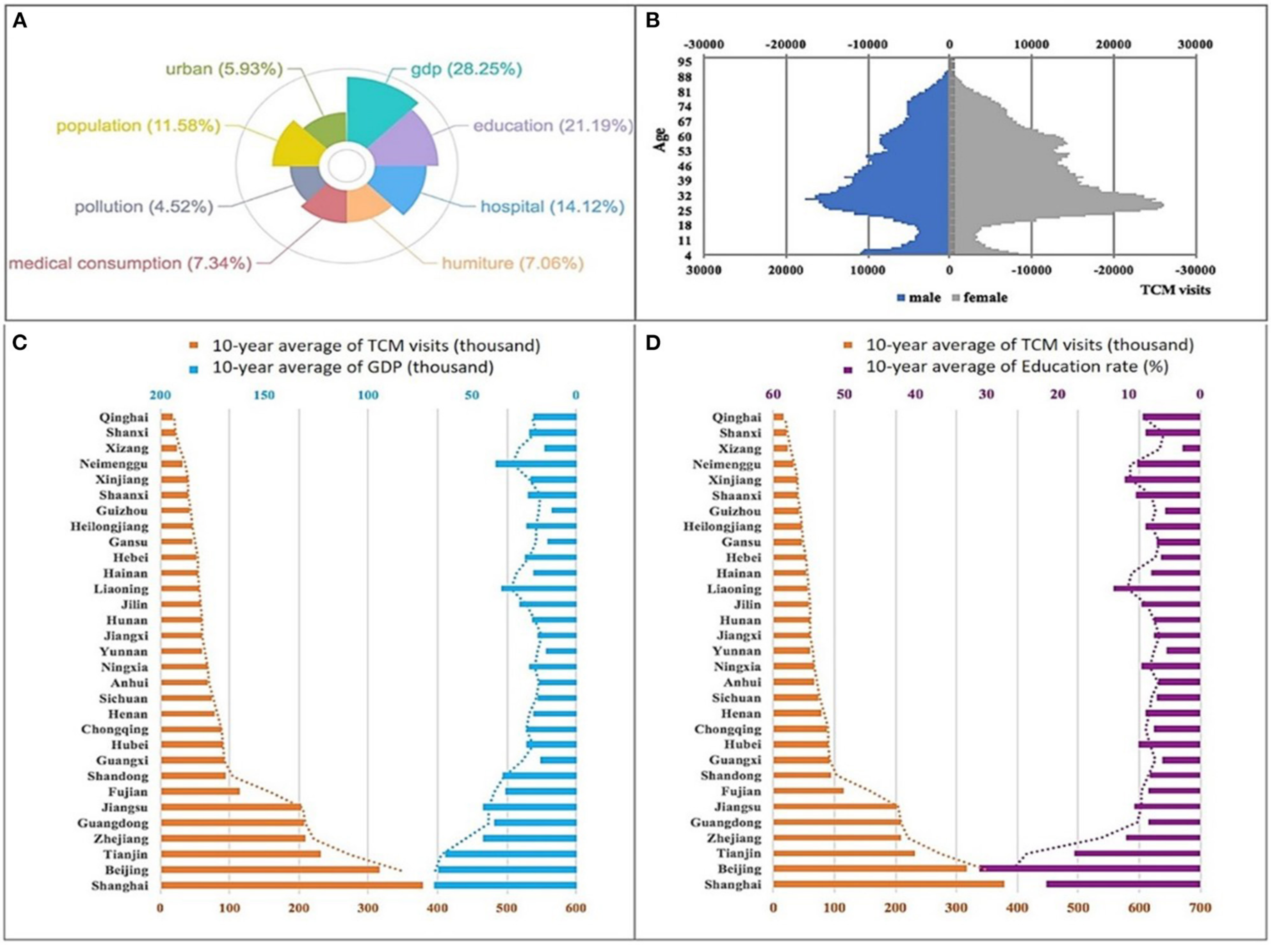


FIGURE 2 | The characteristics of populations utilizing TCM treatments. **(A)** The factor analysis results of Lasso Regression on the NBSPR data. **(B)** The age distribution of Xiyuan Hospital's visit population. **(C)** The relationship analysis between GDP and TCM visits. **(D)** The relationship analysis between education rates and TCM visits.

visits. And people of young, highly educated and economically developed areas have become the mainstay of TCM practitioners.

What Are the Main Reasons That People Support or Oppose TCM?

In addition to the medical visits, cyberspace has generated a huge amount of TCM related experience sharing data (Tianya) in the past 10 years. People's opinions to TCM are implied in these posts. By using sentiment analysis, posts have already been classified into three categories according to their opinions to TCM: positive, negative and neutral. Topics are extracted from positive posts and negative posts respectively by exploring topic mining methods (9, 10). Each top five of all the topics are displayed in **Figure 3**.

Figure 3A shows the positive topics of the top five. The main points people endorse for TCM include the accumulation of culture, years of practical experience, effectiveness with chronic diseases, natural health care efficacy and good curative effect. The co-occurrence relationships of these five positive topics are shown in **Figure 3B**. **Figure 3C** shows the top five negative topics. The criticisms focus on superstition and ignorance, pseudoscience, poor efficacy, large gap with western medicine, and lack of rigorous experimental verification. The co-occurrence relationships of these five negative topics are shown in **Figure 3D** (The details of the topics are shown in **Supplementary Table S4**). People's attitudes toward TCM come from inherent ideas or viewpoints of modern medicine (2, 16). They lack quantitative analysis of the general public opinions. For the first time, the large-scale data analysis has been adopted to calculate people's attitudes toward TCM comprehensively and systematically in this study. The results further verify the contradictory phenomena in this paper.

DISCUSSION

Traditional Chinese Medicine is the general name of all ethnic medicines in China, with a long history, systematic theory and unique techniques. It is a medical system that reflects the Chinese nation's understanding of life, health and disease. During the past 10 years, the number of online visits to TCM content section has doubled, breaking the stigma that only the elderly, people lacking modern medical knowledge and economically underdeveloped areas will be loyal to Traditional Chinese Medicine in treating diseases and maintaining health. In fact, although more interviews also contain many criticisms, the analysis results still show obvious correlation between economy, education and TCM Treatment, that is, higher the economic and education level, more the acceptance of TCM treatment. The young people aged 25–35 years have become the largest audience group of TCM treatment, exceeding the elderly group in number. In addition, children aged 0–8 is another rising group amazedly, which might be due to their parents aged about 25–35, the number 1 audience of TCM treatment. In addition, Traditional Chinese Medicine has advantages in treating and rehabilitating chronic diseases, based on its theory of wholism and treatment according to syndrome differentiation.

The two sides of things evaluation are always accompanied. Although Traditional Chinese Medicine has made continuous achievements in recent years, especially shown favorable efficacy in fighting for COVID-19 prevalence, negative evaluation will still be an indispensable part, which in return promotes the modernization process of TCM. With the improving economy and expansion of educated popularization, the stereotype of TCM still exists, such as uncertain effectiveness, lack of support of modern scientific experiments, pseudoscience and metaphysics, however, more and more clinical and experimental studies with relatively good and scientific design have come into view, and shown favorable effects on many diseases, especially chronic diseases. And the disease treatment mode, especially those of chronic diseases, has turned to the biological social psychological mode, which is more in line with TCM theory. In the process of spreading TCM, it is absolutely not wise to ignore the negative evaluation. Encouraged by positive evaluation, we should accept criticism and analyze its shortcomings, provide more scientific and objective proofs with modern technology and equipments, to serve the people's healthcare worldwidely.

MATERIALS AND METHODS

Although it is widely believed that TCM is full of empirical methods and lack of modern scientific verification, the discovery of the above phenomenon shows that the efficacy of TCM deserves further research. Furthermore, the analysis about the psychology of people's medication selection in modern society will also become an important research topic.

Definition and Terms

Opinion Polarity Ratio

Opinion polarity ratio is the ratio of the total number of negative Tianya posts on TCM topic to the total number of positive Tianya posts. It is abbreviated as OPR. And OPR(s) on the specific topic s is given by the following formula, where $n_+(s)$ is the total number of positive Tianya posts on the topic s, and $n_-(s)$ is the total number of negative Tianya posts on the issues.

$$OPR(s) = \begin{cases} \frac{n_-(s)}{n_+(s)}, n_+(s) \neq 0 \\ \frac{n_-(s)+\delta}{n_+(s)+\delta}, n_+(s) = 0, 0 < \delta \leq 1 \end{cases} \quad (1)$$

In order to deal with the situation of the denominator (the number of positive Tianya posts) being 0, smoothing process is needed in the formula. In the experiments, the smoothing factor δ can be taken from 0 to 1 (excluding 0). Sentiment analysis is applied firstly to obtain the opinion polarities of the Tianya posts, and then calculate the OPR on TCM topics.

Topic Tracking

Topic tracking can overcome the shortcomings of keyword search and minimize the human labeling efforts. We apply a novel multi-label active learning approach with Support Vector Machines (SVM) (8). The detailed process of SVM-based active learning is as follows. For every iteration, we Train a binary SVM classifiers f based on training data Dl. For each instance x in total dataset Du, use the f to assign classification probabilities.

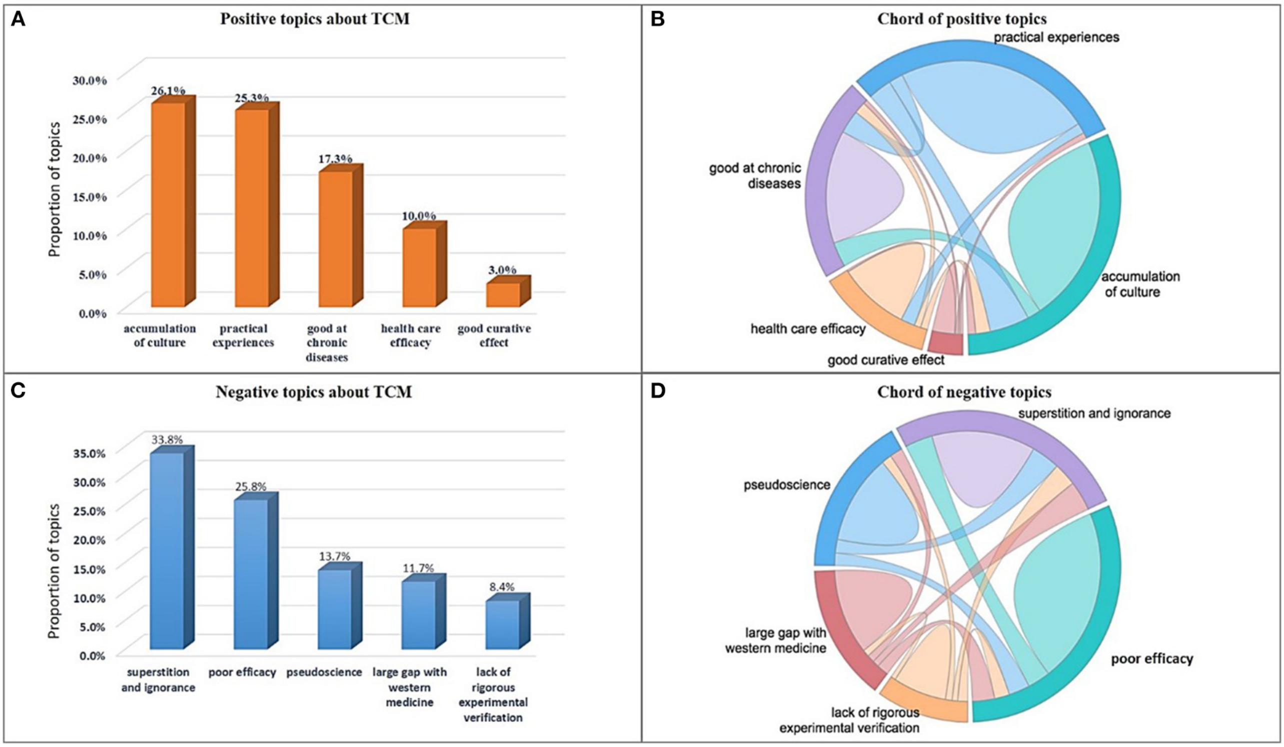


FIGURE 3 | The top five topics about TCM in cyberspace. **(A)** The top five positive topics. **(B)** The co-occurrence relationships between these five positive topics. **(C)** The top five negative topics. **(D)** The co-occurrence relationships between these five negative topics.

Select a set of samples Ds with the largest score, and update the training set Dl = Dl + Ds. At the same time, this set of samples is discarded from Du. If the loop reaches a predetermined number of times, the algorithm terminates and returns to classifier f, otherwise it repeats training SVM. Applying the trained model. In order to process massive posts, the SVM-based active learning classifier is deployed to each data node to do posts classification concurrently. Finally posts related to traditional Chinese medicine are extracted.

Sentiment Analysis

Textual information is categorized into two types: fact and opinion. Fact is about entities, events, and the corresponding objective description. Opinion is usually used to describe the emotion toward entities or events, which can be divided into positive and negative positions. Sentiment analysis method is used to calculate the public opinion on TCM.

Data Preprocessing

To assist the construction of the data sets, beside the word segmentation in topic tracking, the part-of-speech (POS) tagging job is needed additionally.

Training Set Construction

In order to reduce the workload and improve the quality of manual annotation, we run MaxEnt-LDA (17) on data set. MaxEnt-LDA can distinguish entities and sentiment words well by utilizing POS and the co-occurrence information between words. In the labeling process, domain experts who do the labeling job can distinguish these two kinds of words, so as to improve the efficiency and accuracy of labeling.

Feature Selection

Entities and sentiment words are beneficial features for sentiment analysis. In addition, the post provides some emoticons and punctuation marks to help people express their feelings.

Model Training

IIS algorithm (18) is used to train the maximum entropy classifier for post sentiment classification. The main model of classifier is defined as

$$\mathrm{p}(c|d,\lambda)=\frac{\exp[\sum_i \lambda_i f_i(c,d)]}{\sum_{c'}\exp[\sum_i \lambda_i f_i(c',d)]} \tag{2}$$

where c is the category, d is the post and f is the feature vector. The symbol λ denotes the feature weight vector. A larger λi means that the i-th feature is considered as a stronger indicator for class c. The feature-weight vector λ can be obtained through classifier training. We construct an Opinion-aware Knowledge Graph (11) by integrating the opinions and targeted entities extracted from the training set produced by sentiment oriented maximum entropy classification method (12) into an existing structured knowledge base, and perform stance detection of the expression on TCM conveyed by the posts by information propagation on the graph.

Applying the Trained Model

In order to improve the computational efficiency, the trained maximum entropy classifier and OKG are deployed on Hadoop cluster and stance detection is carried out on each data node. Eventually posts are divided into positive, negative and neutral according to its position. An example of sentiment analysis can be seen in **Supplementary Table S6**.

Factor Analysis

Factor analysis method (13, 14) is applied to identify the characteristics of people who adhere to TCM treatment. Because these characteristics also demonstrate the impact of various factors on the visits of TCM. We excavated all the key factors that may affect TCM visits in data from National Bureau of Statistics of the People's Republic of China (NBSPR) and Chinese Family Panel Studies (CFPS). The details of total factors can be referenced to **Supplementary Table S1**. There we use two typical factor analysis methods, Lasso (Least Absolute Shrinkage and Selection Operator) Regression (LR) (13) and Bayesian Network (BN) (14). Factor Analysis includes two stages: data preprocessing and apply the algorithms to select important features.

Data Preprocessing

To assist the training data sets for LR and NB, we first filter data from 2004 to 2014 of 34 provinces. 30 factors are collected, including economic, education, medical, population, environment, weather, social security related and other relevant factors.

Applying the Algorithms

The most typical factor analysis methods, Lasso Regression (13) and Bayesian Network (BN) (14) were used to analyze the relationship between these factors and TCM visits. The independent variable is the factors after data preprocessing, and the dependent variable is the number of TCM visits.

SciKitLearn is applied to train several feature selection model including Random Forst, Lasso Regression, and PCA models on a single machine with 10-fold cross validation method to determine the optimal parameters. The relationship coefficients between these factors and TCM visits are learned. Moreover, to further prove the result of feature selection by Lasso Regression, we also analyze the factors and their conditional dependencies via a directed acyclic graph represented by Bayesian Network, to revise the factors which have strong influence on the number of TCM visits. Bayesian Network method is embedded in Weka, and some important parameters such as estimator, searchAlgorithm, initAsNaiveBayes, maxNrOfParents, etc. are verified multiple times to select the optimal parameters.

Topic Mining

We use topic model to mine topics in 3 billion posts based on the online learning. Topic model can learn the word distributions on topics from the word co-occurrence information in corpus. Specifically, we use its online learning variant10 to deal with the huge number of posts, and adapt it to the dynamic vocabulary of posts.

According to the above method, we dig out the topics on the overall 3 billion posts. Including the top five positive topics and five negative topics, the snippet of all the topics is illustrated in **Supplementary Table S4** in Chinese.

Text Preprocessing

In this paper, many mature natural language processing methods are used for text preprocessing in, topic tracking, stance detection, and topic mining including word segmentation and POS tagging. The following describes how to use these text-preprocessing methods in our work (19).

Word Segmentation

There is no natural division between Chinese words. Chinese post sequences need to be segmented into separate words. Taking user IDs, URLs and email addresses appeared in posts into consideration, we first extract these proper nouns by the predefined regular expressions from the posts and then use ICTCLAS tool to segment words (20). Because the segmentation tool has been trained on a smaller dataset beforehand, and the segmentation on posts can be parallelized, we deploy the segmentation tool onto data nodes and perform word segmentation on the big data platform.

POS Tagging

We mentioned that there are beneficial features to determine the sentiment of a sentence in sentiment analysis section, including sentiment words (adjectives, verbs and others), emoticons and punctuation marks. So it's necessary to determine each word's grammatical rule in the sentence, such as noun, pronoun, verb, adverb, adjective, preposition, conjunction, interjection, and numeral, date, article and determiner. POS tagging is designed for this purpose (17). We use Stanford POS Tagger (21), which is also available for Chinese, to tag posts. And POS tagging is parallelizable for posts. We carry it out on the big data platform in the distributed way. The POS tagging result can be used for selecting sentiment words. It also provides POS information for the MaxEnt-LDA algorithm, which assists experts with building sentiment analysis training sets.

Computing Environment

Discovering the difference of public opinion between the real world and cyberspace requires the support of high performance computing cluster. So we build a distributed computing platform by 100-node servers to support the large-scale data experiments. Machine configurations are shown in **Supplementary Table S5**; **Supplementary Figure S3**. We integrate social survey data and posts into a unified big data platform, which is available for managing and analyzing structured and unstructured data. **Supplementary Figure S5** shows the network configuration information of our cluster. Machines in blue are center servers, machines in orange are name nodes and job trackers, machines in purple run zookeeper to monitor the whole cluster, machines in green are installed with cloud database to restore the computing result, and others are data nodes and task trackers. We firstly train the model on a single machine, and then we deploy the trained model on data nodes in the map-reduce way.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

TW and YG contributed to the overall study design and paper writing. WC contributed to method design and data analysis. TK and XT contributed to medical consultation and method evaluation. XL contributed to data collection and experiment. JY contributed to social issue analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This research is supported by the National Natural Science Foundation of China (Nos. 62176005 and 81774138). Project 2020BD002 supported by PKU-Baidu Fund and Xiyuan Hospital of China Academy of Chinese Medical Sciences.

ACKNOWLEDGMENTS

We owe our thanks to National High Technology Research and Development Program of China for providing support for development of WebEx platform to collect and extract data. Our gratitude also goes to National Science and Technology Major Project of China for providing 100-node computing environment. Finally we wish to thank graduate students at Big Data Center for Arts and Sciences, Peking University: XG, Zhan Li, Wenjing Liao, Changhe Qiu, Yazhou Wang, Lamei Zhang, Ying Zhang, Zhichao Xu, Xiao Zhang, Lei Gai, Gaoyan Ou, Yuxin Chen, Jie Fu and Desi Wen for system development and data preprocessing and Yi Li, Wutao Lin for preparing figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.811730/full#supplementary-material

REFERENCES

1. Chan M. *Supporting the Integration and Modernization of Traditional Medicine*. New York: Amer Assoc Advancement Science (2014). 346(Suppl. 1):2.
2. Cheung, F. TCM Made in China. *Nature*. (2011) 480(Suppl. 1):82. doi: 10.1038/480S82a
3. Schroën Y, van Wietmarschen HA, Wang M, van Wijk EP, Hankemeier T, Xu G, et al. East is east and west is west, and never the twain shall meet? *Science*. (2014) 346(Suppl. 1):10–12.
4. Tian P. Convergence: where west meets east. *Nature*. (2011) 480(Suppl. 1):84-86. doi: 10.1038/480S84a
5. Wang N, Guo Y. Traditional Chinese practice, a promising integrative intervention for chronic non-infectious disease management. *Chin J Integr Med*. (2018) 24:886–90. doi: 10.1007/s11655-018-2572-7
6. Hvistendahl M. Social science: survey to reveal true face of Chinese society. *Science*. (2010) 328:554–555. doi: 10.1126/science.328.5978.554
7. Yang B, Sun J, Wang T, Chen Z. Effective multi-label active learning for text classification. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Sydney. (2009). doi: 10.1145/1557019.1557119
8. Huang W, Chen W, Wang T, Tao S. TaxoPhrase: exploring knowledge base via joint learning of taxonomy and topical phrases. In: *Proceedings of Open Knowledge Base and Question Answering Workshop at SIGIR*. Tokyo (2017).
9. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res*. (2003) 3:993–1022.
10. Wang R, Huang W, Chen W, Wang T, Lei K. ASEM: mining aspects and sentiment of events from microblog. In: *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. Shanghai. (2015). doi: 10.1145/2806416.2806622
11. Chen W, Zhang X, Wang T, Yang B, Li Y. Opinion-aware knowledge graph for political ideology detection. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. (2017). doi: 10.24963/ijcai.2017/510
12. Ou G, Chen W, Li B, Wang T. Exploiting community emotion for microblog event detection. In *Proceedings of the 19th Conference on Empirical Methods in Natural Language Processing*. Doha. (2014).
13. Tibshirani R. regression shrinkage and selection via the Lasso. *J R Stat Soc B*. (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
14. Efron B. Bayes' theorem in the 21st century. *Science*. (2013) 340:1177–8. doi: 10.1126/science.1236536
15. Cameron L, Erkal N, Gangadharan L, Meng X. Little emperors: behavioral impacts of china9s one-child policy. *Science*. (2013) 339:953–7. doi: 10.1126/science.1230221
16. Qiu J. Traditional medicine: a culture in the balance. *Nature*. (2007) 448:126–8. doi: 10.1038/448126a
17. Zhao WX, Jiang J, Yan H, Li X. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA (2010).
18. Pietra SD, Pietra VD, Lafferty J. Inducing features of random fields. *IEEE Trans*. (1997) 19:380–93. doi: 10.1109/34.588021
19. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: University press Cambridge (2008).
20. Zhang HP, Yu HK, Xiong DY, Liu Q. HHMM-based Chinese lexical analyzer ICTCLAS. In: *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*. Sapparo. (2003). doi: 10.3115/1119250.1119280
21. Toutanova K, Klein D, Manning CD, Singer Y. Feature-rich part-of-speech tagging with a cyclic dependency network. In: *Proceedings of the conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*. Edmonton. (2003). doi: 10.3115/1073445.1073478

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

OPINION
published: 23 February 2022
doi: 10.3389/fpubh.2022.804552

Indoor Positioning Systems: A Blessing for Seamless Object Identification, Monitoring, and Tracking

Shilpa Shyam, Sujitha Juliet and Kirubakaran Ezra*

Karunya Institute of Technology and Sciences, Department of Computer Science and Engineering, Coimbatore, India

Keywords: indoor positioning systems (IPS), technique and technology, challenges, features, industry 4.0

OPEN ACCESS

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Victoria Ramos Gonzalez,
Instituto de Salud Carlos III (ISCIII), Spain

****Correspondence:***
Sujitha Juliet
sujitha@karunya.edu

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *29 October 2021*
Accepted: *17 January 2022*
Published: *23 February 2022*

Citation:
Shyam S, Juliet S and Ezra K (2022) Indoor Positioning Systems: A Blessing for Seamless Object Identification, Monitoring, and Tracking.
Front. Public Health 10:804552.
doi: 10.3389/fpubh.2022.804552

INTRODUCTION

Technology is the greatest result of supreme human imagination. The continual proliferation of indoor positioning technology is laid out in this paper along with its advantages and challenges. The authors point out that there are more to be grasped and utilized in this powerful and growing domain.

THE PROMINENCE OF INDOOR POSITIONING SYSTEMS IN THE PAST, PRESENT, AND YEARS TO COME

The global navigation satellite system (GNSS) performs exceedingly well in finding accurate location data anywhere on the planet. It is most sought after for its high accuracy and global coverage. The efficiency of GNSS is only dominant outdoors due to heavy signal multipath and signal attenuation. However, it fails to meet expectations for indoor environments, which is why several indoor localization technologies have popped up. Indoor navigation systems can be wearables, wall mounted devices, or an intelligent model able to calculate the precise location of objects or humans in any sort of sophisticated indoor environment backed up with several obstacles. An indoor navigation system consists of three vital modules: 1. Indoor positioning system module to estimate the object position, 2. Navigation module which helps in routing the object from the current destination, and 3. Object interaction module which helps in providing instructions to the model or system (1). The three module system results in better localization and navigation (modeling, surveying, and mapping of infrastructures) of location-based assets or object tracking, especially in emergency services for disaster management. With the daily invention of new applications, this industry is expected to have a market value of about 24 billion dollars by the year 2023. The aviation industry makes use of this system in helping passengers navigate to lounges, track passenger baggage, and perform other airport related security services. The advertising industry utilizes location-based promotions for the E-commerce sector. The healthcare sector implements location-based services for tracking patient records and whereabouts within the hospital arena. Asset or object-based tracking using the three module system is an inevitable part of the logistics industry. Through this positioning technology, customers can be traced and helped in navigating toward various services available in a railway station, bus stands, etc., benefiting the transportation industry. Indoor positioning technology has also seen a surge in the tourism and automotive industries as easy navigation of tourists and their assets can be monitored along with vehicle identification. The next time one visits the Sydney airport, one could witness the use of apple maps in which navigating through each terminal is made easy using this technology. Indoor positioning technologies not only comply with commercial sector standards but are also made available for day-to-day home services and applications. A tango augmented reality-based indoor location technology has been developed by the technology giant, Google. It would provide detailed and precise location data of the user using their mobile device. Apple has reached far ahead with indoor positioning technologies. They have employed inbuilt ultrawide band (UWB) chips in premium IPhones to calculate the location of a user in real time.

SALIENT TECHNIQUES AND TECHNOLOGIES IN INDOOR POSITIONING SYSTEM

The basic principle behind the indoor navigation and positioning system is to accurately measure the range and distance between two devices. This can be done in two basic methods. The first one is the measurement of the distance using received signal strength (RSS), in which the strength of the signal between the transmitter and receiver determines the location. Though the accuracy is found to be considerable, it is highly influenced and affected by multipath propagation. Conventional but superior technologies, including WiFi, Bluetooth, RFID, Dead Reckoning, Ultrasonic, and ZigBee, fall under this category.

Radio Frequency Identification (RFID) avails the use of radio waves for object detection. The RFID readers and tags undergo interchanging of frequencies during this process. An RFID-based tracking system was implemented for dynamic targets with <1 m localization accuracy which proved it to be a propitious feature for applications where tracking is needed (2). Peer-to-peer communication over shorter distances can be easily established using the most common Bluetooth technology. ZigBee is a sought-after technology when a low cost and low power system has to be implemented. This makes it suitable to be implemented it in smart homes where energy conversation is taken care of Tumlin (3). Dead Reckoning, unlike other technologies, contemplates velocity for measuring position. It determines the present location based on velocity and past position data. A smartphone-based pedestrian dead reckoning system evinced the need for further implementation in this arena by providing exceptional results in indoor positioning systems (4). In ultrasonic systems, the distance is computed using the time of arrival between the emitter and receiver. The coordinates of the emitter are assessed using multilateration to the fixed anchors. The second measurement involves the estimation of the time of flight from several devices. This method comparatively imparts centimeter accuracy and is used by the UWB technology. UWB utilizes both the time difference of arrival (TDOA) and time of arrival (TOA) for measurement purposes. It is also seen to play a significant role in the industrial revolution 4.0.

Several smart factories have emerged by inculcating UWB (5). Khan et al. (6) defend various wireless technologies, including Wi-Fi and LoRa, to be the most worthy of implementing indoor localization applications because of it being vigorous, affordable, and able to utilize a minimum amount of power.

Technologies and techniques in indoor positioning systems go hand in hand. Combinations of technologies or combinations of techniques are perceived to be infused for better accuracy in recent times. Techniques in indoor positioning can be separated into triangulation, proximity, fingerprinting, and vision analysis. The computation of asset location using geometrical features of triangles is known as triangulation. It is further used for computation in two ways: lateration and angulation. Lateration measures distance alone for positioning, unlike angulation which uses both angles and distances. Fingerprinting is conducted in two stages, the online stage (also known as the serving stage) and the offline stage (also known as the training stage) for precise object calculation. Vision analysis is carried out from images received from several points. When an object is detected with respect to a known position, it is known as proximity analysis and requires several fixed detectors for this purpose.

SELECTION OF RELEVANT TECHNIQUES AND TECHNOLOGY ACCORDING TO THE ENVIRONMENT AND NEED OF THE HOUR

There is a huge disparity found in the requirements of indoor positioning when compared to outdoor systems. The dissimilarity in requirements is due to the diverse layout in indoor environments as they have complicated and sophisticated pathways. Hence, the accuracy and coverage demand would vary accordingly. Indoor positioning systems built specifically for assisted living, monitoring patients at home, etc., have a requisite for accuracy within 1 m, whereas systems operated for urban and rural applications demand accuracy of about a few meters. Thus, keeping in mind the application type and its place of execution, a suitable technology has to be chosen. As there is a need for technology in all sectors, no indoor position system can be claimed as the ideal solution (7). Along with accuracy and coverage, maintenance and implementation cost, system size, and power consumption are essential metrics.

The design of an indoor positioning system commences in two stages. First, by determining the principle indoor positioning technology on which it would be based upon, and second, by determining the technique that would be infused along with it. Systems that are supposed to create smart homes, find objects that are misplaced, and track and monitor daily activities are usually implemented using ZigBee, WiFi, and fingerprinting (8). Bluetooth has been adapted for low cost and low power applications (9). Similarly, ZigBee devours minimal power and is inexpensive in most cases and is, thus, used for home applications. Applications that require huge coverage area and centimeter accuracy within larger areas, including industries

TABLE 1 | Comparison of existing indoor localization technologies.

Technology	Accuracy	Range (m)	Power consumption	Noise tolerance	Disadvantages
UWB(Ultra wide band technology)	Very high	1–50	Low	Very high	UWB signal are prone to get obstructed by huge objects.
WiFi	Low	1–50	High	Medium	Utilizes ISM band interferences.
Bluetooth	Medium	1–20	Low	Medium	Suffers from low range
RFID	High	1–50	Low	Medium	Communication security is a question. Suffers from low coverage
ZigBee	Very high	1–50	Low	Medium	Lowe rate of transmission
Dead Reckoning	Medium	1–100	High	Medium	Require high quality sensors

and manufacturing sites, preferably implement UWB (10). They are capable of imparting huge amounts of data using minimal energy. Tracking the motion of a visually impaired person or the movement of humans within a small area can easily be implemented by pedestrian dead reckoning, which is an example of dead reckoning technology (11). Newborn systems based on indoor positioning have been seen to work when using aerial robots, mobile robots (12), and humanoid robots. In such cases, criteria such as battery efficiency and power consumption are vital. Based upon the technology that was chosen to be executed, the suitable technique would be integrated according to the environment and the need of the application. **Table 1** puts forth a comparison of the various technologies used at present in terms of accuracy, range, power consumption, and noise tolerance.

PREVAILING CHALLENGES IN THE IMPLEMENTATION OF INDOOR LOCALIZATION SYSTEMS

Every technology under the indoor positioning system is considered supreme, but it brings inexorable challenges along with it. Once the basic technology and technique are made obvious, the challenges that come with it must be tackled without compromising the requirements of the system. The impediments in an indoor environment should be considered and precision in location data should be accurate (13). Contemporary research proves that UWB technology is largely used for industries and manufacturing sites where it could track and trace both static and dynamic objects at ease. Though its utility is large, the signals of UWB are easily hindered by the indoor obstacles, thereby making error mitigation a necessity (14). Conventional methods such as WiFi and Bluetooth are often considered less often as it offers low range. Ultrasound is often neglected when used for wide ranging locations and has frequency restrictions. Privacy and security are one category that is often abandoned while considering metrics in indoor positioning systems (15). These systems are customized to provide accurate locations of data to the user and its organization alone. The involvement of a third party in such systems is a threat to the user or the organization responsible. Hence, future research is expected to pay more attention to the privacy and the security content of the indoor positioning systems along with the security of the data of users.

DISCUSSION

The needs of humans and technology are swiftly changing. Such needs should acknowledged to bring about changes and revolutions over the course of time. Indoor positioning technology is one such domain that attends to the needs of humans in several ways. Hence, researchers are always on the lookout for new formulations in this arena. Every technology under this system is beneficial and addresses particular complications. It is up to the researcher or the industrialist to select the appropriate technology and technique according to the application needs. From healthcare to travel, indoor positioning technologies are universal and omnipresent. With the internet of things (IoT), intelligent systems and mobile computing are growing at a fast pace as the market of indoor positioning technology has been dramatically increasing. Despite its several advantages, it also comes with several challenges for researchers to improve on. Particularly, the metrics of indoor location systems are its premier challenges. In addition, accuracy, maintenance cost, coverage, scalability, and privacy are major challenges that need to be subdued by implementing efficient measures. Finally, special heed should to be given to the privacy and security of the indoor positioning systems for the personal privacy and security of users.

AUTHOR CONTRIBUTIONS

SS conceived the concept and drafted the manuscript. SJ supervised the study and verified the manuscript. KE performed the review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Department of Science and Technology-Natural Resource Database Management System (DST-NRDMS) [Grant No: NRDMS/UG/NetworkProject/e-13/2019 (C) P-4].

REFERENCES

1. Kunhoth J, Karkar A, Al-Maadeed S, Al-Ali A. Indoor positioning and wayfinding systems: a survey. *Human-centric Comput Inform Sci.* (2020) 10:1–41. doi: 10.1186/s13673-020-00222-0
2. Li J, Feng G, Wei W, Luo C, Cheng L, Wang H, et al. PSOTrack: A RFID-based system for random moving objects tracking in unconstrained indoor environment. *IEEE Intern Things J.* (2018) 5:4632–41. doi: 10.1109/JIOT.2018.2795893
3. Tumlin S. *From Industry to Home: Rapid Development of a ZigBee-Based Indoor Positioning System for Use in Private Residences* (2020).
4. Jeong S, Min J, Park Y. Indoor Positioning Using Deep-Learning-Based Pedestrian Dead Reckoning and Optical Camera Communication. *IEEE Access.* (2021) 9:133725–34. doi: 10.1109/ACCESS.2021.3115808
5. Lumme I. *Indoor Localization in Smart Factory: Utilization of UWB technology in real-case scenario* (2021).
6. Khan FU, Awais M, Rasheed MB, Masood B. A comparison of wireless standards in iot for indoor localization using loPy. *IEEE Access.* (2021) 9:65925–33. doi: 10.1109/ACCESS.2021.3076371
7. Pascacio P, Casteleyn S, Torres-Sospedra J, Lohan ES, Nurmi J. Collaborative indoor positioning systems: a systematic review. *Sensors.* (2021) 21:1002. doi: 10.3390/s21031002
8. Yang C. Design of smart home control system based on wireless voice sensor. *J Sensors*. (2021) 2021:26 doi: 10.1155/2021/8254478
9. Lu X, Yin Y, Zhao N, Wei H. Indoor positioning experiment based on phase ranging with bluetooth low energy (BLE). *J Physics*. 1971:012044. doi: 10.1088/1742-6596/1971/1/012044
10. Xianjia Y, Qingqing L, Queralta JP, Heikkonen J, Westerlund T. Applications of uwb networks and positioning to autonomous robots and industrial systems. In *2021 10th Mediterranean Conference on Embedded Computing (MECO)* (New York, NY: IEEE) (2021). Available online at: https://arxiv.org/pdf/2103.13488.pdf (accessed March 28, 2021).
11. Reyes Leiva KM, Jaén-Vargas M, Codina B, Serrano Olmedo JJ. Inertial measurement unit sensors in assistive technologies for visually impaired people, a review. *Sensors.* (2021) 21:4767. doi: 10.3390/s21144767
12. Su M, Gao S. Design and Implementation of intelligent Home monitoring System based on mobile robot. *J. Phy.* (2021) 1846:12084. doi: 10.1088/1742-6596/1846/1/012084
13. Ashraf I, Hur S, Park Y. Smartphone sensor based indoor positioning: Current status, opportunities, and future challenges. *Electronics.* (2020) 9:891. doi: 10.3390/electronics9060891
14. Ridolfi M, Kaya A, Berkvens R, Weyn M, Joseph W, Poorter ED. Self-calibration and collaborative localization for uwb positioning systems: a survey and future research directions. *ACM Computing Surveys (CSUR).* (2021) 54:1–27. doi: 10.1145/3448303
15. Kim Geok T, Zar Aung K, Sandar Aung M, Thu Soe M, Abdaziz A, Pao Liew C, et al. Review of indoor positioning: Radio wave technology. *Appl Sci.* (2021) 11:279. doi: 10.3390/app11010279

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Public Health

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fpubh.2022.819865

A Multistage Heterogeneous Stacking Ensemble Model for Augmented Infant Cry Classification

Vinayak Ravi Joshi[1], *Kathiravan Srinivasan*[2], *P. M. Durai Raj Vincent*[1], *Venkatesan Rajinikanth*[3] and *Chuan-Yu Chang*[4,5]*

[1] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [2] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, [3] Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Chennai, India, [4] Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan, [5] Service Systems Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan

Understanding the reason for an infant's cry is the most difficult thing for parents. There might be various reasons behind the baby's cry. It may be due to hunger, pain, sleep, or diaper-related problems. The key concept behind identifying the reason behind the infant's cry is mainly based on the varying patterns of the crying audio. The audio file comprises many features, which are highly important in classifying the results. It is important to convert the audio signals into the required spectrograms. In this article, we are trying to find efficient solutions to the problem of predicting the reason behind an infant's cry. In this article, we have used the Mel-frequency cepstral coefficients algorithm to generate the spectrograms and analyzed the varying feature vectors. We then came up with two approaches to obtain the experimental results. In the first approach, we used the Convolution Neural network (CNN) variants like VGG16 and YOLOv4 to classify the infant cry signals. In the second approach, a multistage heterogeneous stacking ensemble model was used for infant cry classification. Its major advantage was the inclusion of various advanced boosting algorithms at various levels. The proposed multistage heterogeneous stacking ensemble model had the edge over the other neural network models, especially in terms of overall performance and computing power. Finally, after many comparisons, the proposed model revealed the virtuoso performance and a mean classification accuracy of up to 93.7%.

Keywords: baby cry, feature vectors, MFCC, spectrograms, stack-based algorithms

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Subha P.,
Sri Sairam Institute of Technology, India
Yunyoung Nam,
Soonchunhyang University, South Korea

***Correspondence:**
Chuan-Yu Chang
chuanyu@yuntech.edu.tw

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *22 November 2021*
Accepted: *05 January 2022*
Published: *24 March 2022*

Citation:
Joshi VR, Srinivasan K, Vincent PMDR, Rajinikanth V and Chang C-Y (2022) A Multistage Heterogeneous Stacking Ensemble Model for Augmented Infant Cry Classification. Front. Public Health 10:819865. doi: 10.3389/fpubh.2022.819865

INTRODUCTION

Globally, around 130 million infants are born every year. Taking good care of babies is a major challenge, particularly in the initial stages of parenting. Many ways and ideas are available in many books and resources, but they do not provide impactful insights into resolving the issues. The principal reason is that it is hard to comprehend the meaning of infant's cries. Newborn children communicate with the world through crying. Experienced guardians, parental figures, specialists, and medical attendants comprehend the cries depending on their experience. Young parents get baffled and experience difficulty calming down their infants since all cry signals sound very similar to them. The major problem faced by many new parents is that they hardly understand the reason for the infant's cry (1–8). It is not possible to identify the reason just by looking at the face or analyzing the emotions of the infant (9–12). Many doctors believe that the reason for the infant's cry is based mainly on the patterns of their voice. If the parents do not know the main reason or the root cause of the infant's cry, then they would not be able to provide the required treatment. So, it would be highly beneficial if significant experiments were performed on the infants' audio signals. Their tone and pitch contribute a lot to finding better results. To find an efficient solution to this problem, a primary emphasis needs to be made on the voice patterns. The voice/audio signals generated by the babies contain many feature vectors that can be used in various deep learning or ensemble models. Machine learning systems possess automated learning capabilities, and their performance improves based on their previous experience devoid of any explicit programming (13–27).

Many convolutional neural network variants, like VGG16, VGG19, and AlexNet, mainly deal with the problems of classifying the results based on the features present in the audio signals. The extracted feature vectors contain much valuable information about the pitch, tone, and amplitude. These can help derive the entropy, energy, and spectral intensity of the audio signals.

Newborn child cry examination expands toward the auditory requirements of the infant's cry signals. Many mechanisms have been exploited for the infant cry order, lumber, and Mel-recurrence cepstral coefficient (MFCC). Using a wide variety of elements to perceive and arrange newborn child crying remains somewhat troublesome, even as vulnerability exists regarding which of these elements is pertinent.

The proposed model makes use of the MFCC algorithm in the initial stages. In strong handling, the Mel-recurrence cepstrum portrays the transient power range of a sound based on a linear cosine change of a log power range on a non-linear Mel scale of recurrence. MFCCs are coefficients that collectivelymake up an MFCC. This is mainly required in the data preprocessing phase because the dataset comprises the audio signals. To perform mathematical calculations and to compute the results, it is vital to convert the audio signals into a 2D feature vector so that various deep learning and advanced classification models can be applied to predict the results.

Our proposed work reveals the following major contributions:

- We performed data preprocessing and generated the spectrograms of the audio signals.
- We worked on the major CNN variants like VGG16 and Yolov4 and used transfer learning approaches to perform multiclass classification. While using the pretrained weights of the models, excessive hyperparameter tuning was performed to enhance the performance of the models.
- A distinctive comparison was made between the CNN variants in terms of accuracy, computational time, and resources. This resulted in Yolov4 being the best model among others. It achieved an accuracy of 75%.
- To improve the performance of the model and provide a more efficient solution, a multistage heterogeneous stacking ensemble model was devised. This model made use of the major ensemble-based boosting algorithms to increase the accuracy and other evaluation metrics of the model.
- On analyzing the cognitive performance of major classifiers, Nu-support vector classification (NuSVC), Random Forest (RF), XGBoost, and AdaBoost were selected for this task. NuSVC is like SVC but uses a parameter to control the number of support vectors. It is based on the implementation of LIBSVM. RF is an outfit learning technique for grouping and relapsing different assignments that work by developing numerous choice trees at the preparing time. For order assignments, the result of RF is the class chosen by most trees. XGBoost is an execution of slope-supported choice trees intended for speed and execution. The AdaBoost algorithm, short for adaptive boosting, is a boosting procedure used as an ensemble method in machine learning. It is called adaptive boosting because the loads are reallotted to each occurrence, with higher loads doled out to inaccurately ordered examples. The proposed multistage heterogeneous stacking ensemble model achieved an accuracy of 93.7%.

RELATED WORK

During the 2000s, most techniques employed in newborn child research were identified with neural organizations, including the scaled form (5). Their review included the details about applying many neural network models and traditional machine learning algorithms like KNN and SVM to predict the reason for the infant's cry. Considering a unique circumstance, the work in (28) zeroed in on making a programmed framework that could recognize diverse newborn child needs dependent on crying. It separated different arrangements of paralinguistic highlights from the child cry sound signals and prepared different rule-based or measurable classifiers.

The work in (29) developed an NonLinear Forcasting (NLF) model that includes the Euclidean distance for its goal work, which is normally a unique instance of difference. In addition, it often experiences slow intermingling. This review proposes a summed up and quick uniting non-negative dormant variable [a generalized and fast-converging non-negative latent factor (GFNLF)] model to resolve these issues. Its primary thought is two-fold: (a) taking on—dissimilarity for its goal work, subsequently improving its portrayal capacity for Host Based Intrusion Detection System (HiDS) information; (b) concluding its energy joined non-negative multiplicative update calculation, along these lines accomplishing its quick intermingling. Experimental investigations on two HiDS grids rising out of genuine RSs show that, with cautiously tuned hyperparameters, the GFNLF model outperforms groundbreaking models in both computational effectiveness and expectation exactness for missing information in a HiDS lattice.

The research in (4) developed a time–frequency-based analysis called STFT. A total of 256 discrete Fourier transform

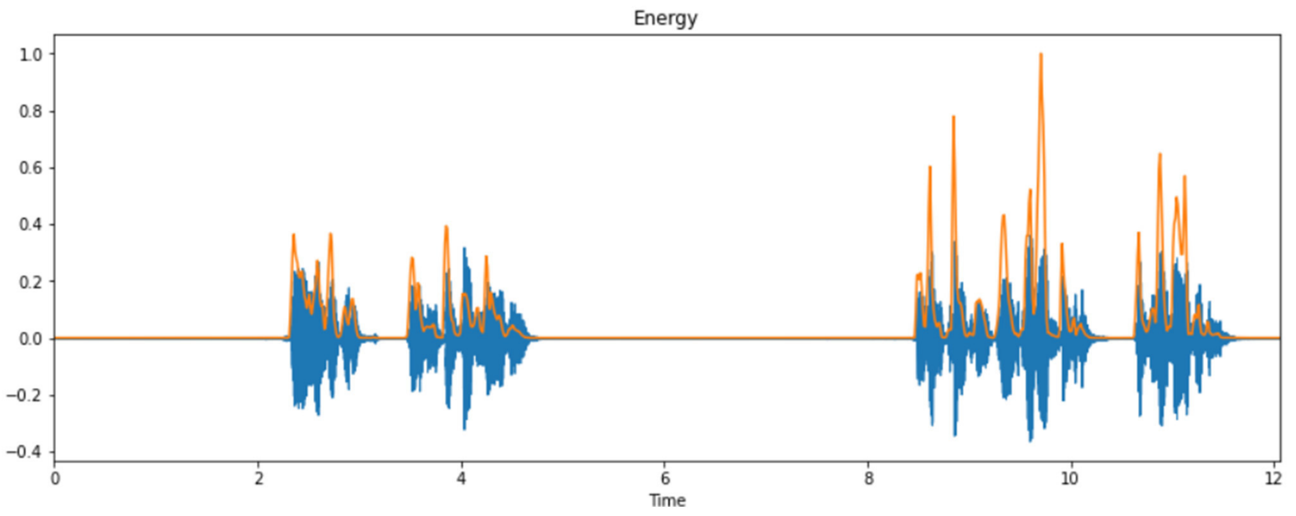


FIGURE 1 | Infant cry acquisition setup and analysis.

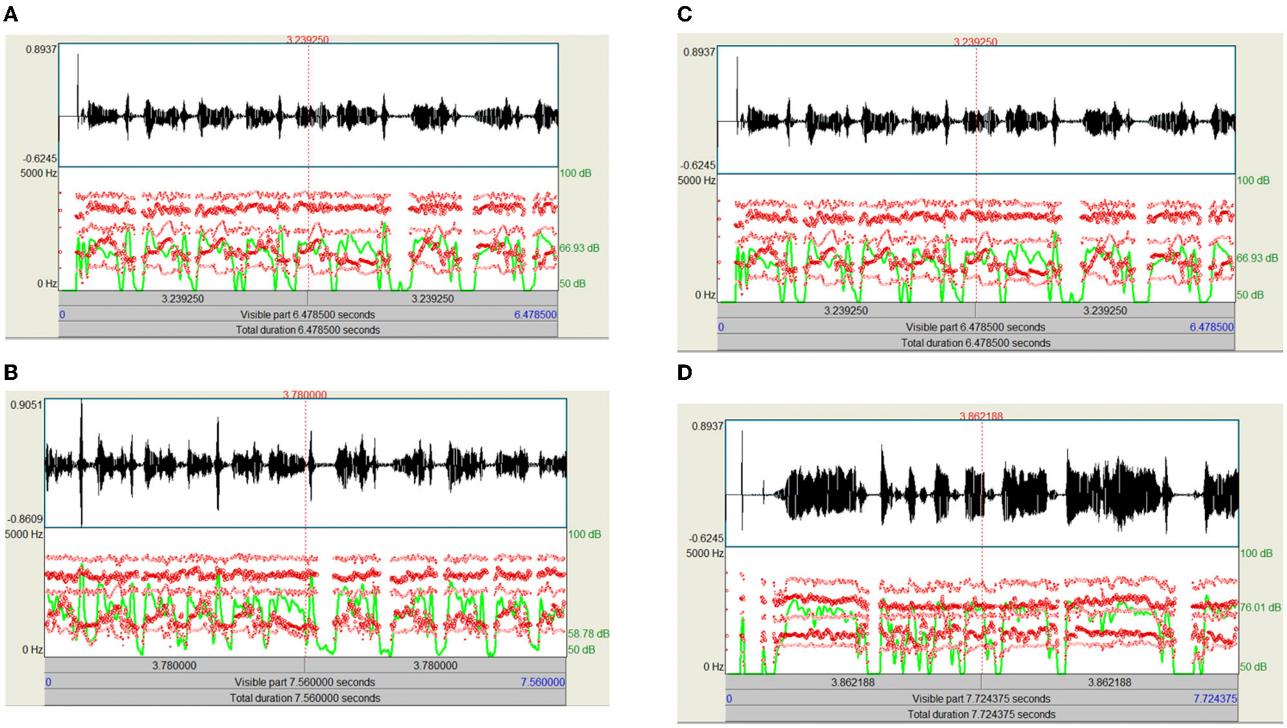


FIGURE 2 | (A) Audio signal of a sample baby cry (diaper) with intensity (green) and formant information (red), **(B)** audio signal of a sample baby cry (hunger) with intensity (green) and formant information (red), **(C)** audio signal of a sample baby cry (pain) with intensity (green) and formant information (red), and **(D)** audio signal of a sample baby cry (sleep) with intensity (green) and formant information (red).

focuses were considered to figure out the Fourier change. It accomplished a deep convolutional neural organization called AlexNet with a few improvements to group the recorded newborn child cry. To work on the viability of the previously mentioned neural organization, stochastic gradient descent with momentum (SGDM) was used to perform the calculation.

The authors in (6) obtained and broke down sound elements of infant's cry signals in schedule and recurrence areas. In view

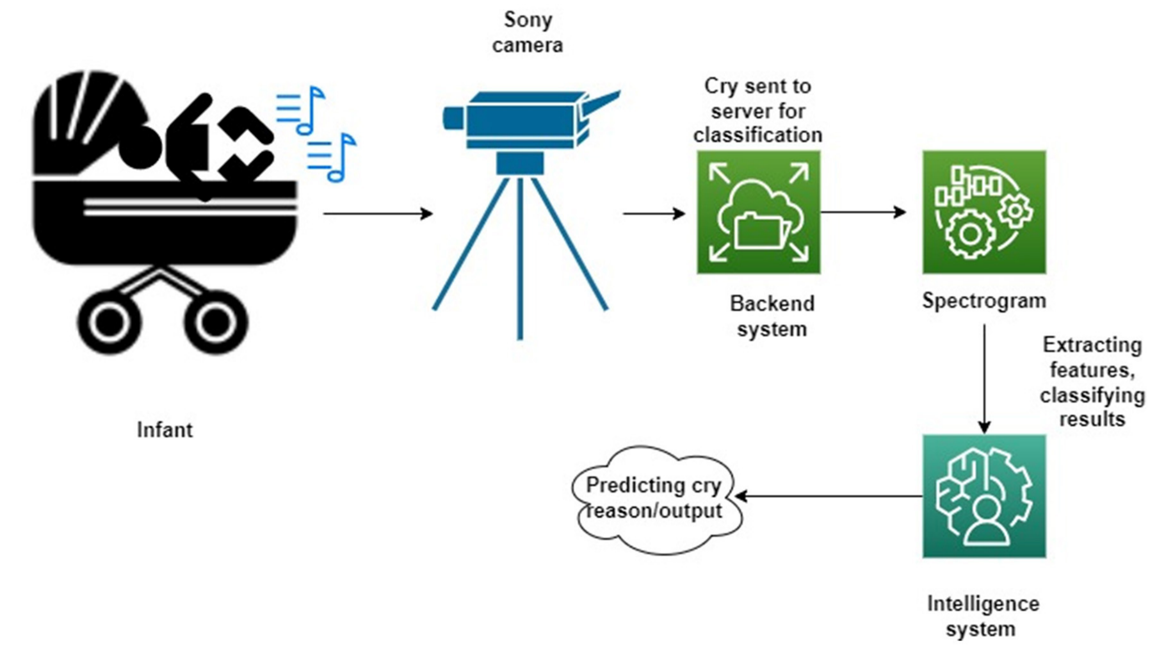


FIGURE 3 | Audio signal of a sample baby cry.

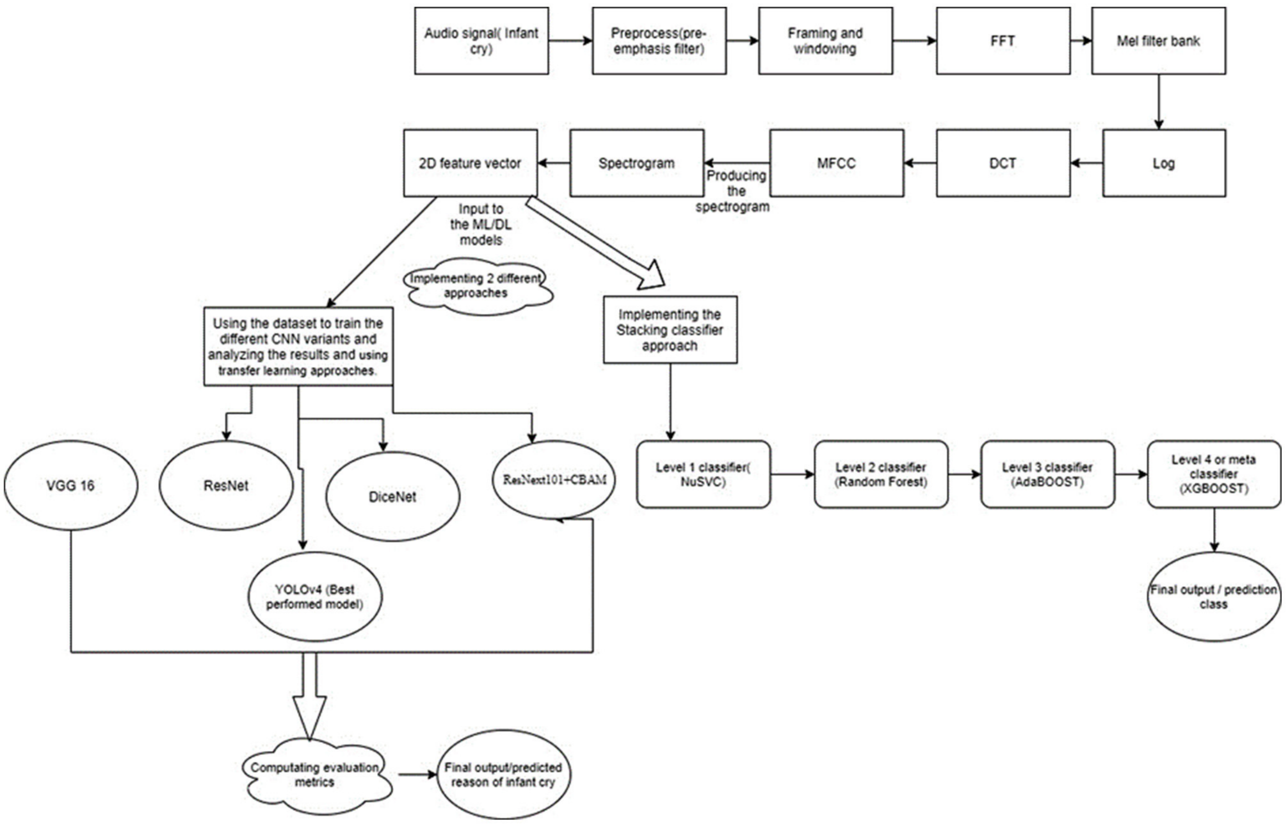


FIGURE 4 | Workflow of the research techniques used in this work.

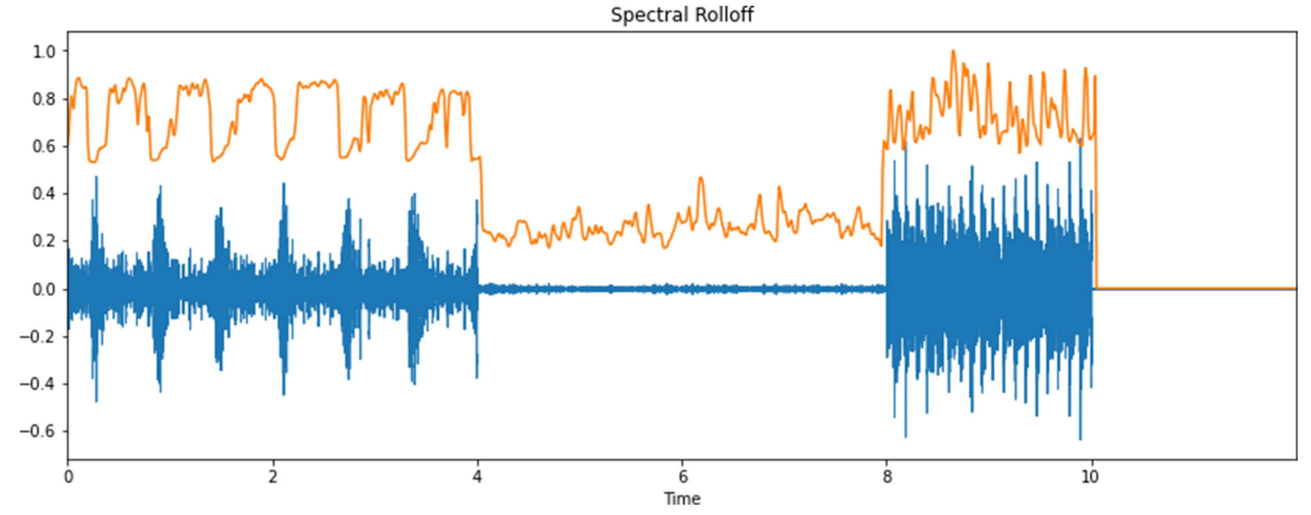


FIGURE 5 | Spectral roll-off of an infant cry's audio signal.

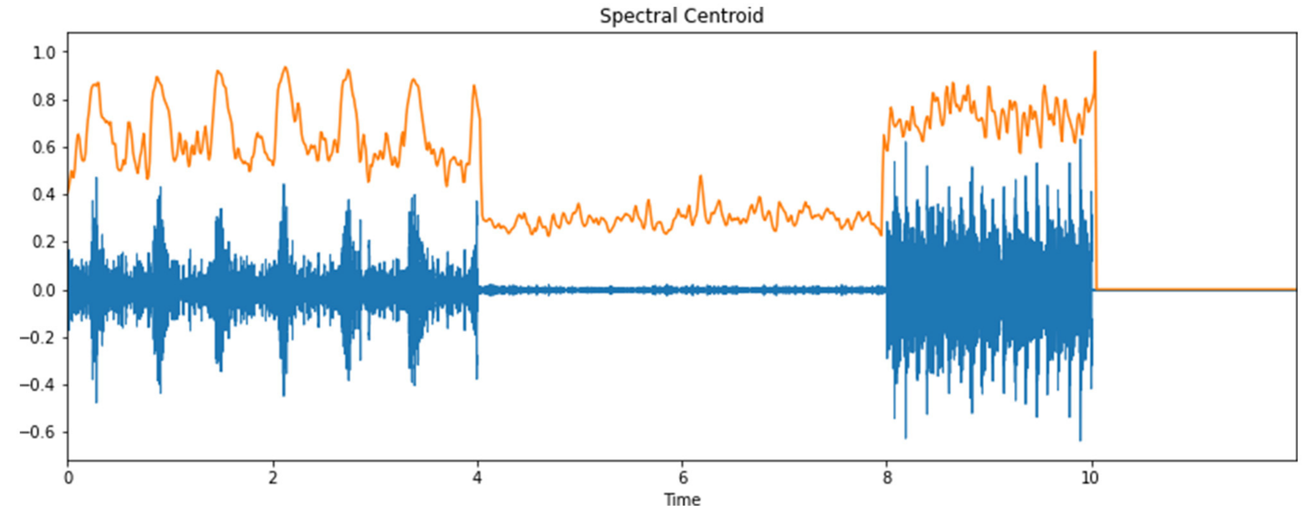


FIGURE 6 | Spectral centroid of an infant cry's audio signal.

of the connected elements, we can arrange cry signals to clear cry implications for cry language acknowledgment. Highlights separated from sound component space incorporate linear predictive coding, linear prediction cepstral coefficients, Bark frequency cepstral coefficients, and MFCCs. Packed detecting method was used for characterization, and useful information was used to plan further and confirm the proposed approaches. Tests showed that the proposed infant's cry detecting approaches offer accurate and promising outcomes.

The work in (7) portraying the advancement of significant information innovation, anticipating clients' buying goals through precise information of their buying practices has turned into a fundamental system for organizations to perform accuracy promotion and increase deal volume. The information of clients' buying behavior is described by an enormous sum, significant changeability, and long haul reliance. Along these lines, the bidirectional long short-term memory (BiLSTM) model is used in this article to examine the client's buying behavior. First, the model accepts client ID as the benchmark of grouping, catching the variance law of the client purchase volume and completely mining the drawn-out reliance of client's buying behavior. Second, the BiLSTM model adaptively extricates highlights, figures out the "start to finish" forecast of client's buy behavior, and diminishes the design subjectivity. This article checks the viability of this strategy depending on the genuine client buying behavior informational indexes. The investigation results show that the BiLSTM technique has high precision in examining the client's buying behavior.

The significant goal of this exploration work (18) was to introduce another procedure to recognize cancer. The

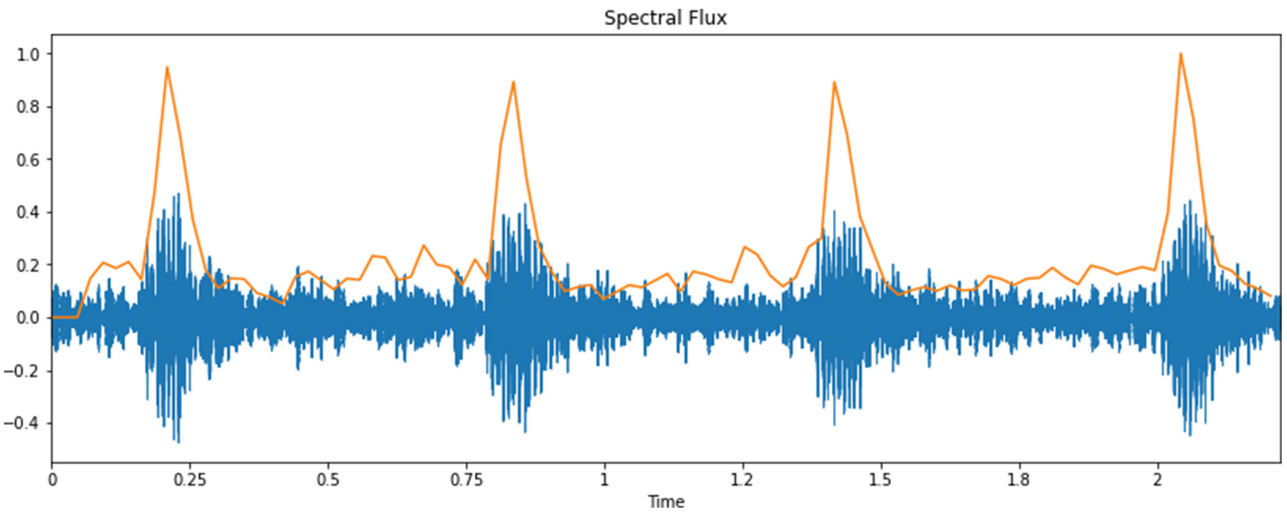


FIGURE 7 | Spectral flux of an infant cry's audio signal.

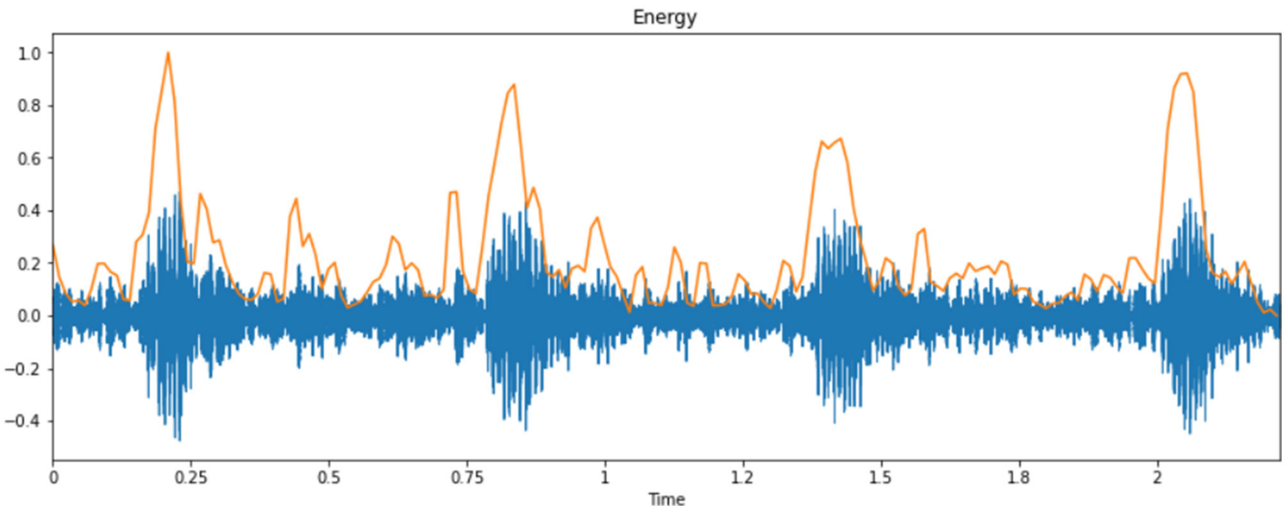


FIGURE 8 | Energy of a sample cry spectrogram.

proposed engineering precisely divided and characterized harmless and dangerous cancer cases. Diverse spatial area techniques to improve and precisely divide the information pictures were applied. Also, AlexNET and GoogLeNet were used for characterization, wherein two score vectors were acquired. Further, both score vectors were melded and provided to many classifiers alongside the Softmax layer. Assessment of this model is done on top medical image computing and computer-assisted intervention (MICCAI) challenge datasets, i.e., multimodal brain tumor image segmentation 2013, 2014, 2015, and 2016 and ischemic stroke lesion segmentation 2018 separately.

The work in (30) emphasized the complete exploration plans to arrange baby's cries into their social characteristics by utilizing evenhanded and insightful AI approaches. Toward this objective, the authors have considered customary AI and later profound learning-based models for child cry arrangement using acoustic elements, spectrograms, and a mix of the two. They have performed a point-by-point experimental review of the open access corpus and the CRIED dataset to feature the adequacy of fitting acoustic elements, signal processing, or AI procedures for this purpose. Major work was done by presuming that acoustic elements and spectrograms together will bring better outcomes. As a side outcome, this work additionally underscored the test of a deficient child cry data set in displaying baby's behavioral attributes.

This study (31) investigates a neural transfer learning way to create precise models for recognizing babies that have

TABLE 1 | Data distribution for different classes.

Class	Number of training samples	Number of testing samples
Sleep	12,820	3,205
Hunger	12,980	3,245
Pain	14,740	3,685
Diaper	14,204	3,551

TABLE 2 | Layer architecture of VGG16.

Layers	Size
A doubly linked convolution layer	It consists of 64 channels—each consisting of a kernel of size 3 × 3
A Maxpool layer	It consists of a kernel having a pool of size 2 × 2 and a stride of size 2 × 2
A doubly linked convolution layer	It consists of 128 channels, each of size 2 × 2
A Maxpool layer	It consists of a kernel having a pool of size 2 × 2 and a stride of size 2 × 2
A triply linked convolution layer	It consists of 512 channels, and each channel has a kernel of size of 3 × 3
A Maxpool layer	It consists of a kernel having a pool of size 2 × 2 and a stride of size 2 × 2
A triply linked convolution layer	It consists of 512 channels, and each channel has a kernel of size of 3 × 3
A Maxpool layer	It consists of a kernel having a pool of size 2 × 2 and a stride of size 2 × 2

TABLE 3 | Parameters—VGG16.

Layer type	Output size	Parameter
Input layer	1 × 256 × 40	0
Convulated 2D layer	128 × 256 × 40	1,280
Batch normalization layer	128 × 256 × 40	512
Activation layer	128 × 256 × 40	0
2D Maxpool layer	128 × 256 × 8	0
Dropout layer	128 × 256 × 8)	0
bidirectional_1	Bidirection (None, 256, 32)	55,488
permute_1 (Permute)	(None, 256, 128, 2)	0
bidirectional_2 (Bidirection)	(None, 256, 32)	12,480
time_distributed_1 (TimeDist)	(None, 256, 32)	1,056
conv2d_3 (Conv2D)	(None, 128, 256, 4)	147,584

TABLE 4 | Hyperparameters settings—VGG16.

Name	Settings
Nodes used per trained layers	1,024
Epochs	300
Optimizers	Adam and SGD
Lr—learning rate	0.0001
Lrd—learning rate decay	Yes
Drop out	0.25

experienced perinatal asphyxia. Specifically, the authors have investigated the speculation that portrayals obtained from grown-up discourse could educate and further develop execution based on models created on newborn baby discourse. Their tests show that models depending on such portrayal moves are resilient to various kinds and levels of commotion, just as to flag misfortune on schedule and recurrence areas. The work analyzes the exhibition of a residual neural organization. Their ResNet model was pretrained on a few discourse assignments in characterizing perinatal asphyxia. Among the implemented models, the model for the word recognition task performed the best, recommending that the varieties learned for this undertaking are generally closely resembling and helpful to their objective assignment. The support vector machine prepared straightforwardly on MFCC highlights ended up being a solid benchmark and, assuming fluctuation in forecasts was of concern, a favored model.

In this article (32), the authors present a safe medical care framework that performs an acoustic examination of messy boisterous baby cry signs to concentrate and gauge specific cry attributes quantitatively and group strong and weak babies indicated only by their cries. In the lead of this infant cry-based indicative framework, the unique MFCC highlights as well as static MFCCs are chosen and removed for both expiratory and inspiratory cry vocalizations to deliver a discriminative and instructive component vector. Then, the authors made a remarkable cry design for each cry vocalization type and neurotic condition by presenting a clever thought utilizing the boosting mixture learning (BML) technique to infer either sound or pathology subclass models independently from the Gaussian mixture model-universal background model. Also, a score-level combination of the proposed expiratory and inspiratory cry-based subsystems was performed to settle on a more dependable choice. The trial results show that the adjusted BML strategy has lower error rates than the Bayesian methodology when considered as a kind of perspective technique.

DATA PREPARATION

Data processing is the first step in designing the model. The dataset comprises several cry signals that have been taken from the National Taiwan University Hospital—Yunlin Branch for research. These signals correspond to the reason for the infants' cry. These four reasons are hunger, pain, tiredness, and diaper. The major steps involved in the preprocessing phase are data cleaning, audio scaling/normalizing, framing, windowing, and later spectrogram formation. In this article, the baby cries were acquired from the Division of Obstetrics

TABLE 5 | Hyperparameter settings—YOLOv4.

Name	Settings
Batch size	10
Epochs	380
Optimizers	Adam
Lr—learning rate	0.001
Lrs—learning rate schedule	Learning rate increases by 0.1
Drop out	0.20
Early stopping	There is a decline in the validation loss for 50 epochs
Momentum	0.924
Weight decay	0.0005
Anchor_t (anchor-multiple threshold)	4.0
F1_gamma	0.2

and Gynecology at the National Taiwan University Hospital Yunlin Branch, Taiwan. They encountered no inconveniences during birth, and their introduction to the world loads, ages, and gestational ages was without neurotic discoveries. Their age was somewhere between 1 and 10 days. Also, all of the baby cries were recorded using a SONY HDR-PJ10 HD computerized video recorder with an underlying mouthpiece, Sony Corporation, Tokyo, Japan. The cries of newborn children were recorded in a supine position. The amplifier was held around 40 cm from the newborn child's mouth. Each recording document lasts somewhere between 10 and 60 s. **Figure 4** illustrates the workflow of the proposed models.

Preprocessing

In this phase, all of the input infant cry audio files are preprocessed. It basically involves applying a double loop on the entire dataset. Here, the noises and other disturbances from the audio files are also removed to get better results. This is because if we train our model on the dataset, which contains noise or unwanted data, then the model might have some significant characteristics and might make wrong interpretations. The algorithm basically converts the input audio files to various keyframes, which are then converted into a series of feature vectors. So, every audio keyframe would have a feature vector, which when combined would cause a 2D feature vector containing the numerical values of all of the relevant features of the audio signal. MFCC is utilized to separate the interesting components of discourse tests. It addresses the momentary force range of human discourse. The MFCC method uses two sorts of channels:directly dispersed channels and logarithmically separated channels. To catch the phonetically significant attributes of discourse, the sign is communicated on the Mel-recurrence scale. The Mel scale is mostly established on examining the pitch or repeat saw by the human. Subsequently, the scale is wrapped up into the units of Mel. The Mel scale is commonly an immediate preparation under 1,000 Hz and logarithmically isolated over 1,000 Hz. MFCC involves six computational advancements (3). This progression likewise includes passing the sign through the channel, which highlights higher recurrence in the frequency band. It likewise underlines the extent of some higher frequencies in regard to other lower frequencies to improve the energy. Librosa is a library available in Python that reads and processes the audio signals. **Figure 1** represents the sample audio signal of an infant's cry. **Figures 2A–D** illustrate the audio signal of a sample baby cry for diaper, hunger, pain, and sleep, respectively. **Figure 3** illustrates the infant cry acquisition setup and analysis.

Frequency Feature Extraction

It pays attention to the recurrence parts of the sound sign. Signals are changed from the time–space to the frequency area using the Fourier transform. Band energy proportion, spectral centroid, and otherworldly motion are models.

Spectral Roll-Off

Roll-off is the steepness of a transfer function with frequency, particularly in electrical network assessment, and most especially with regard to direct circuits in the advancement between a passband and a stopband. It is mostly applied to the expansion loss of the network but can, on a fundamental level, be applied to any appropriate function of frequency, and any technology, not just devices. It is customary to check roll-off as a part of logarithmic frequency; consequently, the units of roll-off are either decibels per decade (dB/decade), where a decade is a 10 times increase in frequency, or decibels per octave (dB/8ve), where an octave is a 2-fold increase in frequency.

Roll-off comes from the fact that in many networks roll-offs tend toward a constant gradient at frequencies well-away from the cut-off point of the frequency curve. Roll-off enables the cut-off performance of such a channel network to be diminished to a single number. Note that roll-off can occur with diminishing frequency as well as as increasing frequency, depending on the bandform of the channel being considered: for instance, a low-pass channel will roll-off with increasing frequency, but a high-pass channel or the lower stopband of a band-pass channel will roll-off with reducing frequency. For conciseness, this article depicts only low-pass channels. This is to be taken in the spirit of model channels; comparative guidelines may be applied to high-pass channels by trading articulations, for instance, "above cut-off frequency" and "below cut-off frequency. **Figure 5** illustrates the spectral roll-off of an infant cry's audio signal.

Spectral Centroid

The spectral centroid is a measure that shows the location of the "focal point of mass" of the spectrum. Perceptually, it has a strong association with the impression of "splendor" of a sound and hence is utilized to portray melodic tone. **Figure 6** portrays the spectral centroid of an infant cry's audio signal.

Spectral Flux

The spectral flux/transition is a helpful measure for differentiating signals whose spectrum changes slowly from signals whose spectrum changes rapidly. It has a lower incentive

TABLE 6 | Comparing the performance of top CNN models.

Variant name	Year (developed)	Category	Role	Parameters	F1 score	Error rate
VGG16	2014	Spatial exploitation and multiclass classification	Uses small-sized kernels and performs better for multiclass problems	140 M	∼72.1%	ImageNet: 27.3%
ResNeXt10 + CBAM	2018	Consideration	Exploits both feature map and spatial information	48.76 M	73.4%	26.6%
ResNet	2016	Depth in learning using multi-path technique	Provides a mapping-based skip connection	25.63 M	72.8%	27.2%
YOLOv4	2020	Spatial exploitation	Performs multiclass classification using pretrained weights—one of the major transfer learning models	60 M	75.2%	24.8%
DiceNET	2021	Dimensional based	It performs dimensional-based CNN, including height, weight, and depth	20 M	74.5%	25.5%

TABLE 7 | Comparing the merits and demerits of the top CNN variants.

CNN variant name	Merits	Demerits
VGG16	Provides the benefits of effective approachable fields in training It also introduces the concept of a simple and homogeneous topology between various audio key frames	Its computational power is dependent on the fully linked layers and is very expensive
ResNeXt101 + CBAM	Provides flexibility in depth and dimensions It has a consistency of maintaining a large amount of data flow between hidden layers, which contribute to information gaining attributes	As the number of feature vectors is very high in this model, the number of parameters and their values will also need to be increased, leading to a slow learning rate
YOLOv4	It is twice as fast in computing results than other top CNN variants In terms of accuracy and other metrics, it provides better performance	Requires a well-scaled and normalized dataset, as its hypertuning variables have restrictive abilities in some cases
ResNet	Error rate reduces for complex architectures The gradient problems are resolved (6)	It has a complex architecture. It has overfitting of hyperparameters
DiceNet	These convolutions use highly enhanced feature filters in every layer The dimension-wise vectors are mixed proficiently It provides high-accuracy multiclass classification (13)	It may take a lot of computational time in prediction

for the previous "slow" class of signals and a higher incentive for the later "quick" class of signals. **Figure 7** shows the spectral flux of an infant cry's audio signal.

TABLE 8 | Number of testing samples for each class.

Class	Number of training samples	Number of testing samples
Sleep	12,924	4,308
Hunger	14,517	4,839
Pain	14,115	4,705
Diaper	14,418	4,806

TABLE 9 | Hyperparameters selection for Nu SVC.

Hyperparameters	Value	Justification
C	150	This value basically determines the training speed of the algorithm. Setting a value less than 100 was taking more time, whereas values above 150 showed less learning by the model. On trying many approaches, C = 150 was considered the most suitable for the model
Degree	4	Due to the 2D feature vector and many attributes of the dataset, a degree of 4 was considered better. Any value above 4 showed a similar performance on consuming more time
Kernel	RBF	According to research, the RBF kernel is more suitable for stacking multiple classifiers (13)
Gamma	auto	This is the default value

These non-linear features help us analyze the patterns and frequency domains of the audio signals. The values derived from these feature vectors would help decide the appropriate methodologies suitable for the data preprocessing phase.

TABLE 10 | Hyperparameter selection for RF.

Hyperparameters	Value	Justification
*n*_estimators	700	It determines the number of trees required to build before making average predictions. The increase in value increases the performance and decreases the speed of training. So, up to 700, the model was performing with consistently increasing accuracy and speed, but after that, the speed started to decrease
max_features	Auto	This is the default value. It improves performance at each node. The model was showing degrading performances at other values
max_depth	50	It represents the depth of each tree in the model. The more the depth, the more the information will be gained from each tree, which will add on to the info_gain parameter of the model. However, values above 50 decreased the speed of the process
min_samples_leaf	0.005	It represents the minimum sample of dataset at each level. It is always beneficial to have fewer samples at each node to prevent complexity in learning

TABLE 11 | Hyperparameter selection for AdaBoost.

Hyperparameter	Value	Justification
*n*_features	40	It represents the number of features seen during the model fit. The value was according to the number of attributes in the dataset
*n*_estimators	200	It represents the number of estimators at which the boosting terminates. Up to the value of 200, the boosting was enhanced without affecting the learning rate. However, above 200, the boosting rate became constant, leading to a decrease in the learning of the model
Learning rate	1	This is the default value. The model performed better at the initial value
Algorithm	SAMME.R	It performs faster than SAMME and achieves lower test errors with fewer boosting iterations
*N*_classes	4	The number of classes was 4

TABLE 12 | Hyperparameter selection for XGBOOST.

Hyperparameter	Value	Justification
Num_boost_round	150	It represents the number of rounds to build the model. Its optimal high value depends largely on other parameters, that is why its value is kept relatively lower to avoid changing it as and when we use other parameters
Early_stopping_rounds	20	It should generally be quite low so that we do not have to worry about improving the accuracy of the model as it has early stopping
*N*_fold	6	It is the number of folds required for cross validation. Initially, it was set to 5, and later, on observing the improvement in the increase in value, it was changed to 6. However, above 7. it did not incline toward the learning curve
Metrics	MAE	The mean absolute error allows us to compare estimates of different sequences in different scales

Framing and Windowing

Framing is the process of segmenting the entire audio signal into small frames of a fixed size. A single audio file is converted into *N* number of frames so that the algorithms work better. Because a single audio file would contain many feature vectors, this would become highly complicated to derive the results. The adjacent frames are also separated by a specific gap so that they do not overlap. Setting a space of 256 or 512 is appropriate for the MFCC process to avoid overlaps.

Fast Fourier Transform

It performs the task of conversion of each single frame of the audio keyframes from the phase of the time domain to the frequency domain, as most of the calculations are performed in this domain.

Mel Filter Bank

There are many frequencies ranges in fast Fourier transform. The most significant one is enormously inclusive, and voice signals do not trail the unswerving gauge.

Discrete Cosine Transform (DCT)

This process involves a cycle that can transform the log Mel spectrum into time domain. The outcome generated by the process is defined as MFCC. The set of coefficients is called acoustic vectors. According to this concept, all of the information values are transformed into a segment of auditory vectors. **Figure 8** represents the energy of the cry spectral waveform.

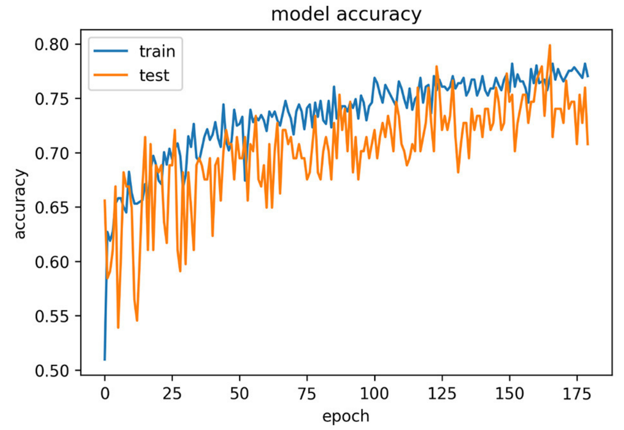


FIGURE 9 | Epochs vs. accuracy graph for the VGG16 model.

APPLYING CNN VARIANTS ON THE CRY SPECTROGRAMS

After data preprocessing and generation of the spectrograms and tabular 2D dataset, we applied the various CNN variants. The expected workflow comprises the following steps.

Data Distribution

First, the dataset was split into training and test sets in the ratio of 80:20. In total, 80% of the entire dataset that comprises the feature vectors of the spectrograms was used for training the different models. The training, testing, and validation accuracies of both the proposed models were compared and analyzed.

Table 1 consists of the number of testing samples for each prediction class. The data distribution sampling is the same for YOLOv4 and VGG16.

Implementing VGG16

The main reason for using VGG16 was its merits in predicting the multiclass outputs of the model (6). According to various research studies, it is one of the best CNN variants that utilizes the transfer learning approaches for multiclass classification problems.

Table 2 represents the information about the layers and the size of the kernels. **Table 3** presents the summary of the VGG16 model, and **Table 4** lists the hyperparameter settings of the VGG16 model.

After performing a lot of comparisons among the various hyperparameter selected values, the proposed VGG16 model was performing at its best on the above-mentioned ones. The observations made on these selected values can provide deep insight into the performance of the model. From this, we can infer that the nodes per trained layers and drop-out values contributed in the most significant way to the training of the entire model. The model was trained on 300 epochs. This was the tuned parameter; the epochs above 300 showed a degrading performance.

Implementing YOLOv4

The accuracy of CNN should be determined through many provisions. The pragmatic testing of such element blends on enormous datasets is needed, similar to the hypothetical pursuit of the outcomes. A few viewpoints, for example, cluster standardizations, are simply enough to explicit inadequately spaced datasets; however, many others are rational to a larger part of representations, assignments, and datasets. To accomplish outstanding results, this proposed work utilizes original provisions, for example, Mish actuation and DropBlock regularization are some of them joining some of them.

Hyperparameter Selection for YOLOv4

The Yolov4 model has 53 convolution layers of size 1 × 1 and 3 × 3 kernels. It also has 30 hyperparameters for improving the training and the performance of the model. Each hyperparameter has a specific potential to create a huge impact on the overall training and learning of the model. The following hyperparameters have been selected after understanding the architecture and learning the observations from various research articles using YOLOv4 (16). After many tests and evaluations, the final YOLOv4 model was subjected to the following hyperparameters. **Table 5** represents the hyperparameter settings of the YOLOv4 model.

Implementing FastGCN

FastGCN is a graph-based model with a purposeful diagram model for semi-supervised learning approaches. Moreover, on account of the recursive neighborhood development of transverse layers, preparing huge and thick diagrams is time- and memory-consuming. This is a hybrid model, and its hyper tuning methodologies are the fundamentals of implanting capacities. Accordingly, it utilizes Monte Carlo strategies to dependably surmise the integrals, which prompts the group preparation plan.

Comparing the Performance of the Top CNN Variants

The performances of top CNN models are compared in **Table 6**.

The performance of these CNN variants was not up to the mark, as these could not handle some outliers even after performing hypertuning. All of these models had low computational power and did not yield high accuracy. It was required to go with another approach that would yield better accuracy and have fast computational power. These are some approaches that combine many ensemble-based algorithms in a stacked form. These algorithms have a higher chance of training the model at a lesser computational time and provide better performance. **Table 7** presents the merits and demerits of the top CNN variants.

TABLE 13 | Evaluation metrics for the VGG16 model.

Class	*n* (truth)	*n* (classified)	Accuracy (%)	Precision	Sensitivity	Specificity	F1 score
Class 1 (sleep)	3,205	3,653	84.5	0.81	0.88	0.856	0.84
Class 2 (hunger)	3,245	3,609	86.3	0.82	0.93	0.902	0.86
Class 3 (pain)	3,685	3,367	84.2	0.78	0.86	0.855	0.83
Class 4 (diaper)	3,551	3,057	77.8	0.74	0.83	0.81	0.77

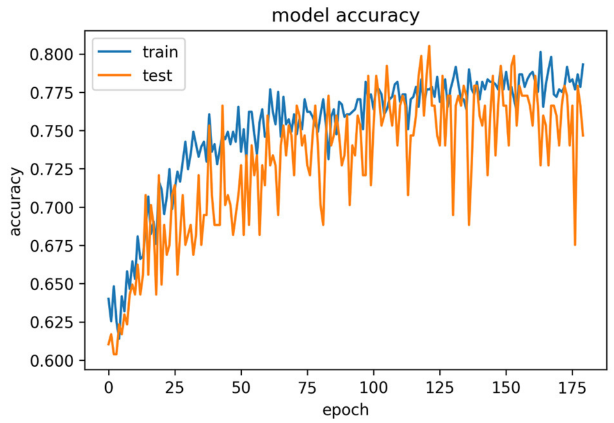


FIGURE 10 | Epochs vs. accuracy graph for the YOLOv4 model.

MULTISTAGE HETEROGENEOUS STACKING ENSEMBLE CLASSIFICATION MODEL FOR HIGHER PREDICTIVE PERFORMANCE

The stacking approach is an ensemble learning technique that uses various classifiers at different levels to produce better outputs. The data are trained at different levels by different classifiers. Since our dataset contains spectrograms of different audio frames, it is convenient to use different classifiers as any one specific classifier will not give the correct output. This was the main reason the top CNN variants were not performing up to the mark and taking a lot of time in training (6).

The following proposed model consists of four different levels. Each level corresponds to one classifier, and the output produced by each level will contribute to the performance of the next levels.

Motivation to Use the Stacking Approach

The major benefit of using the stacking approach is that it can use the features and benefits of the different boosting algorithms to a better extent. There are many boosting algorithms like XGBOOST and AdaBoost that perform much better compared to neural networks and other deep learning models. This is because top CNN variants were performing slowly due to the large dataset. Their process involves various recursive backtracking methods and solves overlapping subproblems. The highly advanced boosting algorithms avoid these recursive calls to some extent.

Implementation

The major objective behind the working of the proposed model is based on the characteristics of the top four classifiers. It is important to choose such classifiers that would increase the accuracy of the model by consuming less computational resources, time, and power. The entire model consists of four levels. Each level corresponds to a classifier. Initially, the spectrogram is fed as input to the first level classifier. Each level has a share of 25%. The classification results of the first classifier are stored in a cache and taken forward to the next level classifier. The next level classifier learns and gets trained on the basis of the prediction outputs of the previous level classifiers. Due to this, they will always try to improve the accuracy of the model as compared to the previous stages. The third and fourth classifiers play a major role in determining the performance and accuracy of the model.

Model 1

This model uses SVC, multilayer perceptron (MLP), NuSVC, and RF classifiers. These classifiers are accessible in Scikit-learn. The most reasonable tuning will be applied to this model. To put it obtusely, if some classifiers are underperforming, a pile of them would most likely be garbage as well. To produce appropriate results, the hyperparameters of each classifier are fixed. With this model, an accuracy of 82–83% was achieved.

Model 2 (Improvement of Model 1)

This model will support ensemble learning algorithms like XGBOOST, AdaBOOST, RF, and NuSVC at four different levels. The performance and the accuracy produced by this model were comparatively better than the previous approach. The main reason for choosing the boosting algorithms over the traditional machine learning algorithms was that these algorithms are highly flexible and support the parallel processing technique. Parallel processing is very important when we have a large dataset because it avoids computing the results in a recursive manner. Unlike the deep learning models, they are faster than gradient boosting. They can perform cross-validation after every iteration (5). XGBoost utilizes choice trees as base students, joining numerous powerless students to make a solid student. Accordingly, it is alluded to as an outfit learning technique since it utilizes the yield of many models in the last forecast. XGBoost or extreme gradient boosting may be very well-placed into different use cases like positioning, ordering, relapsing, and client-characterized forecast issues. It is an ideal mix of programming and equipment advancement methods to yield common results by utilizing fewer

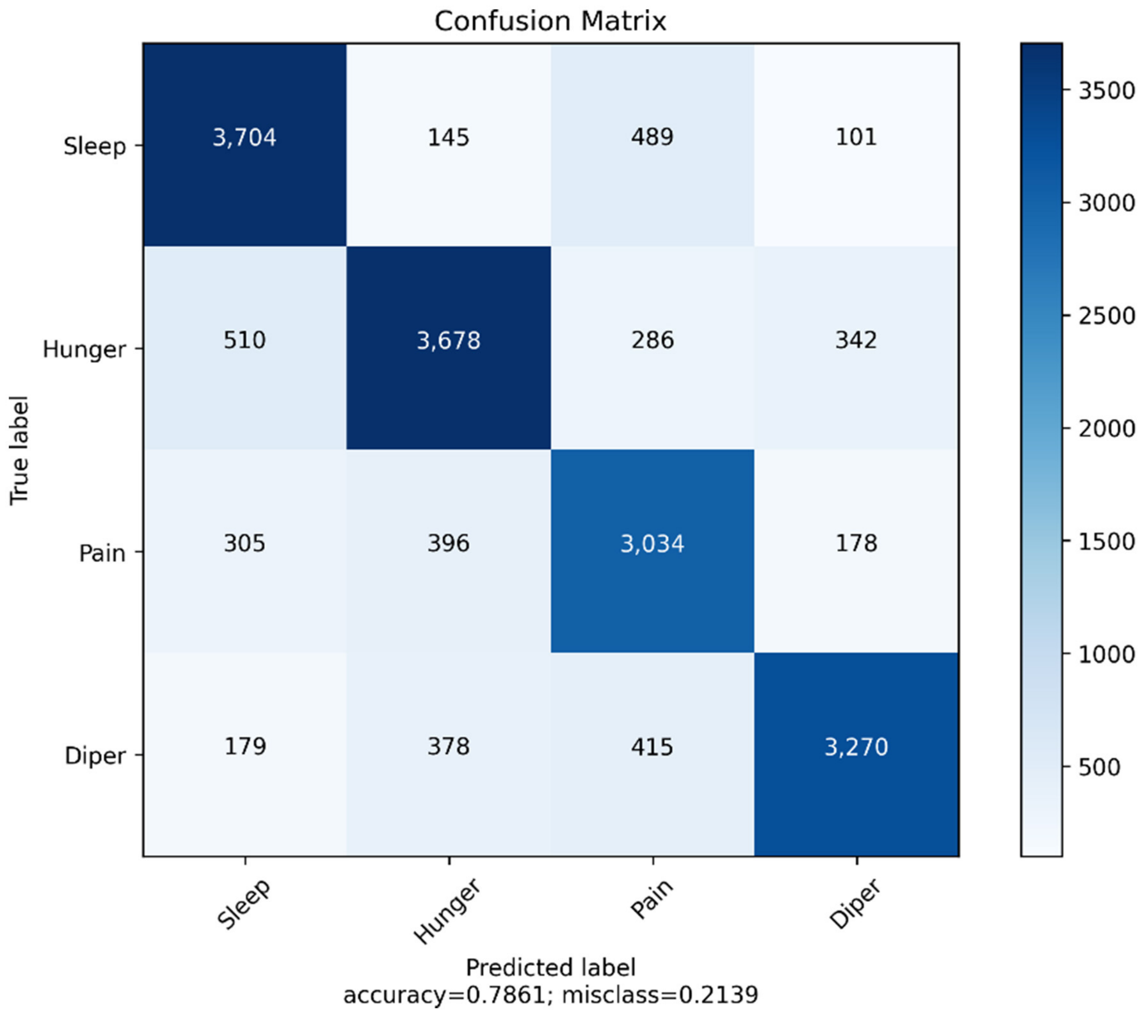


FIGURE 11 | Confusion matrix for the YOLOv4 model.

TABLE 14 | Evaluation metrics for the YOLOv4 model.

Class	*n* (truth)	*n* (classified)	Accuracy (%)	Precision	Sensitivity	Specificity	F1 score
Class 1 (sleep)	3,205	3,678	87.8	0.83	0.90	0.83	0.85
Class 2 (hunger)	3,245	3,704	86.7	0.81	0.92	0.88	0.83
Class 3 (pain)	3,685	3,270	82.3	0.77	0.85	0.81	0.80
Class 4 (diaper)	3,551	3,034	75.6	0.74	0.84	0.79	0.77

processing assets in the briefest measure of time. With this approach, an accuracy of ∼87% was achieved with 300 epochs.

Data Distribution

The dataset is split into training and testing sets in the ratio of 75:25. This the distribution of the testing samples of all the classes. **Table 8** represents the number of testing samples for each class.

Hyperparameter Tuning in XGBOOST and AdaBoost

XGBoost is a highly advanced implementation of the gradient boosting algorithm. It provides a lot of flexibility to enhance the performance of the model. The main advantage offered by XGBoost is that it helps improve the training of the model. Also, the proposed deep learning models had some overfitting instances. These could not be removed or reduced due to the limitations in the parameters of the various hidden layers. The training speed can be increased by increasing the max_Depth value to some extent. We had to implement some trail-error methods to fix the best value of max_Depth. This variable has a huge impact on the training speed of the model. According to a research article (8), a learning rate of 0.0001 was considered to be the most optimal for the initial stage. Later, the learning rate was increased to 0.001 on making some relevant observations in training accuracy. The parameter named objective was set to multi:softmax. The main reason for considering this value was the structure of our dataset. The dataset had multiclass output; hence, using multi:softmax was considered appropriate. Since XGBoost can predict the output much better than AdaBoost in terms of accuracy and precision, it is used in the last or fourth level of the proposed architecture.

The AdaBoost algorithm has a better learning rate and better performance among multiple *k*-fold cross validations. When fitting the last model, it could be attractive to either expand the number of trees until the difference of the model is decreased across rehashed assessments or to fit numerous last models and normalize their expectations. A significant hyperparameter for AdaBoost is *n*_estimator. By changing the number of base models or frail students regularly, we can change the precision of the model. The number of trees added to the model should be high for the model to function admirably, frequently hundreds, if not thousands. After all, the more the number of feeble students, the more the model will change from being high one-sided to low one-sided. The learning rate relies profoundly on the number of *n*_estimators. Naturally, it is set to 1, yet it tends to be expanded or diminished depending upon the assessors utilized. For the most part, for countless *n*_estimators, we utilize a more modest benefit of the learning rate. For instance, whenever our powerless classifier gets the opportunities of right expectations just somewhat more than arbitrary theory, the learning rate is 0.5. It is normal to utilize a more modest benefit of the learning rate going somewhere in the range of 0 and 1, as 0.1 and 0.001 because, in any case, it brings about the issue of overfitting.

Hyperparameters of classifier 1 (NuSVC) are represented in **Table 9**.

Hyperparameter Selection for Random Forest

The hyperparameters selected for RF are illustrated in **Table 10**.

Hyperparameter Selection for AdaBoost

The hyperparameters selected for AdaBoost are illustrated in **Table 11**.

Hyperparameter Selection for XGBOOST

The hyperparameters selected for XGBOOST are illustrated in **Table 12**.

RESULTS AND DISCUSSION

Results for the VGG16 Model

Figure 9 represents the epoch vs. accuracy graph for the VGG16 model using the transfer learning approach. Further, **Table 13** represents the evaluation metrics for the VGG16 model.

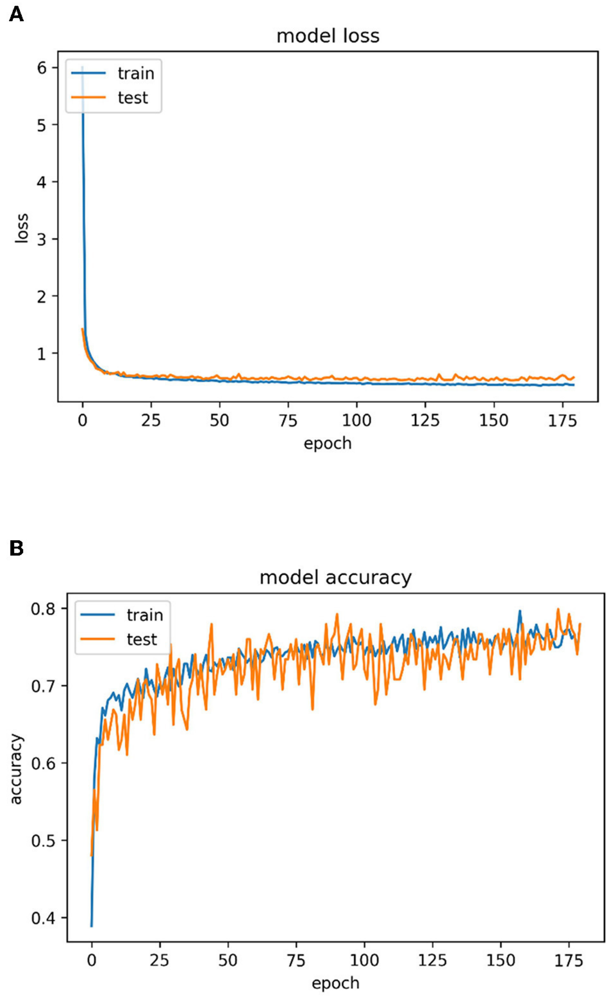


FIGURE 12 | (A) Epoch vs. loss graph for multi-stage heterogeneous stacking ensemble classification model 2 and **(B)** epoch vs. accuracy for multistage heterogeneous stacking ensemble classification model 2.

From the above table, we can analyze that the model performs the best in classifying the cries due to hunger. The model has an imbalanced nature in prediction metrics of the other three classes. This is one of the limitations of the model as the model classifies hunger cries more even though the samples are fewer than the pain and diaper cries. Hence, the model needs to have a balance in prediction metrics among all of the classes. Tragically, there are two significant disadvantages with VGGNet: It is agonizing to prepare. The organization's engineering loads themselves are enormous (concerning circle/data transmission). Because of its profundity and number of completely associated hubs, VGG16 is over 533MB. This makes sending VGG a tedious errand. VGG16 is utilized in many profound learning picture arrangement issues; in any case, more modest organization models are frequently more alluring (like SqueezeNet, GoogLeNet, and so on). Yet, it

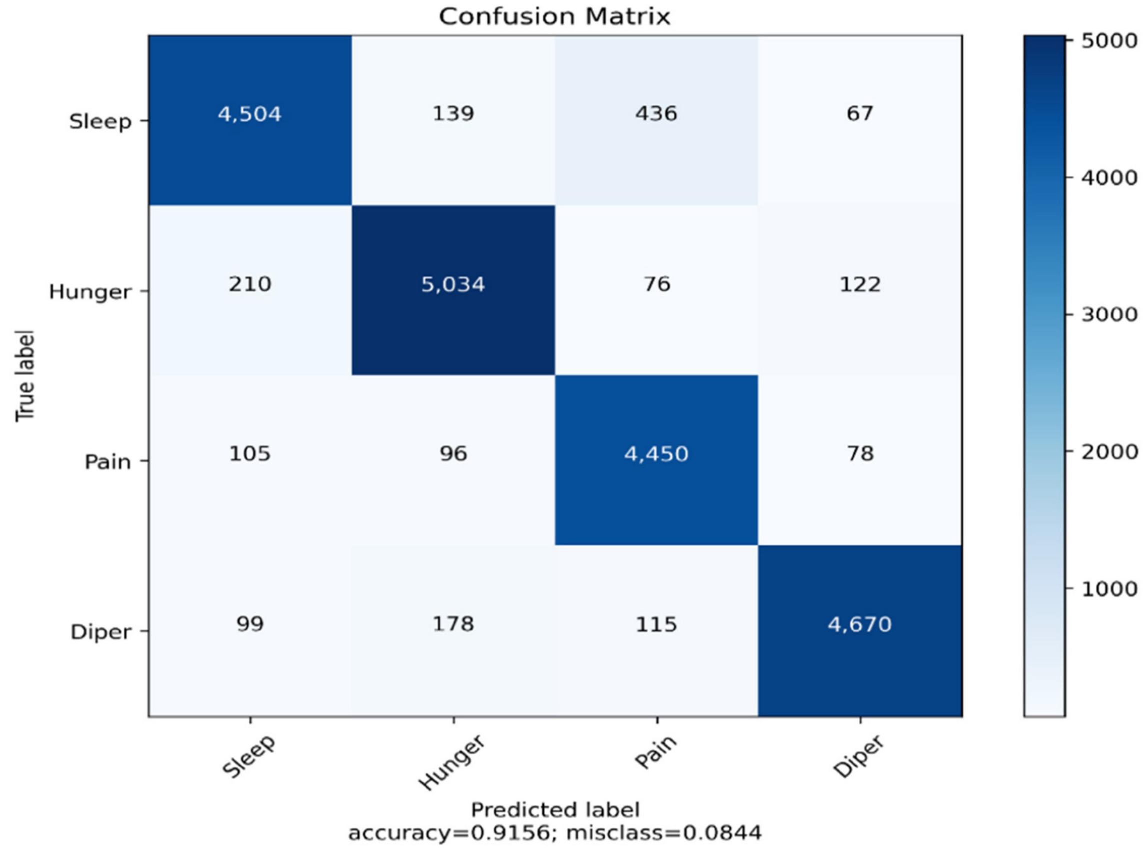


FIGURE 13 | Confusion matrix of multi-tage heterogeneous stacking ensemble classification model 2.

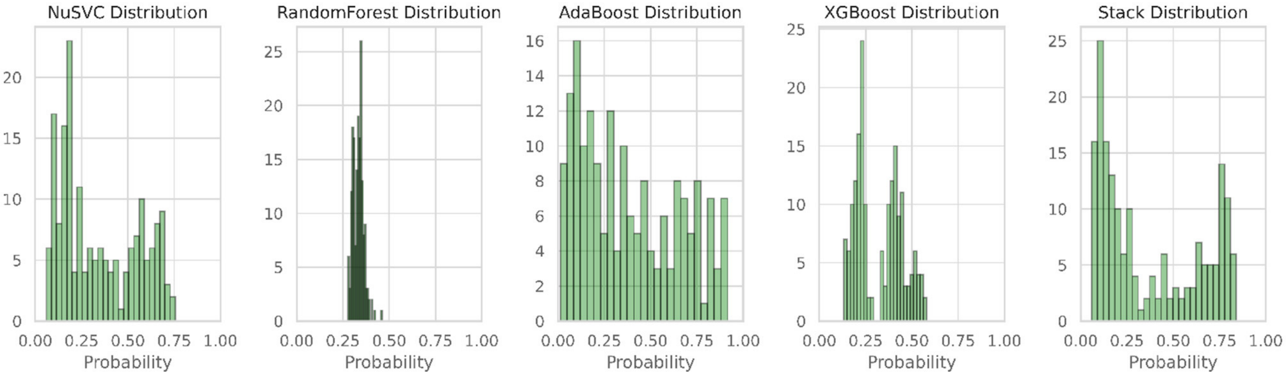


FIGURE 14 | Probability distributions of the four classifiers and multistage heterogeneous stacking ensemble classification model 2.

is an extraordinary structure block for learning purposes as it is not difficult to carry out.

Results for the YOLOv4 Model

Figure 10 represents the epochs vs. accuracy graph for the YOLOv4 model. From this graph, we can observe the improvement in results of the YOLOv4 model compared to those of the VGG16 model.

We have generated the confusion matrix and the evaluation metrics for the YOLOv4 model. The main objective of developing the confusion matrix is to observe the number of true positive predictions for every class. The output consists of four different classes. They are sleep, hunger, pain, and diaper. **Figure 11** illustrates the confusion matrix for the YOLO v4 model. **Table 14** represents the evaluation metrics of the YOLOv4 model.

TABLE 15 | Multistage heterogeneous stacking ensemble classification model 2.

Class	Accuracy (%)	Precision	Sensitivity	Specificity	F1 score
Class 1 (sleep)	94.5	0.87	0.88	0.87	0.93
Class 2 (hunger)	96.3	0.95	0.93	0.89	0.95
Class 3 (pain)	91.2	0.92	0.86	0.80	0.91
Class 4 (diaper)	92.8	0.94	0.83	0.76	0.90

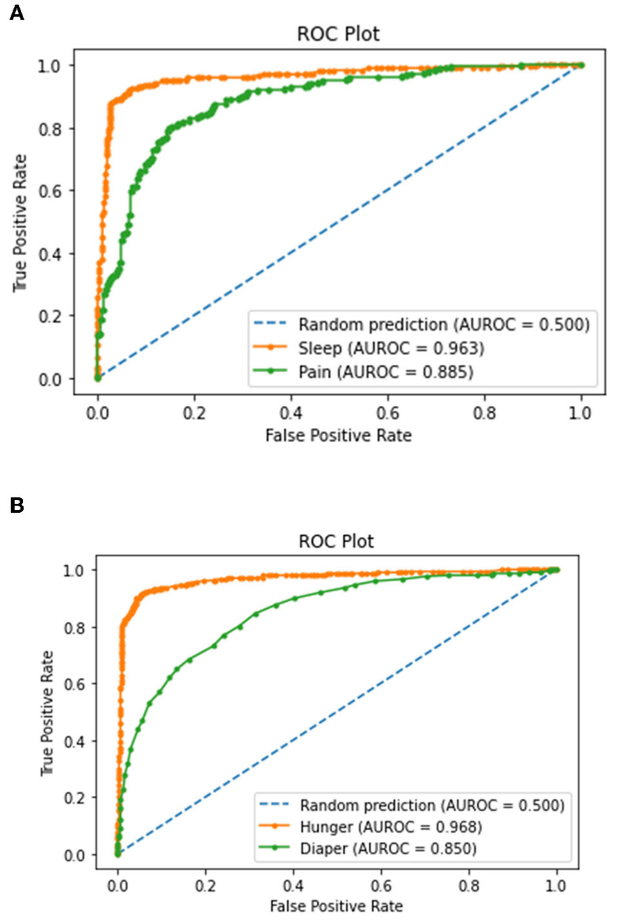


FIGURE 15 | Receiver operator characteristic (ROC) curve for **(A)** sleep and pain classes and **(B)** hunger and diaper classes.

The overall accuracy of the YOLOv4 model is 78.61%. This is better than the VGG16 model as this model is balancing its validation metrics for two classes (hunger and sleep), which accounts for 50% of the entire test data. From the cross-validation of YOLOv4, we can infer the performance of the model in predicting all four classes.

From **Table 14**, we can understand the evaluation metrics of the prediction classes. We can infer that sleep cries have the highest F1 score and the model works almost the same in predicting sleep and hunger cries. This is an improvement from the VGG16 model, as it was more inclined toward hunger cries. However, an imbalance still exists for the pain and diaper classes. The evaluation metrics need to be improved for these two classes. Affectability estimates how frequently a test accurately produces a positive outcome for individuals who have the condition that is being tried for (otherwise called the "genuine positive" rate). An exceptionally delicate test will hail nearly every individual who has the infection and will not create some bogus adverse outcomes. Particularity estimates a test's capacity to accurately produce an adverse outcome for individuals who do not have the condition that is being tried for (otherwise called the "genuine negative" rate). The F1 score is the weighted normal of precision and recall. Naturally, it is not as straightforward as precision; however, F1 is typically more valuable than exactness, particularly on the off chance that we have skewed class circulation.

Results for Multistage Heterogeneous Stacking Ensemble Classification Model 2

Figure 12A represents the epoch vs. loss graph and **Figure 12B** illustrates the epoch vs. accuracy graph of the proposed multistage heterogeneous stacking ensemble classification model 2 for a maximum of 175 epochs.

Exploring the Evaluation Metrics of the Proposed Multistage Heterogeneous Stacking Ensemble Classification Model 2

Figure 13 illustrates the confusion matrix of the multistage heterogeneous stacking ensemble classification model 2, and **Figure 14** portrays the probability distributions of the four classifiers along with the probability distribution of multi-stage heterogeneous stacking ensemble classification model 2.

From the cross-validation of the proposed model, we can infer that the model is performing at its best in predicting all of the classes. A perfect balance is maintained between all of the prediction labels. The overall accuracy of the model is 91.56%, which is far better than the hypertuned and optimized YOLOv4 model. The accuracies of all of the classes have also been increased compared to the previously implemented CNN variants.

Table 15 represents the evaluation metrics of the proposed multistage heterogeneous stacking ensemble classification model 2. By observing the computed values of the different metrics, we can infer that it has performed much better than the top CNN variants. The overall validation accuracy produced by the model is around 92%. With respect to the computational time, it was almost four times faster than CNN variants due to the advantage of highly advanced ensemble algorithms.

From **Table 15**, we can infer that all of the classes are classified with the best precision values after performing many hypertuning

TABLE 16 | Comparison with the state-of-the-art models.

Reference	Year	Method	Dataset	Number of emotion classes	Emotion classes	Mean F1 score
—	—	Multistage heterogeneous stacking ensemble model (proposed model)	Infant cry dataset—National Taiwan University Hospital Yunlin Branch, Taiwan	4	Diaper, sleep, hunger, pain	0.923
Jian et al. (33)	2021	LSTM + deepf_3	Fau Aibo children's emotion corpus database	5	Angry, hungry, pain, sad, tired	0.604
Ashwini et al. (4)	2021	Linear support vector machine	Infant cry dataset—National Taiwan University Hospital Yunlin Branch, Taiwan	3	Hunger, pain, sleepy	0.844
Jiang et al. (34)	2021	Gaussian mixture model-universal background model	Donate-A-Cry corpus	4	Hungry, discomfort, scared, tired	0.828
Boersma et al. (35)	2021	Convolutional neural net with self-attention	CRIED dataset	3	Fussing, screaming, high-pitched screeching	0.797

techniques. The hunger cries are performing the best compared to all other classes. The F1 scores of all of the classes have been improved compared to the previous two models. The proposed algorithm provides a balance in classifying the results of all of the prediction classes.

Plotting and Analyses of the Receiver Operator Characteristic Curve for the Four Prediction Classes

The area under ROC curve (AUC) is a valuable metric for assessing the nature of class intervals for delicate classifiers. In the multiclass setting, we can envision the exhibition of multiclass models as indicated by their one-vs.-all accuracy review curves. The AUC can likewise be summed up to the multiclass setting.

Figure 15A represents the receiver operator characteristic (ROC) curve and AUC scores for the sleep and pain prediction classes. From this, we can infer that the sleep class is getting predicted more accurately and perfectly compared to the pain class. This is because the higher the AUC score, the better the performance of the model in differentiating the true and false outcomes for the particular class. The AUC score for the sleep class is 0.963, and for the pain class is 0.885. From **Figure 15B**, we can infer that the ROC curve and AUC scores for the hunger class are much better than those for the diaper class. This is also because the model is more inclined toward the hunger class as compared to the diaper class.

DISCUSSION

The proposed hybrid model performed much better compared to the previously mentioned and implemented CNN variants in terms of accuracy and computational time. It achieved an overall accuracy of 92%. The stacking approach is highly suitable for the multiclass classification problem. Since the dataset consisted of audio signals and four different output prediction classes, consideration of four different classifiers simultaneously was highly beneficial. All of the classifiers were dependent on one another, unlike the top CNN variants. Due to this, the overall architecture of the model is more complex than the CNN variants. The mean AUC score of the proposed hybrid model is approximately 92%.

Comparison With State-of-the-Art Models

Table 16 presents the comparison of the proposed multistage heterogeneous stacking ensemble classification model 2 with the state-of-the-art models. The superior performance of the proposed model 2 in terms of the mean F1 score, could be witnessed from this table.

CONCLUSIONS

The multistage heterogeneous stacking ensemble model two consisting of four highly advanced boosting algorithms was roughly able to predict the classes on the basis of the feature vectors present in the spectrogram of the audio signals. However, these boosting algorithms are sensitive to some outliers. Our dataset comprises only four major reasons for the baby's cry. In the future, if some other major reason arises, then the model may not identify it and would produce incorrect results as it is a supervised learning model. The classifiers are trained on the feature vectors and these four classes. They will not be able to identify a new class, and this would require the model to undergo training once again to meet the updated changes. The proposed model can be extended to a semi-supervised or unsupervised approach by using some pretrained weights in a similar stacking-based model. This would save some time and efficiently provide better outcomes in terms of computational time and accuracy. In this research article, we were able to predict the reason for the infant's cry using two different approaches. In the first approach, we used pretrained weights and transfer learning approaches to predict the results. These included various CNN variants. A complete analysis and comparison were performed on the top CNN models. However, in the second approach, we made use of the multistage heterogeneous stacking ensemble classification model for enhancing the performance of the model using boosting algorithms. This approach produced much better results in terms of computational time, power, and accuracy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Research Ethics Committee at National Cheng Kung University (NCKU HREC) authorized by Ministry of Education, Taiwan. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

C-YC conceptualized and supervised the research. C-YC carried out funding acquisition. VJ investigated the data, performed the methodology, and implemented the software code. KS and C-YC carried out the project administration and validated the results. VJ and KS wrote the manuscript. VJ, KS, PV, and VR reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Ministry of Science and Technology, Taiwan (MOST 109-2221-E-224-045-MY3) and the Ministry of Education: Higher Education Sprout Project.

REFERENCES

1. Sallah A, Sundaravadivel P. Tot-mon: a real-time internet of things based affective framework for monitoring infants. *2020 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. Texas (2020) 93:600–601. doi: 10.1109/ISVLSI49217.2020.00093
2. Vankudre G, Ghulaxe V, Dhomane A, Badlani S, Rane T. A survey on infant emotion recognition through video clips. *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. Dubai (2021). p. 296–300. doi: 10.1109/ICCIKE51210.2021.9410786
3. Dewi SP, Prasasti AL, Irawan B. The study of baby crying analysis using MFCC and LFCC in different classification methods. *2019 IEEE International Conference on Signals and Systems (ICSigSys)*. Pune (2019). p. 18–23. doi: 10.1109/ICSIGSYS.2019.8811070
4. Ashwini K, Vincent PMDR, Srinivasan K, Chang CY. Deep learning assisted neonatal cry classification via support vector machine models. *Front Public Health.* (2021) 9:670352. doi: 10.3389/fpubh.2021.670352
5. Ji C, Mudiyanselage TB, Gao Y, Pan Y. A review of infant cry analysis and classification. *J Audio Speech Music Proc.* (2021) 2021:8. doi: 10.1186/s13636-021-00197-5
6. Liu L, Li W, Wu X, Zhou BX. Infant cry language analysis and recognition: an experimental approach. *IEEE/CAA J Autom Sin.* (2019) 6:778–88. doi: 10.1109/JAS.2019.1911435
7. Bgnicg I, Cucu H, Buzo A, Burileanu D, Burileanu C. Baby cry recognition in real-world conditions. *2016 39th International Conference on Telecommunications and Signal Processing (TSP).* Vienna (2016). p. 315–8. doi: 10.1109/TSP.2016.7760887
8. Asthana S, Varma N, Mittal VK. Preliminary analysis of causes of infant cry. *2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. Noida (2014). p. 000468–73. doi: 10.1109/ISSPIT.2014.7300634
9. Chunyan J, Chen M, Bin L, Pan Y. Infant cry classification with graph convolutional networks. *2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS).* (2021). p. 322–7. doi: 10.1109/ICCCS52626.2021.9449246
10. Xie J, Long X, Otte RA, Shan C. Convolutional neural networks for audio-based continuous infant cry monitoring at home. *IEEE Sens J.* doi: 10.1109/JSEN.2021.3123906
11. Lahmiri S, Tadj C, Gargour C. Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks. *IEEE Instrum Meas Mag.* (2021) 24:24–9. doi: 10.1109/MIM.2021.9400952
12. Lahmiri S, Tadj C, Gargour C, Bekiros S. Characterization of infant healthy and pathological cry signals in cepstrum domain based on approximate entropy and correlation dimension. *Chaos, Solitons Fractals.* (2021) 143:110639. doi: 10.1016/j.chaos.2020.110639
13. DŽeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one? *Mach Learn.* (2004) 54:255–73. doi: 10.1023/B:MACH.0000015881.36452.6e
14. Jain G, Prasad RR. Machine learning, Prophet and XGBoost algorithm: analysis of traffic forecasting in telecom networks with time series data. *8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. Bandung (2020). p. 893–7. doi: 10.1109/ICRITO48877.2020.9197864
15. XingFen W, Xiangbin Y, Yangchun M. Research on user consumption behavior prediction based on improved XGBoost algorithm. *2018 IEEE International Conference on Big Data (Big Data)*. New York (2018). p. 4169–75. doi: 10.1109/BigData.2018.8622235
16. Kabiraj S, Raihan M, Alvi N, Afrin M, Akter L, Sohagi SA. Breast cancer risk prediction using XGBoost and random forest algorithm. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. Kharagpur (2020). p. 1–4. doi: 10.1109/ICCCNT49239.2020.9225451
17. Jain VK, Kumar S, Fernandes SL. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J Comput Sci.* (2017) 21:316–26. doi: 10.1016/j.jocs.2017.01.010
18. Amin J, Sharif M, Yasmin M, Saba T, Anjum MA, Fernandes SL, et al. new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *J Med Syst.* (2019) 43:326. doi: 10.1007/s10916-019-1453-8

19. Amin J, Sharif M, Yasmin M, Fernandes SL, A. distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognit Lett.* (2020) 139:118–27. doi: 10.1016/j.patrec.2017.10.036
20. Shah JH, Sharif M, Yasmin M, Fernandes SL. Facial expressions classification and false label reduction using LDA and 3fold SVM. *Pattern Recognit Lett.* (2020) 139:166–73. doi: 10.1016/j.patrec.2017.06.021
21. Chang C-Y, Bhattacharya S, Durai Raj Vincent PM, Lakshmanna K, Srinivasan K. An efficient classification of neonates cry using extreme gradient boosting-assisted grouped-support-vector network. *J Healthcare Eng.* (2021) 2021:7517313. doi: 10.1155/2021/7517313
22. Chang CY, Chang CW, Kathiravan S, Lin C, Chen ST. DAG-SVM based infant cry classification system using sequential forward floating feature selection. *Multidim Syst Sign Process.* (2017) 28:961–76. doi: 10.1007/s11045-016-0404-5
23. Kaliappan J, Srinivasan K, Qaisar SM, Sundararajan K, Chang CY. Performance evaluation of regression models for the prediction of the COVID-19 reproduction rate. *Front Public Health.* (2021) 9:729795. doi: 10.3389/fpubh.2021.729795
24. Srinivasan K, Garg L, Chen B, Alaboudi AA, Jhanjhi NZ, Chang CT, et al. Expert system for stable power generation prediction in microbial fuel cell. *Intell Autom Soft Comput.* (2021) 30:17–30. doi: 10.32604/iasc.2021.018380
25. Srinivasan K, Garg L, Datta D, Alaboudi AA, Jhanjhi NZ, Agarwal R, et al. Performance comparison of deep CNN models for detecting driver's distraction. *Comput Mater Continua.* (2021) 68: 4109–24. doi: 10.32604/cmc.2021.016736
26. Ashwini K, Durai Raj Vincent PM, Srinivasan K, Chang CY. Deep convolutional neural network based feature extraction with optimized machine learning classifier in infant cry classification. *2020 International Conference on Decision Aid Sciences and Application (DASA)*. (2020). p. 27–32. doi: 10.1109/DASA51403.2020.9317240
27. Srinivasan K, Mahendran N, Vincent DR, Chang C-Y, Syed-Abdul S. Realizing an integrated multistage support vector machine model for augmented recognition of unipolar depression. *Electronics.* (2020) 9:647. doi: 10.3390/electronics9040647
28. Tuduce RI, Cucu H, Burileanu C. Why is my baby crying? An in-depth analysis of paralinguistic features and classical machine learning algorithms for baby cry classification. *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*. Vienna (2018). p. 1–4. doi: 10.1109/TSP.2018.8441363
29. Yuan Y, Luo X, Shang M, Wu D. A generalized and fast-converging non-negative latent factor model for predicting user preferences in recommender systems. *Proceedings of the Web Conference 2020 (WWW '20)*. New York, NY: Association for Computing Machinery (2020). p. 498–507. doi: 10.1145/3366423.3380133
30. Jindal S, Nathwani K, Abrol V. Classification of infant behavioural traits using acoustic cry: an empirical study. *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*. Illinois (2021). p. 97–102. doi: 10.1109/ISPA52656.2021.9552159
31. Onu CC, Lebensold J, Hamilton WL, Precup D. Neural transfer learning for cry-based diagnosis of perinatal asphyxia. *ArXiv, abs/1906.10199* (2019).
32. Alaie HF, Abou-Abbas L, Tadj C. Cry-based infant pathology classification using GMMs. *Speech Commun.* (2016) 77:28–52. doi: 10.1016/j.specom.2015.12.001
33. Jian T, Peng Y, Peng W, Yang Z. Research on LSTM + attention model of infant cry classification. *J Rob Networking Artif Life.* (2021) 8:218–23. doi: 10.2991/jrnal.k.210922.013
34. Jiang L, Yi Y, Chen D, Tan P, Liu X, A. novel infant cry recognition system using auditory model-based robust feature and GMM-UBM. *Concurr Comput Pract Exp.* (2021) 33:e5405. doi: 10.1002/cpe.5405
35. Li J, Hasegawa-Johnson M, McElwain NL. Analysis of acoustic and voice quality features for the classification of infant and mother vocalizations. *Speech Commun.* (2021) 133:41–61. doi: 10.1016/j.specom.2021.07.010
36. Boersma P. Praat, a system for doing phonetics by computer. *Glot Int.* (2001) 5:341–5.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fpubh.2022.838438

System Architecture of a European Platform for Health Policy Decision Making: MIDAS

Xi Shi[1,2*†], Gorana Nikolic[1†], Scott Fischaber[3], Michaela Black[4], Debbie Rankin[4], Gorka Epelde[5,6], Andoni Beristain[5,6], Roberto Alvarez[5,6], Monica Arrue[5,6], Joao Pita Costa[7,8], Marko Grobelnik[7,8], Luka Stopar[7,8], Juha Pajula[9], Adil Umer[9], Peter Poliwoda[10], Jonathan Wallace[11], Paul Carlin[12], Jarmo Pääkkönen[13] and Bart De Moor[1]

[1] Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium, [2] Vlerick Business School, Leuven, Belgium, [3] Analytics Engines, Belfast, United Kingdom, [4] School of Computing, Engineering and Intelligent Systems, Ulster University, Londonderry, United Kingdom, [5] Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, Spain, [6] EHealth Group, Biodonostia Health Research Institute, Donostia-San Sebastián, Spain, [7] Quintelligence, Ljubljana, Slovenia, [8] AI Lab, Institute Jozef Stefan, Ljubljana, Slovenia, [9] Data-Driven Solutions, Smart Health, VTT Technical Research Centre of Finland, Tampere, Finland, [10] IBM Ireland Lab, Innovation Exchange, International Business Machines Corporation, Dublin, Ireland, [11] School of Computing, Ulster University, Jordanstown, United Kingdom, [12] Faculty of Wellbeing, Education and Language Studies, Open University, Belfast, United Kingdom, [13] Centre for Health and Technology, University of Oulu, Oulu, Finland

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Victoria L. Lemieux,
University of British Columbia, Canada
Abhishek Gupta,
Shri Mata Vaishno Devi
University, India

***Correspondence:**
Xi Shi
xi.shi@vlerick.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: 17 December 2021
Accepted: 13 January 2022
Published: 31 March 2022

Citation:
Shi X, Nikolic G, Fischaber S, Black M, Rankin D, Epelde G, Beristain A, Alvarez R, Arrue M, Pita Costa J, Grobelnik M, Stopar L, Pajula J, Umer A, Poliwoda P, Wallace J, Carlin P, Pääkkönen J and De Moor B (2022) System Architecture of a European Platform for Health Policy Decision Making: MIDAS. Front. Public Health 10:838438. doi: 10.3389/fpubh.2022.838438

Background: Healthcare data is a rich yet underutilized resource due to its disconnected, heterogeneous nature. A means of connecting healthcare data and integrating it with additional open and social data in a secure way can support the monumental challenge policy-makers face in safely accessing all relevant data to assist in managing the health and wellbeing of all. The goal of this study was to develop a novel health data platform within the MIDAS (Meaningful Integration of Data Analytics and Services) project, that harnesses the potential of latent healthcare data in combination with open and social data to support evidence-based health policy decision-making in a privacy-preserving manner.

Methods: The MIDAS platform was developed in an iterative and collaborative way with close involvement of academia, industry, healthcare staff and policy-makers, to solve tasks including data storage, data harmonization, data analytics and visualizations, and open and social data analytics. The platform has been piloted and tested by health departments in four European countries, each focusing on different region-specific health challenges and related data sources.

Results: A novel health data platform solving the needs of Public Health decision-makers was successfully implemented within the four pilot regions connecting heterogeneous healthcare datasets and open datasets and turning large amounts of previously isolated data into actionable information allowing for evidence-based health policy-making and risk stratification through the application and visualization of advanced analytics.

Conclusions: The MIDAS platform delivers a secure, effective and integrated solution to deal with health data, providing support for health policy decision-making, planning of public health activities and the implementation of the Health in All Policies approach. The platform has proven transferable, sustainable and scalable across policies, data and regions.

Keywords: public health, decision support system, epidemiology, data visualization, machine learning

INTRODUCTION

We live in a data-rich society, which provides extensive opportunities for the development of big data and artificial intelligence technologies to provide new insights to enhance decision-making. Such technologies have particular importance in healthcare and health policy making. Despite the urgent need and opportunity, their use has not reached full potential in this field for various reasons, for example, healthcare data is typically heterogeneous and disconnected, existing in isolated silos, making meaningful analysis difficult. Privacy concerns create an additional barrier in exploiting the potential of healthcare data, preventing data sharing in a timely manner.

A systematic review on big data applications biomedical research and healthcare summarized the big data applications for clinical informatics and public health information (1). Among the studies on clinical informatics applications, most of the platforms were developed for data storage and retrieval (2, 3), data sharing (4, 5), and data security (6), which could not provide simulation, forecast or other analytics. Similarly, when the platform was developed for data analysis (7–9), data storage and data processing lost its priority. There are some platforms using social media to track and monitor public opinions, thereby providing evidence for policy decision making (10, 11). These platforms were mainly for infectious disease surveillance. In general, the platforms mentioned above have a focus on one aspect, such as data storage or processing, data analytics, or social media analysis. However, the whole process is all important to support the health policy decision making. An integrated platform including all these functions is in need.

A means of connecting healthcare data and integrating it with additional open and social data in a secure way did not exist prior to the MIDAS platform release. Such a solution can support the monumental challenge policy-makers face in safely accessing all relevant data to assist in managing the health and wellbeing of all.

The MIDAS project set out to address this challenge and has developed a novel health data platform that connects a range of heterogeneous health-related data with open and social data and applies advanced analytics techniques to provide a visual data-driven decision making tool that enhances healthcare policy making, whilst ensuring key aspects of ethics, security and privacy are adhered to (12). The platform has been piloted across four European regions: Basque Country (Spain), Finland, Northern Ireland (United Kingdom), and Ireland, addressing major health challenges in each region including mental health issues of young adults, diabetes and the aging population, childhood obesity, and social care for children, respectively.

This paper will present the system architecture of the MIDAS platform, which integrates data warehouse, data analytics, data visualization, and external applications for social media analysis, and enables rapid adjustments to new pilots.

METHODS

The MIDAS platform was developed in an iterative and collaborative way with close involvement of policy-makers and experts who informed data exploration and analysis based on their expertise. The co-created platform solves the practical policy questions proposed by the policy-makers and provides the possibility of being applied to a wider range of topics in a generally automated process. Moreover, the MIDAS platform addresses the problem of how the data can be linked, harmonized, analyzed, and visualized in a multinational framework. The scope of the specification encompasses user-interface integration, authentication and authorization, data storage, data preparation, analytics backend, visualization, and connection with external resources (**Figure 1**).

Pilots

MIDAS was developed in the light of the needs of four very different pilot sites with different research topics and data sources, namely the Basque Country (Spain), Finland, Northern Ireland (United Kingdom), and Ireland. The research objectives for each pilot are listed in the table (**Table 1**). The MIDAS project aimed to develop a platform that could deal with a wide range of topics in the international context using machine learning models. Therefore, each pilot had a unique research topic, and separate tailored dashboards were developed for them built upon a uniform architecture.

Platform Overview

As shown in **Figure 1**, the MIDAS Platform consists of a Policy Site Network and an External Network. The back-end analytics platform includes the Core Data Platform, tools for data harmonization (GYDRA), Analytics Backend, tools for data visualization (OpenVA), and three open and social data analytics and engagement tools in External Analytics Platforms. The policy-makers in a pilot site could adopt tools in the User Interface (UI), including the ISAACUS Metadata server, GYDRA, MIDAS Dashboard, News Media Dashboard, MEDLINE Dashboard, and Social Campaign Manager.

The MIDAS Platform is a collection of standard open-source big data processing tools, which is a modular, scalable data

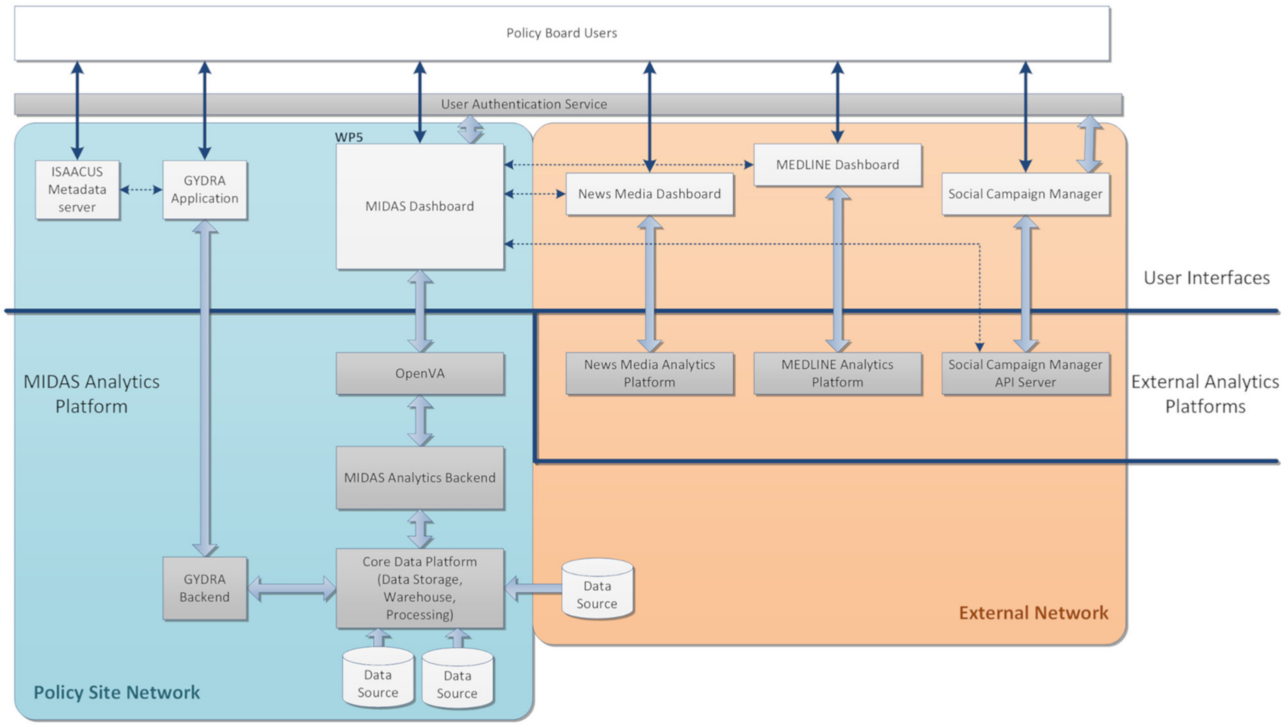


FIGURE 1 | MIDAS Platform Overview. The platform consists of software hosted within a Policy Site Network for analyzing local data and applications hosted externally (External Network) for analyzing open and social data as indicated by the colored boxes. The light-gray boxes indicate end user-facing web-applications, connected to back-end applications and analytics platforms shown in the dark-gray boxes. The white boxes are the data sources, either internally from different pilots or externally from open and social data. The arrows between different software indicate they have direct interactions.

TABLE 1 | Research topics for all pilots.

Pilot	Research objectives	Data source
Basque	To understand what drives childhood obesity and the etiology of the childhood obesity	Controlled and open data
Finland	To understand mental health issues of young people with the support of visual analytics and analysis of available diverse datasets	Controlled and open data
Northern Ireland	To analyze the anonymized data extracted from children care system to provide new insights into a child's journey through the care system	Controlled and open data
Republic of Ireland	To study the cohort of persons with diabetes and determine the best distribution for diabetes services	Controlled and open data

analytics platform along with the tools for packaging, deploying and configuring these applications in a bespoke manner.

The core services can be divided up into those which are necessary for the operation of the MIDAS Platform, and those which have been used for the development of the platform or which are optional depending on the desired usage. The required services are Hive, Spark, and HDFS; in addition, in the deployed MIDAS Platform, PostgreSQL is used for the Hive Metastore (database), but this can be changed to other database technologies. The rest of the services are deployed as part of the pilot site deployments, but these are optional services:

- OpenLDAP–used for service level authentication
- Hue–Web-UI for Hive/HDFS
- Jupyter–Web-UI for analytics notebook development
- Zeppelin–Web-UI for analytics notebook development
- PgAdmin–Web-UI for PostgreSQL
- PostgreSQL–used for Hive Metastore and/or Unified Data View data virtualization

An overview of the Core Data Platform configured for MIDAS is given in **Figure 2**, including the core services of data storage and processing, user applications for interacting with these services, local user authorization and authentication, and data virtualization.

Data Storage and Processing

The underlying data storage for the MIDAS Platform is HDFS-based. Where data virtualization is desired this can also be provided through Hive or via PostgreSQL.

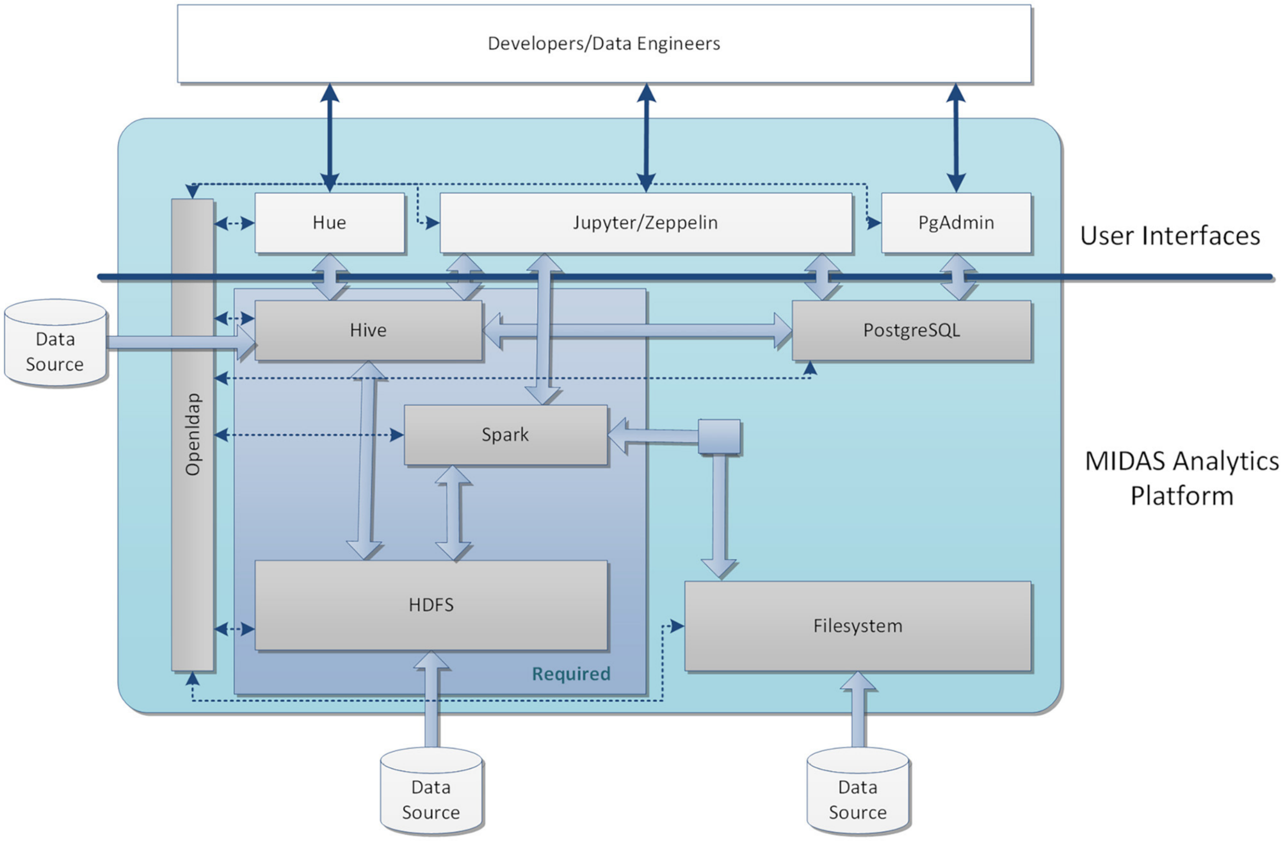


FIGURE 2 | An overview of the system configured for MIDAS. The core data platform for the MIDAS stack was based on HDFS, Hive, and Spark. The data can be imported into the system through Filesystem, HDFS, or externally to Hive. HDFS was applied to store files and raw data and Hive was employed as a data warehouse for the structured data after processing. External data assets were also virtualized through the Hive interface and they could be accessed by the MIDAS tools similarly to locally loaded data assets and used within GYDRA. The UI of the analytic platform includes Jupyter Notebook with Python and PySpark for developing and testing the underlying analytics models before being implemented within the MIDAS Analytics Backend. For managing and querying the databases in Hive and PostgreSql, an open sourced interactive editor Hue was used, and Zeppelin provided support for running Spark applications. User access was managed by a local LDAP server, which provided role-based access to the user applications and underlying data stores. © [2021] IEEE. Reprinted, with permission, from (31).

Data processing engines include MapReduce, Spark and Celery for running distributed analytics workloads on the data, with Hive being employed as a data warehouse for the data within HDFS to structure it so that it can be analyzed and results provided to the MIDAS Dashboard.

User Applications

For development of the platform, a number of web-based applications are provided to technical users to access various services within the MIDAS Platform, including Jupyter Notebook web-application which provides entry-points to access data within HDFS/Hive and is used to develop the underlying analytics models and code before being implemented within the MIDAS Analytics Platform; Hue for working with Hive and viewing the underlying HDFS file structure; PgAdmin for interacting with PostgreSQL; and Zeppelin notebook web-application for running code on various services.

User Authorization and Authentication

Access to the underlying data stores and services within the MIDAS Platform is managed by a local LDAP server (running OpenLDAP), although this could be replaced with a user-specific local server or a centralized server (e.g., Active Directory) within a pilot site. This provides role-based access to the user applications as well as HDFS and Hive. Access to data within HDFS can be limited to a specific user-group or MIDAS applications, for instance, restricting access to the raw data to a pre-processing group of users or the GYDRA application.

Data Virtualization

Data virtualization to external data sources outside of the MIDAS platform uses Hive. This provides access to external data assets that may be held outside the MIDAS platform. External data assets will likely be existing databases (PostgreSQL, SQL Server, Oracle, etc.) which have already been preprocessed (e.g., to create a register). Once access to these external assets has been set up in Hive, they can be accessed by the MIDAS tools similarly to locally loaded data assets and used within GYDRA or pulled through to the MIDAS Dashboard.

Data Preparation and Harmonization

The data preparation and harmonization task aimed to develop appropriate pre-processing modules for preparing the raw data to ensure that they were compatible with the agreed data representations and could be used for analysis, including for instance: data cleansing, normalization, transformation, joining, and missing value imputation. The GYDRA software (renamed from TAQIH) was developed and applied for data preprocessing and transformation (13, 14). The GYDRA is a customizable tool for facilitating the data wrangling process through interactive and visual tools, taking advantage of machine learning algorithms. The aim is to simplify the tedious and time consuming part of data analysis, allowing non-technical users to transform raw data into information ready for analysis.

The GYDRA provides web interfaces to understand the content, structure and distribution of the dataset through an easy-to-use tab-based navigation approach following common data assessment and preparation steps. **Figures 3A,B** presents screenshots for two representative sections for general statistics and missing values, respectively. Moreover, on each tab or section of the application, a visual transformation pipeline allows the users to add a dataset transformation action after knowing the dataset's content.

As a python-centered solution, with an easy-to-use interactive UI, the GYDRA uses Celery for asynchronous distributed data-processing suitable for handling big HDFS datasets that do not fit into system memory. Additionally, through the web-based GYDRA tool, a data synchronization function allows the data owners and policy-makers to efficiently deploy prepared datasets to the analytics platform. The synchronization logic aligns the GYDRA metadata tool with the ISAACUS metadata server and updates the data warehouses further through the GYDRA backend depicted in **Figure 1**. The raw data of each health policy area was prepared and processed using the GYDRA tool thus making the data ready for the MIDAS Analytics Backend.

The details of data sources and data types are listed in the **Supplementary Material**. The technical details of the data processing section have been published (13), and another published use case can be used as an example to show how the data was processed and prepared for data analytics (15).

Data Analytics

The MIDAS Analytics Backend provides the back-end analytics and simulation results required for the MIDAS dashboard. Apart from being a middle layer linking the data preparation and the data visualization, it supervises the user in selecting the correct data tables and data variables for chosen analytics and visualization scenarios.

The communication between the analytics and visualization layers was managed through a REST API server developed with the Flask microframework for Python. The Analytics APIs were developed to support generic exploratory data analysis (EDA), uniform across all pilot sites, as well as more specialized cross-filter dashboards and health policy simulators specific to each pilot-site.

The EDA was uniform for all pilots, providing eight types of basic visualizations for the selected variables from the harmonized data, i.e., scatter plot, heatmap, histogram, bar chart, pie chart, bubble plot and choropleth map. The cross-filter analytics for each pilot platform (**Figure 4**), which are interactive visualization tools (16), had the same basic principle to update their content when the user selects different values on the displayed graphs. The associated visualizations were flexible for different pilots, with the layout and categorical variables proposed by policy-makers, including components such as line chart, bar chart, and tables.

Different machine learning methods were applied for each pilot to solve their unique research questions (**Table 2**). Because of data protection regulations, the data-related results cannot be shown. As the main focus of the paper is on system architecture, the detailed results are not discussed and shown in this paper.

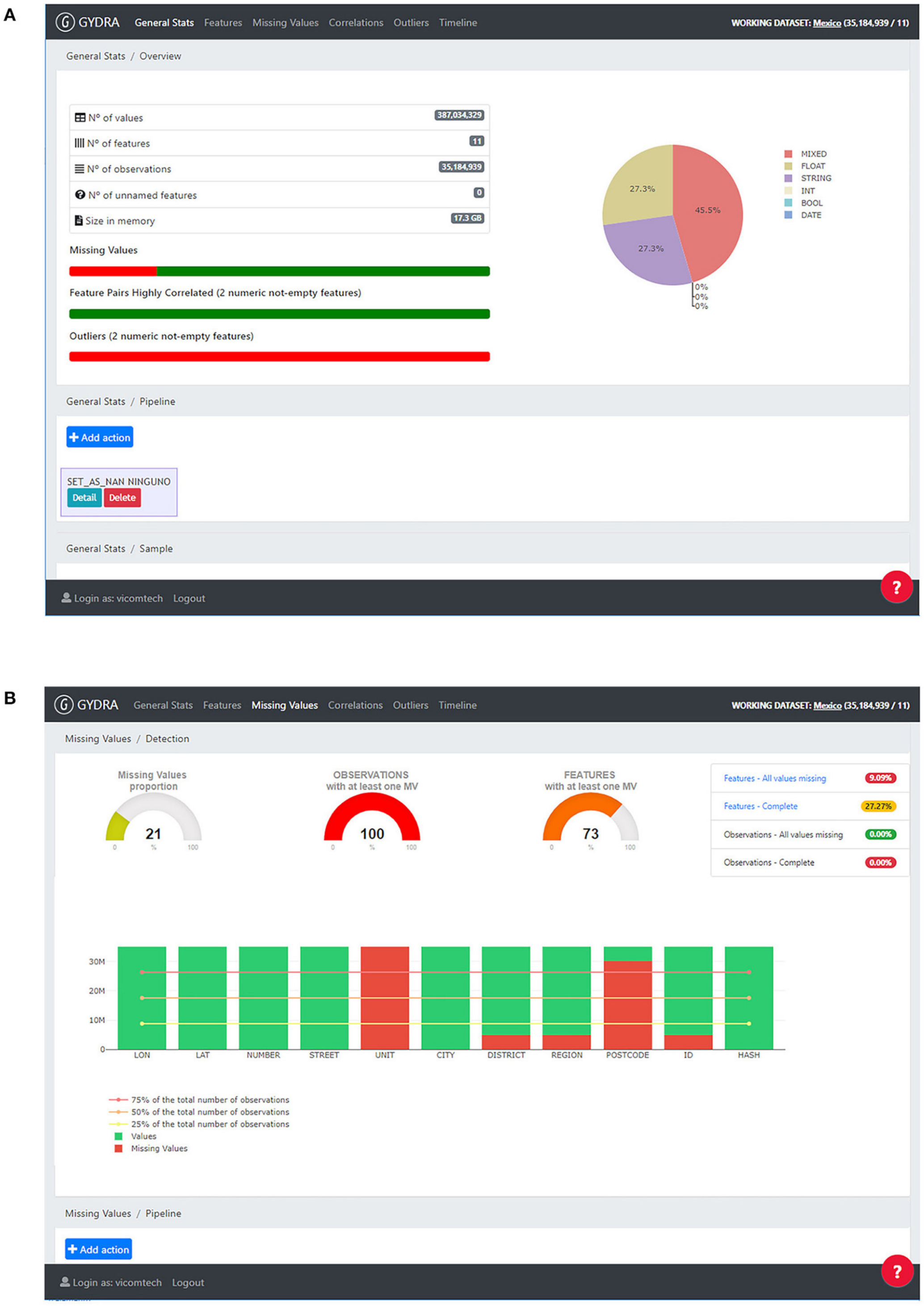
The private MIDAS GitHub repository contains branches of each pilot, consisting of API endpoints for generic EDA, cross-filter, and pilot-specific analytics. Different types of cross-filter and pilot-specific analytics were deployed on each of the pilots in an iterative process. Feedback from policy-makers on the required analytics with evaluation of results was collected in each deployment iteration, making it possible to meet the real needs of the policy-makers.

Data Visualization

Data visualization was Deployed utilizing a three-tier architecture in the MIDAS Dashboard, including the MS Azure AD B2C authentication service (17), the OpenVA middleware framework (18), and the dedicated MIDAS UI single page application (SPA), which provides decision-making support for policy-makers with data-driven analytics from the internal and external resources.

A Single-Sign-On service was implemented between the MIDAS Dashboard and external resources through the common authentication service, mentioned above. The OpenVA framework handles the connectivity of shared SPA to local resources and dedicated external components. The MIDAS UI SPA is shared by all instances from a centralized web server and it connects to the local OpenVA instance in line with the account details of the current user. The external resources include the Social Campaign Manager, MEDLINE Publication search and News Media search dashboard.

Through the MIDAS UI, users can generate a dashboard and interact with a widget wizard to generate the specific visualization widgets that can help them with policy decision support. Furthermore, additional pilot-specific dashboards and analytics tools were developed for each pilot, supporting each user in exploring and understanding their main research question. The MIDAS UI (**Figure 5A**) shows the visualized analytics results for selected datasets, together with the reporting tool illustrated in **Figure 5B** to allow users to generate a PDF report.



WORKING DATASET: Mexico (35,184,939 / 11)

N° of values	387,034,329
N° of features	11
N° of observations	35,184,939
N° of unnamed features	0
Size in memory	17.3 GB

Features - All values missing	9.09%
Features - Complete	27.27%
Observations - All values missing	0.00%
Observations - Complete	0.00%

FIGURE 3 | GYDRA data preparation tool UI, provided through an easy-to-use tab-based navigation approach following a common data assessment and preparation steps (i.e., General Stats, Features, Missing Values, Correlation and Outliers analysis tasks). Screenshots for two representative sections are included: **(A)** General stats–On the left side, general statistics on features and observations are provided, on the right side the variables type distribution is shown on a pie chart. On the lower part a transformations pipeline is included to add dataset transformations as their need is identified. **(B)** Missing values–On the top left area, missing value proportion is depicted for values, features and observations, on the right side indicators for complete and completely empty features and observations is provided. On the lower part, each feature is analyzed separately on bar charts representing their missing value percentage. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Business Information System Workshops, Chapter Enhancing the Interactive Visualization of a Data Preparation Tool from in-Memory Fitting to Big Data Sets by (14).

Age Groups	2000-2002	2003-2005	2006-2008	2009-2011	2012-2014	2015-2016
37-39	NaN	NaN	NaN	NaN	0.04	0.55
34-36	NaN	NaN	NaN	0.03	0.1	0.37
31-33	NaN	NaN	0.02	0.06	0.09	0.38
28-30	NaN	0	0.05	0.06	0.1	0.39
25-27	0.02	0.02	0.03	0.04	0.1	0.41
22-24	0.03	0.02	0.04	0.04	0.08	0.41
19-21	0.04	0.04	0.03	0.06	0.08	0
16-18	0.04	0.03	0.03	0.05	0	NaN
13-15	0.04	0.02	0.02	0	NaN	NaN
10-12	0.03	0.02	0	NaN	NaN	NaN
7-9	0.03	0	NaN	NaN	NaN	NaN

mouthpharynx

● Yes ○ No

Gender

Male Female [✓ Both]

Region

Helsinki-Uusimaa

North & East Finland

South Finland

West Finland

FIGURE 4 | Cross-filter of the Finnish Pilot. Users can select the county or region in the map on the right and all the charts will update automatically. Gender, Region, and Age Group are the categorical variables shown in the lower panel next to the map, and users could select subgroups either by pressing the buttons on the panel or selecting the subgroups in the line chart or bar chart. The Finnish cross-filter consists of a matrix heatmap, a bar chart, a line chart and the regional map of Finland. The cross-filter of other pilots have different components.

TABLE 2 | Pilot-specific analytics.

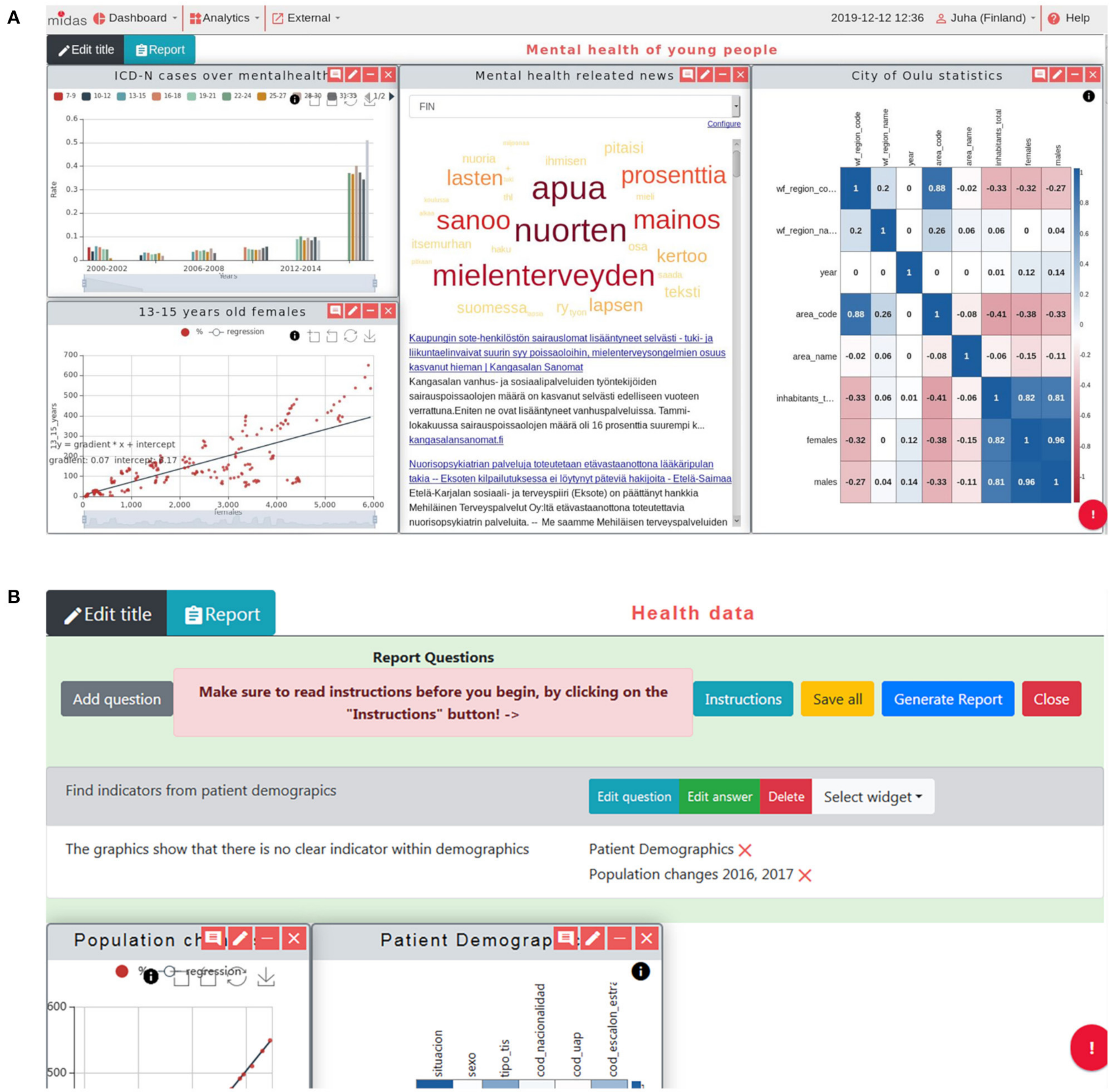
Pilot	Method	Purposes
Basque	RandomForests/LASSO	To identify the risk factors of childhood obesity
Finland	Lexis diagram analysis	To aggregate, summarize and visualize the selected risk factors in a secure way to protect the privacy of patients
Finland	Descriptive analysis	To intuitively evaluate the health, social, and education status of the inhabitants in regional level on a yearly basis by using open data
Northern Ireland	Markov chain	To track patterns of behavior over time and to give better visualization to intuitively present how children move in and out of different types of care by estimating the probability of the transition between different types of care
Northern Ireland	LSTM Network	To predict the future status of children to improve the protection for children from the policy level
Republic of Ireland	ARIMA	To forecast the consumption of diabetic drugs

Open and Social Data

Social Media Analysis

Social media is considered as an important source of information for policy making, to help better understand motivations and determine public acceptance of implemented policies. The social media analysis provides insights into the public's perception and sentiments toward health policies by having members of the public engage with the created chatbots through a series of questions about a specific health policy. The questions included a number of multiple choice questions or open questions to be answered in free-form text. The free-form text responses are analyzed in real time as they are entered into the system using IBM Watson Natural Language Understanding (NLU) APIs (19), and the sentiment and emotional analysis are then displayed on the Social Campaign Manager dashboard or widget on the MIDAS UI dashboard (**Figure 6A**). In order to avoid bias and protect the participants' privacy, the analysis was not done on the individual level, but on the aggregated level. The bot extracted emotions from the free-form text, only giving the potential inclination of the participants. The aggregated view of these responses is the percentage of one type of emotions or opinions, which alleviated the bias generated from individual response.

The Social Campaign Manager was hosted as a microservice on the IBM Cloud platform. Twitter was used to interact with the public and the IBM Watson Assistant and Watson NLU services were used for the chatbot. The Social Campaign Manager was



[Kaupungin sote-henkilöstön sairauslomat lisääntyneet selvästi - tuki- ja liikuntaelinvaivat suurin syy poissaoloihin, mielenterveysongelmien osuus kasvanut hieman | Kangasalan Sanomat]
Kangasalan vanhus- ja sosiaalipalveluiden työntekijöiden sairauspoissaolojen määrä on kasvanut selvästi edelliseen vuoteen verrattuna.Eniten ne ovat lisääntyneet vanhuspalveluissa. Tammi-lokakuussa sairauspoissaolojen määrä oli 16 prosenttia suurempi k...
kangasalansanomat.fi

[Nuorisopsykiatrian palveluja toteutetaan etävastaanottona lääkäripulan takia -- Eksoten kilpailutuksessa ei löytynyt päteviä hakijoita - Etelä-Saimaa]
Etelä-Karjalan sosiaali- ja terveyspiiri (Eksote) on päättänyt hankkia Mehiläinen Terveyspalvelut Oy:ltä etävastaanottona toteutettavia nuorisopsykiatrin palveluita. -- Me saamme Mehiläisen terveyspalveluiden

	wf_region_code	wf_region_name	year	area_code	area_name	inhabitants_total	females	males
wf_region_co...	1	0.2	0	0.88	-0.02	-0.33	-0.32	-0.27
wf_region_na...	0.2	1	0	0.26	0.06	0.06	0	0.04
year	0	0	1	0	0	0.01	0.12	0.14
area_code	0.88	0.26	0	1	-0.08	-0.41	-0.38	-0.33
area_name	-0.02	0.06	0	-0.08	1	-0.06	-0.15	-0.11
inhabitants_t...	-0.33	0.06	0.01	-0.41	-0.06	1	0.82	0.81
females	-0.32	0	0.12	-0.38	-0.15	0.82	1	0.96
males	-0.27	0.04	0.14	-0.33	-0.11	0.81	0.96	1

Make sure to read instructions before you begin, by clicking on the "Instructions" button! ->

Find indicators from patient demograpics

The graphics show that there is no clear indicator within demographics

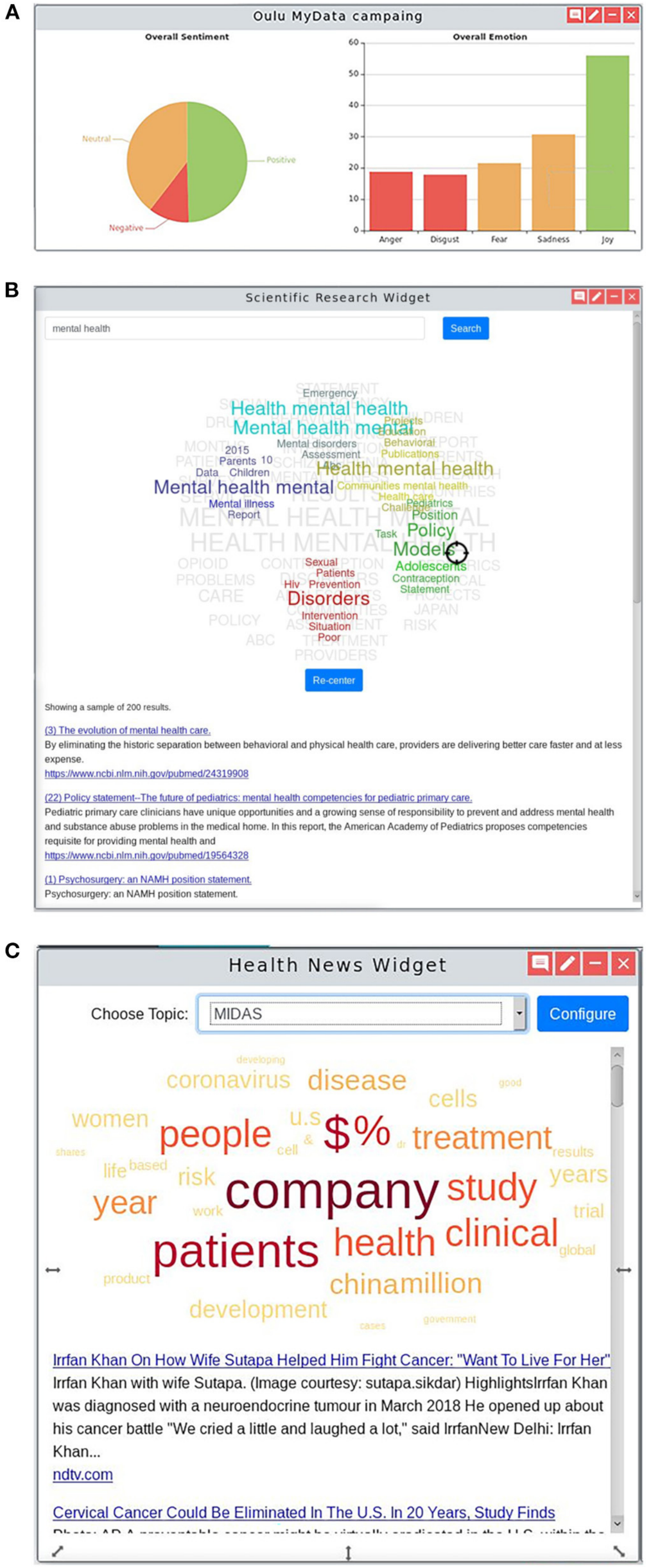
Patient Demographics ✕
Population changes 2016, 2017 ✕

FIGURE 5 | MIDAS UI screenshots. **(A)** Common view of generated dashboard. On top of the view are the menus to manage dashboards, add analytics and use the external resources. The reporting tool is located in the middle (hidden in figure) and the rest of the view is the open space for widgets. Users can freely resize and organize widgets in this space. The analytic results come from the MIDAS Analytics Backend, together with the widgets developed externally. **(B)** Reporting tool open with a single research question and an answer with two associated widgets. This tool is used to generate a PDF file, defining the research questions and answers and attaching the most suitable figures describing them from the available widgets. © [2021] IEEE. Reprinted, with permission, from (31).

a separate web application for creating, running, and managing the individual campaigns. These provided the intelligence and dialog capability to interact with the user as well as performing the analysis of the conversations. The Social Campaign Manager API Server was the core application connecting these various services, providing data to the MIDAS Dashboard and the Social Campaign Manager web application. The authentication used within MIDAS platform layers is OAuth 2.0, a common industry-standard protocol for authorization.

MEDLINE Analytics

The MEDLINE dashboard accessible through the MIDAS platform was developed to provide dedicated text-mining tools and visualizations to enable users to extract meaningful



Showing a sample of 200 results.

(3) The evolution of mental health care.

By eliminating the historic separation between behavioral and physical health care, providers are delivering better care faster and at less expense.

https://www.ncbi.nlm.nih.gov/pubmed/24319908

(22) Policy statement--The future of pediatrics: mental health competencies for pediatric primary care.

Pediatric primary care clinicians have unique opportunities and a growing sense of responsibility to prevent and address mental health and substance abuse problems in the medical home. In this report, the American Academy of Pediatrics proposes competencies requisite for providing mental health and

https://www.ncbi.nlm.nih.gov/pubmed/19564328

(1) Psychosurgery: an NAMH position statement.

Psychosurgery: an NAMH position statement.

Irrfan Khan On How Wife Sutapa Helped Him Fight Cancer: "Want To Live For Her"

Irrfan Khan with wife Sutapa. (Image courtesy: sutapa.sikdar) HighlightsIrrfan Khan was diagnosed with a neuroendocrine tumour in March 2018 He opened up about his cancer battle "We cried a little and laughed a lot," said IrrfanNew Delhi: Irrfan Khan...

ndtv.com

Cervical Cancer Could Be Eliminated In The U.S. In 20 Years, Study Finds

FIGURE 6 | Open and Social Analytics. **(A)** Social Campaign Manager shows a high-level overview of the (i) sentiment and (ii) emotional analysis of the policy being studied by a Twitter social media campaign. The sentiment and emotions found in responses to a public online survey reaching out to the public to gather their voice on

(Continued)

FIGURE 6 | a specific health policy being considered on the dashboard. Clicking into the dashboard provides further insight including responses to particular questions and results processed using Natural Language Processing techniques showing the most common topics of conversation mentioned in the responses and the sentiment in which they were made. **(B)** MEDLINE custom widget that includes: (i) a list of the top ten MEDLINE articles with the first part of the abstract serving as a short description; (ii) a tag-cloud representing clusters of topics extracted from the MEDLINE articles including the searched keywords; and (iii) a target-shaped pointer that the user can move through the tag-cloud and by that, change the ranking of the listed articles. **(C)** News custom widget that consists of: (i) a word cloud that represents the main topics of the listed news, enabling a global perspective of the key topics before further activity; (ii) a list of news titles and first lines that are linked to the original news source; and (iii) the search choices where news are based on, defined by the filter and search options at the "Media Monitoring" menu of the external news dashboard.

information from the MEDLINE dataset (20, 21). The MEDLINE dataset was indexed using the ElasticSearch, and visualized through an open source tool Kibana (22). The purpose of the dashboard was to provide users with tools to explore the insights of published biomedical research, in an intuitive manner. The main advantage is the dynamic article prioritization (ranking). The user enters a few keywords in the search box and results are shown (**Figure 6B**). This visual interactive widget helps surface information that one is looking for by re-ranking the top 10 articles, letting the users interact with the index of the results, getting them closer to the scientific information that they are looking for.

Each topic dashboard was developed through extensive interactions with the pilot sites, improving the understanding of how the tools could be used to address specific use cases. The MEDLINE knowledge was also served directly at the MIDAS platform by a widget that also allowed for Lucene queries and for the user to interact with a pointer over a tag cloud of related topics to alter the order of scientific articles provided as result of the query.

News Media Analysis

MIDAS provides users with tools to monitor specific health topics in the worldwide and local news. The news media analysis tool is available through the platform (**Figure 6C**), enabling the monitoring of worldwide news outlets and the enriching of these news articles with data from the MEDLINE knowledge base (23). Each pilot region in the MIDAS project has its own live news source which can be accessed via the dedicated news data exploration dashboard served by the Event Registry news engine and through a widget within the MIDAS Dashboard UI (24). In addition to setting up the pilot-specific data streams, the underlying data sources for Event Registry were improved to better support Finnish and Basque language news coverage, adding to the 60+ languages available. In addition to the news media tools, a MeSH Classifier tool was developed which enables classification of news articles (and any text snippets) with MeSH terms. The system is available through a web portal and a REST API, and includes a NodeJS wrapper for direct inclusion into other systems (25).

Implementation

Given the heterogeneous nature of the various data sources, policy environments and stakeholder perspectives, the platform development followed an agile, user- centerd design approach to ensure that user needs were met across the consortium and beyond. User-centered design approach included a co-design workshop, an iterative platform evaluation, and feedback integration. The co-design workshop was attended by approximately 80 participants, including a mixture of consortium members and external stakeholders. The professional backgrounds of attendees were diverse and included policy-makers, civil servants, academic experts, and industry representatives. The workshop took participants through a staged process, which included the development of "personas" (i.e., typical users of the system), the identification of "user stories" (simple, non-technical descriptions of user requirements), and the brainstorming of "wireframes" (interface design ideas) on paper and online. The results from the workshop were subsequently collated, analyzed and distributed among consortium partners to inform the future development of the MIDAS platform (26).

We conducted three rounds of user experience testing to help improve the intermediate prototypes, methodology and results of the initial round are reported in (27). A combination of heuristic and formative user-centered evaluation methods was employed, providing feedback from both usability experts and evaluating prototypes with real users. A rigorous test protocol was jointly developed by consortium members, led by usability testers from Ulster University's UX Lab. The usability testing protocol was informed by Ulster's UX-Lab having carried out a range of usability tests on medical devices, software and data visualizations (28, 29). The participants included data analysts and policy makers, a more detailed demographic statistics can be seen in **Supplementary Table 2.1** in the **Supplementary Material**. We guided the participants to finish a list of tasks and collected their feedback and suggestions for further improvements. The user experience testing helped successfully identify the potential problems, and improvements were achieved after incorporating user feedback.

RESULTS

The developed MIDAS platform consists of several dashboards, including Exploratory Data Analysis (EDA) (**Figure 5A**), cross-filter dashboard (**Figure 4**), pilot-specific analytics dashboard, and social media dashboards (**Figure 6**).

The final versions of the pilot platforms were evaluated by policy makers from all pilots based on the Key Performance Indicators (KPI) (**Supplementary Table 2.2** in the **Supplementary Material**). The second column is the demands proposed by policy makers, and the third column is the corresponding function on the MIDAS platform. All KPIs were successfully achieved and the platform has received positive feedback from stakeholders on its capacity to integrate and analyze previously fractured heterogeneous data. Furthermore, the ability to produce new knowledge and results that are actionable by health policy-makers was demonstrated. The custom-tailored analytics solved the practical questions for the health policy-makers and gave them insights for possible future interventions. The platform can be easily manipulated by users without technical background by following the User Guide (30).

DISCUSSION

Principal Results

The core user groups of these tools are mainly business users, dashboard users and in-house analytics teams. In contrast, the MIDAS platform was co-created by academia, industry, and crucially, healthcare staff, health policy-makers, patients and citizens thus ensuring the solution's design and development has been user-led. With this user-centered approach, the MIDAS platform guides its users through all steps of the data analytics pipelines. Besides, data blending is restricted according to the prior knowledge of the original data in the data processing procedure. These restrictions assist the user in selecting only suitable variables for a chosen visualization, thus producing meaningful analytics and visualization results.

Because of the flexibility of the open data tools, they can be quickly adjusted to study the most urgent topics, as a result, MIDAS recently presented a fast response to the COVID-19 global initiative (31). This impactful public health event was addressed through the worldwide news, offering the customized news streams through the MIDAS news widget, to help the pilot site use cases to better track news and relate it to their own priorities.

In order to maximize the sustainability of the MIDAS platform beyond the lifetime of the project, we explored a range of mechanisms for coordinating further development and marketing activities among the project contributors post-project. After detailed partner and stakeholder engagement we determined that the establishment of a MIDAS Open Source Foundation would be the most suitable approach. New regions, cities, and organizations from Scotland, France, and Spain have confirmed their interest, with more public sector policy departments noticing the platform capability of addressing similar problems in their area in future.

Comparison With Prior Work

The MIDAS platform tries to maintain the privacy of each stakeholder by keeping their sensitive health data in-house. Other commercial tools like Tableau, PowerBI, or QlikView often require a connection to external services, while all layers of the MIDAS platform are hosted inside the stakeholder's trusted zone. Moreover, they are general purpose solutions that do not consider the specific challenges of public health data, nor the user stories of the target MIDAS audiences. Therefore, each layer of the MIDAS platform supports a secure data analytics pipeline and minimizes data-leakage. Additionally, the learning curve of some commercial tools can be steep, requiring specialized training. In terms of advanced analytics capabilities, Tableau provides some advanced analytics features but with external integration, PowerBI has core competency and integration, while QlikView does not offer any advanced analytics features.

Conclusions

This study has demonstrated the value of a secure, effective and integrated solution that deals with health data to harness the potential of underutilized healthcare data and provide support for health policy decision making. The MIDAS platform was successfully implemented within the four pilot regions and has received positive feedback from stakeholders on its capacity to turn large amounts of previously isolated data into actionable information to inform health policy making and risk stratification through the applications and visualizations of advanced analytics. By delivering the MIDAS platform as an innovative and state-of-the-art solution, we have successfully provided a tool with fully functioning architecture that can potentially transform the way health policies are developed, evaluated and implemented, which will ultimately enable impactful improvements in public health and the quality of life amongst European citizens and beyond. Besides, the platform has successfully demonstrated that it is transferable, sustainable and scalable across policies, data and regions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from MIDAS but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

AUTHOR CONTRIBUTIONS

XS, GN, SF, MB, DR, GE, JPC, JuP, PP, and JW: analyzed the data, contributed reagents, materials, analysis tools, and wrote the paper. All authors were involved to conceive and design the study, reviewed and interpreted the results, commented on manuscript, contributed to revision, and read and approved the final version.

FUNDING

This work was supported by KU Leuven: Research Fund (projects C16/15/059, C3/19/053, C32/16/013, and C24/18/022), Industrial Research Fund (Fellowship 13-0260), and several Leuven Research and Development bilateral industrial projects, Flemish Government Agencies: FWO [EOS Project no 30468160 (SeLMA), SBO project S005319N, Infrastructure project I013218N, TBM Project T001919N; Ph.D. Grants (SB/1SA1319N, SB/1S93918, SB/151622)]. This research received funding from the Flemish Government (AI Research Program). BD and XS are affiliated to Leuven.AI–KU Leuven institute for AI, B-3000, Leuven, Belgium. VLAIO [City of Things (COT.2018.018), Ph.D. grants: Baekeland (HBC.20192204) and Innovation mandate (HBC.2019.2209), Industrial Projects (HBC.2018.0405)], European Commission: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No 885682), (EU H2020-SC1-2016-2017 Grant Agreement No.727721: MIDAS), and KOTK foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.838438/full#supplementary-material

REFERENCES

1. Luo J, Wu M, Gopukumar D, Zhao, Y. Big data application in biomedical research and health care: a literature review. *Biomed Inform Insights.* (2016) 8:1–10. doi: 10.4137/BII.S31559
2. Jin Y, Deyu T, Yi Z. A distributed storage model for EHR based on HBase. In: *International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII).* Shenzhen, China; IEEE. (2011) p. 369–72. doi: 10.1109/ICIII.2011.234
3. Sahoo SS, Jayapandian C, Garg G, Kaffashi F, Chung S, Bozonrgi, et al. A Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data' using cloud computing for epilepsy clinical research. *J Am Med Inform Assoc.* (2014) 21:263–71. doi: 10.1136/amiajnl-2013-002156
4. Bahga A, Madisetti VK, A. cloud-based approach for interoperable electronic health records (EHRs). *IEEE J Biomed.* (2013) 17:894–906. doi: 10.1109/JBHI.2013.2257818
5. Sharp J. An application architecture to facilitate multi-site clinical trial collaboration in the cloud. In: *Proceedings of the 2nd International Workshop on Software Engineering for Cloud Computing.* Honolulu, Hawaii: ACM. (2011) p. 64–8. doi: 10.1145/1985500.1985511
6. Schultz T. Turning healthcare challenges into big data opportunities: a use-case review across the pharmaceutical development lifecycle. *J Assoc Inf Sci Technol.* (2013) 39:34–40. doi: 10.1002/bult.2013.1720390508
7. Ng K, Ghoting A, Steinhubl SR, Stewart WF, Makin B, Sun J. PARAMO: a PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records. *J Biomed Inform.* (2014) 48:160–70. doi: 10.1016/j.jbi.2013.12.012
8. Zolfaghar K, Meadem N, Teredesai A, Basu Roy S, Chin SC, Muckian B. Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In: *IEEE International Conference on Big Data.* Santa Clara, California: IEEE. (2013) p. 64–71. doi: 10.1109/BigData.2013.6691760
9. Deligiannis P, Loidl H-W, Kouidi E. Improving the diagnosis of mild hypertrophic cardiomyopathy with MapReduce. In: *Proceedings of Third International Workshop on MapReduce and its Applications Date.* Delft, Netherlands: ACM. (2012) p. 41–8. doi: 10.1145/2287016.2287025
10. Hay SI, George DB, Moyes CL, Brownstein, JS. Big data opportunities for global infectious disease surveillance. *PLoS Med.* (2013) 10:e1001413. doi: 10.1371/journal.pmed.1001413
11. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Prevent Med.* (2014) 63:112–5. doi: 10.1016/j.ypmed.2014.01.024
12. Black M, Rankin D, Wallace J, Carlin P, Bond R, Mulvenna M, et al. Meaningful integration of data, analytics and services of computer-based medical systems: The MIDAS touch. *Proceedings of the IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS); 2019 Jun 5-7; Cordoba, Spain.* New York: IEEE. (2019).
13. Álvarez Sanchez R, Beristain Iraola A, Epelde Unanue G, Carlin P. TAQIH, a tool for tabular data quality assessment and improvement in the context of health data. *Comput Methods Programs Biomed.* (2019) 181:104824. doi: 10.1016/j.cmpb.2018.12.029
14. Epelde G, Álvarez R, Beristain A, Arrue M, Arangoa I, Rankin D. Enhancing the interactive visualisation of a data preparation tool from in-memory fitting to big data sets. In: Abramowicz W, Klein G, editors. *Business Information Systems Workshops.* New York: Springer International Publishing. (2020) p. 272–84. doi: 10.1007/978-3-030-61146-0_22
15. Shi X, Nikolic G, Epelde G, Arrue M, Van-Dierdonck JB, Bilbao, R, De Moor B. An ensemble-based feature selection framework to select risk factors of childhood obesity for policy decision making. *BMC Medical Inform Decis Mak.* (2021) 21:222. doi: 10.1186/s12911-021-01580-0
16. Cross-filtering dashboards. Available online at: https://docs.looker.com/dashboards/cross-filtering (accessed May 7, 2020).
17. Microsoft Azure. Azure Active Directory B2C. Available online at: https://azure.microsoft.com/en-us/services/active-directory-b2c/ (accessed May 7, 2020).
18. OpenVA. Available online at: https://github.com/pekka-siltanen/vttopenva (accessed May 7, 2020).
19. Watson Natural Language Understanding. Available online at: https://www.ibm.com/cloud/watson-natural-language-understanding (accessed May 7, 2020).
20. Greenhalgh T. How to read a paper. The medline database. *BMJ.* (1997) 315:180–3. doi: 10.1136/bmj.315.7101.180
21. Srinivasan S, Libbus B. Mining MEDLINE for implicit links between dietary substances and diseases. *Bioinformatics.* (2004) 20:i290–6. doi: 10.1093/bioinformatics/bth914
22. Kibana. Available online at: https://www.elastic.co/kibana (accessed May 7, 2020).
23. Pita Costa J, Fuart F, Grobelnik M, Santanam R, Chenlu S, Carlin P, et al. Text mining open datasets to support public health. In: *Conf Proceedings of WITS.* Soeul: WITS Research Community (2017).
24. Leban G, Fortuna B, Brank J, Grobelnik M. Event registry: learning about world events from news. *Proceedings of the 23rd International Conference on World Wide Web; 2014 Apr; Seoul, Korea.* New York: Association for Computing Machinery. (2014).
25. Pita Costa J, Fuart F, Stopar L, Grobelnik M, Mladenic D, Kosmerlj A, et al. Health news bias and epidemic intelligence for public health. In: *Proceedings of the SiKDD.* Ljubljana: Institute Jozef Stefan (2019).
26. Cleland B, Wallace J, Bond R, Black M, Mulvenna M, Rankin D, et al. Meaningful integration of data analytics and services in MIDAS project: engaging users in the co-design of a health analytics platform. *Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI-2018); 2018 Jul 4-6; Belfast, Northern Ireland.* Swindon: BCS Learning & Development Ltd. (2018).
27. Cleland B, Wallace J, Bond R, Muuraiskangas S, Pajula J, Eplede G, et al. Usability evaluation of a co-created big data analytics platform for health policy-making. In: Yamamoto S, Mori H, editors. *Human Interface and the Management of Information. Visual Information and Knowledge Management.* Cham: Springer: Lecture Notes in Computer Science. (2019).
28. Bond RR, Finlay DD, Nugent CD, Moore G, Guldeniring D. A usability evaluation of medical software at an expert conference setting. *Comput Methods Programs Biomed.* (2014) 113:383–95. doi: 10.1016/j.cmpb.2013.10.006
29. Bond RR, Van Dam E, Van Dam P, Finlay D and Guldenring D. Evaluating the human-computer interaction of 'ECGSim': a virtual simulator to aid learning in electro- cardiology. In: *Computing in Cardiology Conference (CinC).* (2015) p. 409–412. doi: 10.1109/CIC.2015.7408673
30. CORDIS EU Research Results. MIDAS Framework User Guide. Available online at: https://cordis.europa.eu/project/id/727721/results (accessed Feb 10, 2021).
31. Costa JP, Grobelnik M, Fuart F, Stopar L, Epelde G, Fischaber S, et al. Meaningful big data integration for a global COVID-19 strategy. *IEEE Comput Intellig Magazine.* (2020) 15:51–61. doi: 10.1109/MCI.2020.3019898

Conflict of Interest: SF was employed by Analytics Engines. JPC, MG, and LS were employed by Quintelligence. PP was employed by International Business Machines Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Public Health

OPINION
published: 12 April 2022
doi: 10.3389/fpubh.2022.839838

Efficacy of Single-Channel EEG: A Propitious Approach for In-home Sleep Monitoring

B. L. Radhakrishnan[1]*, **E. Kirubakaran**[2], **Immanuel Johnraja Jebadurai**[1], **A. Immanuel Selvakumar**[3] and **J. Dinesh Peter**[1]

[1] Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Karunya Nagar, Coimbatore, India, [2] Department of Computer Science and Engineering, Grace College of Engineering, HWP Colony, Thoothukudi, India, [3] Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

Keywords: sleep, sleep disorders, sleep monitoring, remote sleep monitoring, sleep quality, sleep stage classification, single-channel EEG

1. INTRODUCTION

Sleep is a pivotal biological process and has generally been accepted as a critical factor in human health. Even though the whole function of sleep is not very well studied, it is associated with physical and mental wellness (1, 2). The transient disruptions in regular sleep patterns known as acute sleep deprivation impair cognitive skills. In contrast, long-term sleep abnormalities such as chronic sleep deprivation are related to disease development (3, 4).

Poor sleep quality (5) (as outlined by National Sleep Foundation recommended parameters sleep latency, wake after sleep onset, number of >5 minutes awakenings, and sleep efficiency) has an association with a slew of major medical conditions ranging from obesity and diabetes to neuropsychiatric disorders (6–8). Recent research findings also reveal the associations of poor sleep quality with cardiometabolic risk (9), diabetes (7), weight gain (6), impaired appetite (10), cognitive decline (11), mood changes (12), depression (13), immune function (14), and cancer (15). Global sleep trends indicate that average sleep time is diminishing (16, 17). Besides, sleep-related disorders are on the rise (18, 19). Considering these trends and the importance of sleep for health, a better understanding of sleep characteristics is a public health goal (20, 21).

Polysomnography (PSG) is the gold standard for objective sleep physiology evaluation and has proven to be the most helpful tool for diagnosing sleep-related breathing disorders such as obstructive sleep apnoea, central apnoea, hypopnea, and other respiratory disorders, and also for screening less common sleep disorders, including neurological disorders, such as narcolepsy, parasomnias and seizure disorders, restless legs syndrome and periodic limb movement disorder, depression with insomnia, and circadian rhythm sleep disorder in-clinic sleep assessment and sleep disturbances treatment (22, 23). A PSG study records distinct physiological signals, including electroencephalogram (EEG), electrooculogram (EOG), electromyography (EMG), electrocardiogram (ECG), respiration, pulse oximetry, and other parameters (24). After acquiring the PSG recordings, it is converted into 30-s epochs, and each epoch is mapped to a particular sleep stage (such as N1, N2, SWS, W, REM) manually by a sleep expert or technician, termed as sleep scoring (sleep staging/sleep stage classification). Traditionally more than one technician is involved in this process to avoid biases in marking sleep stages. The accuracy of sleep scoring depends on the expertise of the technicians (25). Although PSG use in clinical sleep medicine has significant benefits, the high cost is a barrier to its accessibility to many populations. Moreover, while undergoing an overnight PSG test, factors such as the unfamiliar sleeping environment with limited privacy, skin irritation due to electrode adhesion, and the numerous leads attached

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Venkatesan Rajinikanth,
St. Joseph's College of Engineering,
India

***Correspondence:**
B. L. Radhakrishnan
blrk.research@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 20 December 2021
Accepted: 21 March 2022
Published: 12 April 2022

Citation:
Radhakrishnan BL, Kirubakaran E, Jebadurai IJ, Selvakumar AI and Peter JD (2022) Efficacy of Single-Channel EEG: A Propitious Approach for In-home Sleep Monitoring. Front. Public Health 10:839838. doi: 10.3389/fpubh.2022.839838

to the person, could obstruct sleep, undermining the accuracy of the recordings (26). Recent advancements in technology play an indispensable role in developing reliable portable monitors (PMs) that support the evaluation of sleep in-home and assist in overcoming the limitations of in-clinic PSG assessment (27). In general PMs are categorised as type 2 (at least seven channels), type 3 (minimum of four channels) and type 4 (either one or two channels) (28).

After considering the underpinning proof of a connection between sleep and wellness, challenges in the traditional sleep evaluation, the need for quality data for the sleep clinicians, this article discusses the significant importance of single-channel EEG in home-based sleep monitoring and analysis. Besides, to further explore the opportunities behind single-channel EEG this article highlights the need for in-home sleep monitoring and its challenges and the role of AI in sleep monitoring. Finally, the challenges in the presented technique and opportunities for future research are presented.

2. IN-HOME SLEEP MONITORING

In-home sleep monitoring is gaining popularity because of its convenience, non-invasive, and self-administrable. In-home sleep monitoring usually uses type 4 sleep monitoring devices. These devices are either picked up by the patients from the clinic or delivered to their homes (29, 30).

2.1. Type 4 Sleep Monitoring

A wearable sleep monitoring system (type 4) is widely used in-home sleep monitoring; these systems detect sleep stages based on any one or two of the signals such as brain waves, heart rate, pulse rate, respiration rate, movements, and other types of signals (body temperature, snoring, etc.) Even though different signals are used for sleep stage analysis, brain signals(EEG) provide more accuracy in specific sleep stage detection and analysis (27). For sleep professionals looking for a more trustworthy, long-term documented total sleep time evaluation, single-channel EEG could be a helpful tool (31). Affordable medical-grade EEG on a large scale is conceived by reducing the electrodes, making it more comfortable, miniaturised, and low-cost. Neurosky (single channel system), Muse and Melon (3-4 electrodes), iBrain, Zeo, and Ear-EEG are consumer-grade products available in the market (32–34).

2.2. COVID and Sleep

Due to the COVID-19 pandemic, people are staying indoors; sleep habits have changed. Sleep disorders are common during the COVID-19 pandemic, affecting roughly 40% of the public and healthcare population (35, 36). Furthermore, non-emergency health services like sleep labs were shut to reduce the infection rate (37). During the pandemic, there has been a significant increase in the usage of PMs for sleep monitoring. Most healthcare providers are expecting this trend to continue in the future (38).

2.3. Need for In-home Sleep Monitoring

Sleep must be tracked in a free-living setting and in an unobtrusive manner to ensure that the sleep captured is as representative of regular sleep as feasible in order to understand the role of sleep in health and disease. Due to the limitations of PSG, most people are only monitored for a single night. Monitoring during one night, on the other hand, is insufficient to ascertain the actual sleep condition. Long-term, at-home monitoring is required to optimise effectiveness and obtain proper follow-up (29). There are currently various options available for sleep monitoring outside the laboratory using type 4 PM devices (39). Compared to the traditional PSG, the single-channel scheme will save money (40) and make data collection much more straightforward, and valuable in a situation like the COVID-19 pandemic (41, 42). A comparison work done by Lucey et al. (43) show that single-channel EEG can assess REM, combined Stages N2 and N3 sleep, and a variety of other indicators, including frontal slow-wave activity, in a way that is equivalent to polysomnography. Therefore, single-channel EEG based PMs can serve as a better alternative to traditional sleep monitoring, and a better solution for in-home monitoring (44). The significant advantage of in-home sleep monitoring is convenience, level of comfort, and less expensive (45).

2.4. In-home Sleep Monitoring Challenges

A home sleep study records your breathing patterns while you sleep in your own bed using small, portable monitoring equipment (30). Even though in-home sleep study is convenient, comfortable, easily accessible and less expensive, it is having a set of challenges listed below.

2.4.1. End User Requirements and Acceptance

An excellent in-home sleep monitoring system design must consider the way the sensor is connected to the body and its attachment method. Users may not prefer the inconvenient method (using adhesive) of connecting sensors to the body and connecting too many sensors. Hence, striking a better balance between user's requirements and acceptance is necessary to design an in-home sleep monitoring device (30, 46).

2.4.2. Long-Term Monitoring

The capacity to do prolonged monitoring is essential for an effective sleep monitoring system. For reliable results and early detection of aberrant sleep abnormalities, long-term monitoring is required. In order to accomplish this, sleep monitoring devices should be low-cost, simple-to-use, and easily accessible (46). The capacity to do prolonged monitoring is essential for an effective sleep monitoring system. For reliable results and early detection of aberrant sleep abnormalities, long-term monitoring is required. In order to accomplish this, sleep monitoring devices should be low-cost, simple-to-use, and easily accessible. To this end, a bed based sensor (47), ear-based EEG (48), Wireless polysomnography system based on the Internet of Things (49), and posture recognition based sleep monitoring (50) proposed.

2.4.3. Seamless Data Sharing With Healthcare Providers

Sleep can be monitored using wearables for a wide range of physical and mental disorders. The majority of the studies relied on commercially accessible gadgets that are linked to smartphones or tablets (34, 51). Wearables can be used to track sleep data. Sleep data can be collected and sent over the internet to a remote clinical on-premise server or cloud server for further analysis, evaluation, decision-making, and treatment. In the captured data, applying machine or deep learning to evolving trends, and instantly alerting patients, nurses, and physicians is a powerful ability (46). Sharing data from a remote location is still a challenging task due to connectivity issues (52).

2.4.4. Data Privacy Issues

Long-term sleep data collection is more comprehensive and diverse. There is an increased chance of user personal information being leaked. Even though, informed consent is used before data collection; given the great value and growing popularity of big data apps, data sharing privacy is a big concern (29). Blockchain technology has the potential to address the issue of privacy and secure data sharing. Implementation of blockchain in real-life applications are in the inception stage and more understanding is needed (53).

3. ROLE OF AI IN SLEEP STAGE CLASSIFICATION

Sleep technicians must verify each epoch manually to perform the sleep scoring, and it has limitations such as labour-intensive and time-consuming and inter-rater variability (25). Kappa (κ) measures the manual sleep scoring performance to estimate interrater reliability, representing an agreement between epoch-to-epoch. The benchmark κ value against human-AI algorithm agreement is 0.68–0.76 approximately (54). Analyzing hours of patient sleep records is not easy due to the aforementioned limitations. Therefore, researchers extensively use machine-learning (ML) and deep-learning (DL) techniques to score the sleep stages from the sleep data automatically (55).

ML models' performance depends on the representative features extracted from the EEG signals. The features are either extracted from 30-s data (epoch) or sub-bands of decomposed signals. Specific studies have extracted features from both the 30-s epochs as well as sub-bands or only from the decomposed signals. The commonly used EEG signal decomposition techniques used in the literature include EMD (empirical-mode decomposition), EEMD (ensemble-empirical-mode decomposition), wavelet-based, FFT (fast Fourier-transform), and frequency such as alpha, gamma, etc., based decomposition (56–58). Once the features are extracted from the EEG signal, it is given as an input (training/test data) to the ML model. Sleep scoring is a multi-class classification problem (mapping five/six sleep stages based on input features). Researchers have tested multiple ML models for automated sleep scoring systems *viz.*: tree-based models (Random forest, SVM, XGBoost, etc.), clustering (KNN), and an amalgamation of distinct models known as ensemble learning (stacking, boosting, bagging, and blending). These approaches achieved 74.5 to 91.9 % accuracy levels for a standard five-class classification (57, 59–63).

Data dimensionality is a notable concern in ML-based systems when dealing with high data-dimension like PSG, which overfit the ML model. Besides, ML-based sleep scoring systems have distinct phases, such as feature extraction, selection, and classification that run as separate tasks. Recent improvements in ML facilitate ways to run discrete tasks together. DL models provide encouraging results in sleep scoring using single-channel EEG based systems (64). The literature's proposed DL-based sleep scoring systems generate features from the input sequence or employ the manually extracted features. A few existing works have used 90- and 60-s epochs instead of traditional 30-second epochs for classification in order to represent the temporal relations between the sleep stages and enhance the sleep scoring results (65). Authors have explored many individual and cascaded DL models for automated sleep scoring systems. Some of the recently proposed DL based works are: convolutional neural network (CNN) (65, 66), IITNet, a CNN and Recurrent Neural Networks (RNN) based network (67), SingleChannelNet (SCNet), a CNN based model. The DL based approaches achieved 83.9 to 92.9% for a standard five-class classification (68), SleepStageNet (69), Long short-term memory (LSTM)-RNN (70). The **Figure 1** illustrates the overview of sleep data extraction and analysis pipeline using cloud computing. The EEG signals are extracted from the individuals suspected of sleep-related issues. Then the signals are filtered to remove the noise and converted into 30-s epochs for identifying the specific sleep stages. Next, features are extracted manually using the methods mentioned above or fed the signals to the DL model for automatic feature extraction. The trained model is used to detect sleep stages automatically.

4. ROLE OF CLOUD IN SLEEP MONITORING

Remote health monitoring is one of the key benefits of the digital era; it allows for remote tracking and monitoring of an individual's health-related factors, as well as sharing with healthcare specialists (71). Recent advancements in sensors, microcontrollers, and communication systems have made significant advances in remote healthcare, allowing for collecting health data from individuals. The Internet of Things (IoT) has pushed the healthcare industry to adopt completely digitised e-health systems and will undoubtedly play a role in remote healthcare (72). The application of wearable sensors to assess an individual's health or well-being status is most attractive in out-of-hospital circumstances (73).

The IoT and Cloud integration is gaining traction in digital healthcare and is used for digital sleep monitoring and assessment. The modern digital sleep health applications comprise data acquisition, fog, and cloud layers depicted in **Figure 1**. The data acquisition layer uses wearable EEG devices for EEG data acquisition. The fog layer brings the cloud to the end-user devices to perform basic preprocessing of data and initial assessment at the user end. Also, It supplies the

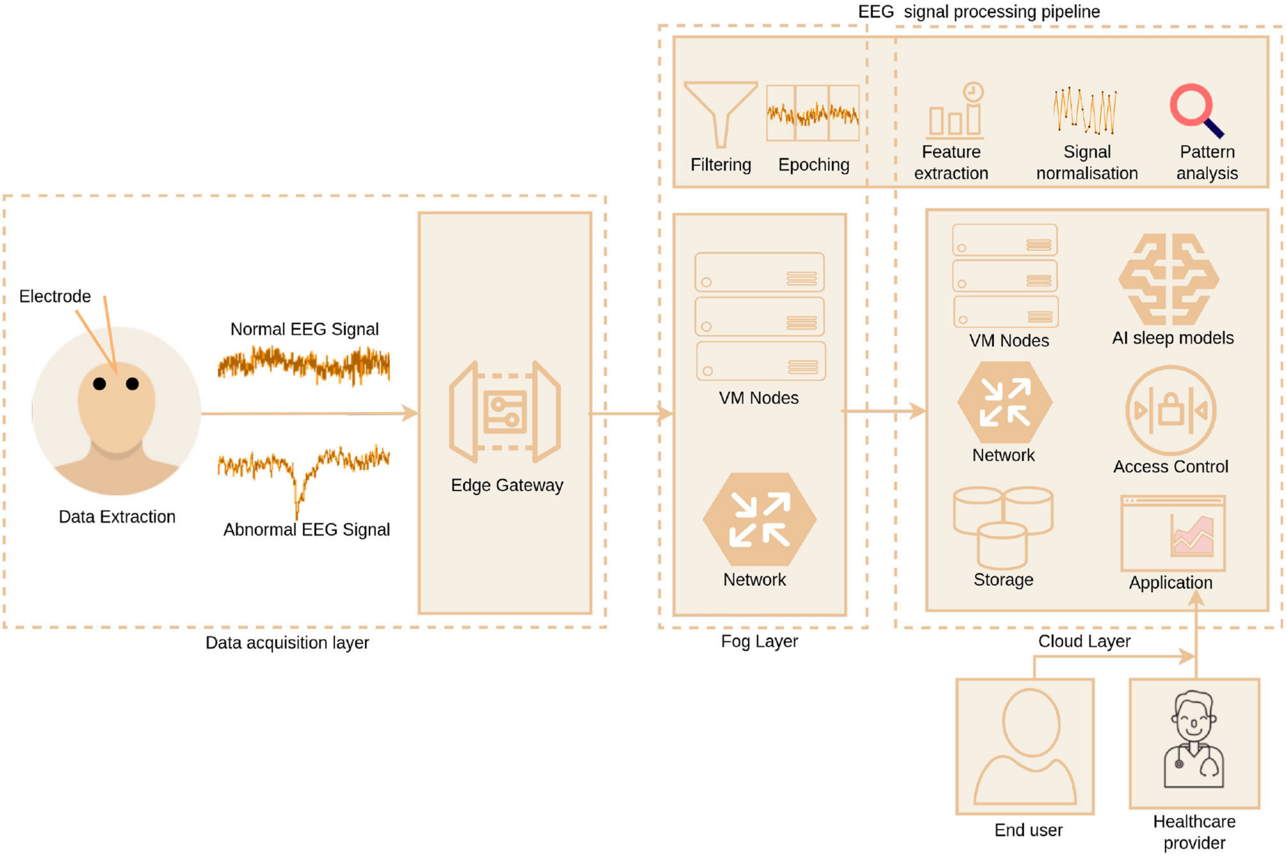


FIGURE 1 | Overview of sleep data extraction and analysis using cloud computing.

homogeneous data to the sleep application running in the cloud that simplifies the processing. The cloud layer provides the compute and storage for the applications. The data acquired by the wearable EEG devices needs to be accumulated and kept safely for future reference, trend analysis and retraining the deployed models. Cloud storage is the viable option to store the sleep data since it offers cost-efficiency, secure sharing, synchronisation and scalability. Healthcare providers leverage ML or DL models using the EEG data stored in the cloud to perform automated sleep stage analysis, resulting in more accurate insightful findings, visualisation, and diagnosis. Analysing EEG data and running sleep stage predictive models support real-time decision making. Despite the merits of the cloud, there are specific challenges when connecting EEG devices in real-time and storing data to the cloud. In the fog layer, EEG data may miss due to improper wearing of the device or malfunction of the sensor. Cloud layer stores and shares EEG data among applications; however, there is growing concern concerning privacy, security and data access (74–76).

5. DISCUSSION

In a technology-driven world, sleep is usually the first thing individuals compromise when they feel pressed against time. Chronic sleep deficiency is associated with cognitive skills, as well as health consequences, such as obesity, diabetes, neuropsychiatric disorders, cardiometabolic risk, impaired appetite, mood changes, depression, immune function, and cancer (6–10, 13–15). COVID-19 pandemic phenomenon increased the sleep-related issues and inaccessibility to sleep clinics. Despite continuous sleep monitoring having its benefits, COVID-19 highlighted its importance (35, 36). Contemporary advancements in sensing techniques, data analytics, and AI systems allow sleep monitoring ubiquitously and unobtrusively (38, 54, 65). Sleep monitoring research using single-channel EEG gains an excess of attraction and momentum since it supports continuous monitoring non-obstructive, aids detecting specific sleep stages accurately and is easy to employ at home (44, 45).

Many ML and DL-based models proposed in the literature achieved better accuracy and kappa (κ) values. Although ASSC systems produce better results, a set of specific challenges exist, such as database variability, channel mismatch, class imbalance, inter-class distinction, computational complexity, and scoring issues (63, 65). Most of the ML and DL-based models proposed in the literature used distinct datasets; these data had been collected from different individuals. Hence, this creates a bias during the comparison of results, and there is a need to investigate the robustness of the model (65). There has been a considerable interest among researchers to design sound in-home single-channel EEG-based sleep monitoring systems. Channel mismatch is another significant factor that hampers the performance of portable ASSC systems (65).

Effectively managing class imbalance (among stages of sleep) is an obligation of an ASSC system. All PSG recordings from healthy persons are generally imbalanced due to the less representation of the S1 stage. Consequently, both ML and DL models render limited performance in classifying the N1 stage (65). To handle this issue, researchers took different approaches. Zhou et al. (62) proposed a method that adjusts the class weights to achieve class balancing, and it has significantly improved the N1 stage recognition to 72.52%. Another author Jiang et al. (58) balanced only the training dataset to make all the classes equal. This approach improved the S1 stage detection from 0.44 to 0.58 (recall score). Moreover, this approach slightly improved the overall detection of less represented classes. In Sors et al. (66) class balancing is done using cost-sensitive learning. That improves N1 and N3 stage results significantly.

Sleep is referred to as a continuous event, and there is no clear-cut boundary between sleep stages. Therefore, different sleep stages, particularly transitioning stages, are tough to distinguish due to subtle inter-class distinctions (65). The sleep stages N1and N2 have similar features. Similarly, N1 and REM also have similar features. This similarity confuses the classifier and even human experts. Moreover, the S1 stage is the transitional stage between W and REM. Hence, among the tested classifiers in literature, N1 stage detection is still challenging (58). Some studies have combined the N1 and N2 stages as light sleep (LS), improving detection accuracy. This limitation can be further improved by incorporating EMG or EOG signals (69). The study by Michielli et al. (70) used a multi-class (N1 and REM combined) and binary class (N1 and REM) approach, this approach improving the N1 stage detection considerably. The performance of the ASSC systems depends on the ML/DL model's complexity. Hence, it is necessary to strike a balance between model complexity and performance (65). Training neural networks like RNN on GPU have tight memory size. This limitation is managed using lesser training sequences.

An in-home ASSC system based on single-channel EEG is required to alleviate the problems in manual scoring and enable the development of a convenient, comfortable, and less expensive in-home sleep monitoring system (63). Single-channel systems have excellent scope in terms of convenience and cost-effectiveness. When comparing the results of recent studies, single-channel EEG shows significant performance. However, there are fewer number studies that validate the results against PSG. Conducting more validation studies among the diverse group (including normal and person with a sleep disorder) using single-channel EEG and validating against PSG would improve the reliability and validity of the ASSC systems. A cloud-based trained and tested framework is necessary to provide accurate multi-model sleep scoring and analysis, seamless data sharing, and facilitate connecting health providers. The importance of continuous monitoring is evident in the literature. Continuous monitoring accumulates more data, detects sleep anomalies, and predicts health-related consequences. Cloud providers offer HIPPA compliance cloud storage that supports securely storing and sharing data. All the services offered by the cloud providers may not be HIPPA compliant. Therefore, before adopting a cloud service, it is essential to verify its HIPPA compliance. Convenience and accuracy is the prime objective of single-channel EEG systems. The missing data issue in the fog layer can be addressed by various methods such as tensor factorization (77). Some studies suggested that adding EMG or EOG sensors may improve the discrepancies in detecting accuracy among N1, N2, and REM stages. However, it causes inconvenience and obstructs regular sleep. Therefore, future research could incorporate heart rate and body movement signals (smartwatches or wearables) with single-channel EEG to improve sleep scoring.

AUTHOR CONTRIBUTIONS

BR, EK, IJ, AS, and JP contributed to the study conception and design, literature review, interpretation, and manuscript preparation. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Schwartz JR, Roth T. Neurophysiology of sleep and wakefulness: basic science and clinical implications. *Curr Neuropharmacol.* (2008) 6:367–78. doi: 10.2174/157015908787386050
2. Clement-Carbonell V, Portilla-Tamarit I, Rubio-Aparicio M, Madrid-Valero JJ. Sleep quality, mental and physical health: a differential relationship. *Int J Environ Res Publ Health.* (2021) 18:460. doi: 10.3390/ijerph18020460
3. Dawson D, Reid K. Fatigue, alcohol and performance impairment. *Nature.* (1997) 388:235.
4. Bertisch SM, Pollock BD, Mittleman MA, Buysse DJ, Bazzano LA, Gottlieb DJ, et al. Insomnia with objective short sleep duration and risk of incident cardiovascular disease and all-cause mortality: sleep heart health study. *Sleep.* (2018) 41:zsy047. doi: 10.1093/sleep/zsy047
5. Ohayon M, Wickwire EM, Hirshkowitz M, Albert SM, Avidan A, Daly FJ, et al. National Sleep Foundation's sleep quality recommendations: first report. *Sleep Health.* (2017) 3:6–19. doi: 10.1016/j.sleh.2016.11.006
6. Kim HO, Kang I, Choe W, Yoon KS. Sleep duration and risk of obesity: a genome and epidemiological study. *World Acad Sci J.* (2021) 3:1. doi: 10.3892/wasj.2021.91
7. Lu H, Yang Q, Tian F, Lyu Y, He H, Xin X, et al. A meta-analysis of a cohort study on the association between sleep duration and type 2 diabetes mellitus. *J Diabetes Res.* (2021) 2021, 8861038. doi: 10.1155/2021/8861038
8. Kammerer MK, Mehl S, Ludwig L, Lincoln TM. Sleep and circadian rhythm disruption predict persecutory symptom severity in day-to-day life: a combined actigraphy and experience sampling study. *J Abnormal Psychol.* (2021) 130:78. doi: 10.1037/abn0000645
9. Nagai M, Hoshide S, Kario K. Sleep duration as a risk factor for cardiovascular disease-a review of the recent literature. *Curr Cardiol Rev.* (2010) 6:54–61. doi: 10.2174/157340310790231635
10. Knutson KL, Spiegel K, Penev P, Van Cauter E. The metabolic consequences of sleep deprivation. *Sleep Med Rev.* (2007) 11:163–78. doi: 10.1016/j.smrv.2007.01.002

11. Spira AP, Chen-Edinboro LP, Wu MN, Yaffe K. Impact of sleep on the risk of cognitive decline and dementia. *Curr Opin Psychiatry.* (2014) 27:478. doi: 10.1097/YCO.0000000000000106
12. Selvi Y, Gulec M, Agargun MY, Besiroglu L. Mood changes after sleep deprivation in morningness–eveningness chronotypes in healthy individuals. *J Sleep Res.* (2007) 16:241–4. doi: 10.1111/j.1365-2869.2007.00596.x
13. Becker NB, de Jesus SN, Viseu JN, Stobäus CD, Guerreiro M, Domingues RB. Depression and quality of life in older adults: Mediation effect of sleep quality. *Int J Clin Health Psychol.* (2018) 18:8–17. doi: 10.1016/j.ijchp.2017.10.002
14. Tan HL, Kheirandish-Gozal L, Gozal D. Sleep, sleep disorders, and immune function. In: *Allergy and Sleep*. Cham: Springer (2019). p. 3–15.
15. Lu Y, Tian N, Yin J, Shi Y, Huang Z. Association between sleep duration and cancer risk: a meta-analysis of prospective cohort studies. *PloS ONE.* (2013) 8:e74723. doi: 10.1371/journal.pone.0074723
16. Bonnet MH, Arand DL. We are chronically sleep deprived. *Sleep.* (1995) 18:908–911.
17. Fry C. Sleep deprived but socially connected: balancing the risks and benefits of adolescent screen time during COVID-19. *J Children Media.* (2021) 15:37–40. doi: 10.1080/17482798.2020.1858907
18. K Pavlova M, Latreille V. *Sleep Disorders*. Elsevier. (2019). doi: 10.1016/j.amjmed.2018.09.021
19. Zitting KM, Lammers-van der Holst HM, Yuan RK, Wang W, Quan SF, Duffy JF. Google Trends reveals increases in internet searches for insomnia during the 2019 coronavirus disease (COVID-19) global pandemic. *J Clin Sleep Med.* (2021) 17:177–84. doi: 10.5664/jcsm.8810
20. Stranges S, Tigbe W, Gómez-Olivé FX, Thorogood M, Kandala NB. Sleep problems: an emerging global epidemic? Findings from the INDEPTH WHO-SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia. *Sleep.* (2012) 35:1173–81. doi: 10.5665/sleep.2012
21. Hafner M, Stepanek M, Taylor J, Troxel W, Stolk C. *Why Sleep Matters – the Economic Costs of Insufficient Sleep: A Cross-Country Comparative Analysis*. vol. 6. RAND Corporation. (2016). Available online at: https://circadiansleepcoaching.com/wp-content/uploads/sites/1753/2022/01/RAND-REPORT-2016.pdf
22. Thorpy M. International classification of sleep disorders. In: *Sleep Disorders Medicine.* Springer. (2017). p. 475–84. doi: 10.1007/978-1-4939-6578-6_27
23. Sharma M, Goyal D, Achuth P, Acharya UR. An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank. *Comput Biol Med.* (2018) 98:58–75. doi: 10.1016/j.compbiomed.2018.04.025
24. Berthomier C, Drouot X, Herman-Stoïca M, Berthomier P, Prado J, Bokar-Thire D, et al. Automatic analysis of single-channel sleep EEG: validation in healthy individuals. *Sleep.* (2007) 30:1587–95. doi: 10.1093/sleep/30.11.1587
25. Cesari M, Stefani A, Penzel T, Ibrahim A, Hackner H, Heidbreder A, et al. Interrater sleep stage scoring reliability between manual scoring from two European sleep centers and automatic scoring performed by the artificial intelligence–based Stanford-STAGES algorithm. *J Clin Sleep Med.* (2021) 17:1237–47. doi: 10.5664/jcsm.9174
26. Zhang J, Tang Z, Gao J, Lin L, Liu Z, Wu H, et al. Automatic detection of obstructive sleep apnea events using a deep CNN-LSTM model. *Comput. Intell Neurosci.* (2021) 2021:5594733. doi: 10.1155/2021/5594733
27. Kwon S, Kim H, Yeo WH. Recent advances in wearable sensors and portable electronics for sleep monitoring. *Iscience.* (2021) 24:102461. doi: 10.1016/j.isci.2021.102461
28. Krishnaswamy U, Aneja A, Kumar RM, Kumar TP. Utility of portable monitoring in the diagnosis of obstructive sleep apnea. *J Postgraduate Med.* (2015) 61:223. doi: 10.4103/0022-3859.166509
29. Pan Q, Brulin D, Campo E. Current Status and Future Challenges of Sleep Monitoring Systems: Systematic Review. *JMIR Biomed Eng.* (2020) 5, e20921. doi: 10.2196/20921
30. Shustak S, Inzelberg L, Steinberg S, Rand D, Pur MD, Hillel I, et al. Home monitoring of sleep with a temporary-tattoo EEG, EOG and EMG electrode array: a feasibility study. *J Neural Eng.* (2019) 16:026024. doi: 10.1088/1741-2552/aafa05
31. Hof zum Berge A, Ferrauti A, Meyer T, Pfeiffer M, Kellmann M. Portable polysomnography for sleep monitoring in elite youth rowing: An athlete's gain or the sleep's thief? *Transl Sports Med.* (2021) 4:289–96. doi: 10.1002/tsm2.205
32. NeuroSky. *EEG Sensors - EEG Headsets | NeuroSky*. (2020). Available Online at: http://neurosky.com/biosensors/eeg-sensor/biosensors/
33. Muse. *Muse^{TM} - Meditation Made Easy With the Muse Headband*. (2021). Available Online at: https://choosemuse.com/
34. Looney D, Goverdovsky V, Rosenzweig I, Morrell MJ, Mandic DP. Wearable in-ear encephalography sensor for monitoring sleep. preliminary observations from nap studies. *Ann Am Thoracic Soc*. (2016) 13:2229–33. doi: 10.1513/AnnalsATS.201605-342BC
35. Jahrami H, BaHammam AS, Bragazzi NL, Saif Z, Faris M, Vitiello MV. Sleep problems during the COVID-19 pandemic by population: a systematic review and meta-analysis. *J Clin Sleep Med.* (2021) 17 :299–313. doi: 10.5664/jcsm.8930
36. Radhakrishnan B, Kirubakaran E, Belfin R, Selvam S, Sagayam KM, Elngar AA. Mental health issues and sleep quality of Indian employees and higher education students during COVID-19 lockdown. *Int J Intell Eng Inf.* (2021) 9:193–210. doi: 10.1504/IJIEI.2021.10040086
37. Lee M, You M. Avoidance of healthcare utilization in south korea during the coronavirus disease 2019 (COVID-19) pandemic. *Int J Environ Res Publ Health.* (2021) 18:4363. doi: 10.3390/ijerph18084363
38. Johnson KG, Sullivan SS, Nti A, Rastegar V, Gurubhagavatula I. The impact of the COVID-19 pandemic on sleep medicine practices. *J Clin Sleep Med.* (2021) 17:79–87. doi: 10.5664/jcsm.8830
39. Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell.* (2018) 41:423–43. doi: 10.1109/TPAMI.2018.2798607
40. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans Neural Syst Rehabil Eng.* (2017) 25:1998–2008. doi: 10.1109/TNSRE.2017.2721116
41. Toedebusch CD, McLeland JS, Schaibley CM, Banks IR, Boyd J, Morris JC, et al. Multi-modal home sleep monitoring in older adults. *J. Visual. Exp. JoVE.* (2019) 143: 10.3791/58823. doi: 10.3791/58823
42. Faezipour M, Faezipour M. Efficacy of smart EEG monitoring amidst the COVID-19 pandemic. *Electronics.* (2021) 10:1001. doi: 10.3390/electronics10091001
43. Lucey BP, Mcleland JS, Toedebusch CD, Boyd J, Morris JC, Landsness EC, et al. Comparison of a single-channel EEG sleep study to polysomnography. *J Sleep Res.* (2016) 25:625–35. doi: 10.1111/jsr.12417
44. Hussain Z, Sheng QZ, Zhang WE, Ortiz J, Pouriyeh S. A review of the non-invasive techniques for monitoring different aspects of sleep. *arXiv preprint arXiv:210412964.* (2021).
45. Kundel V, Shah N. Impact of portable sleep testing. *Sleep Med Clin.* (2017) 12:137. doi: 10.1016/j.jsmc.2016.10.006
46. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, et al. Clinical applications of mobile health wearable–based sleep monitoring: systematic review. *JMIR mHealth uHealth.* (2020) 8:e10733. doi: 10.2196/10733
47. Shin J, Chee Y, Park K. Long-term sleep monitoring system and long-term sleep parameters using unconstrained method. In: *Proceedings of Intl Special Topic Conf on Info Tech in BME.* New York, NY (2006).
48. Nakamura T, Goverdovsky V, Morrell MJ, Mandic DP. Automatic sleep monitoring using ear-EEG. *IEEE J Transl Eng Health Med.* (2017) 5:1–8. doi: 10.1109/JTEHM.2017.2702558
49. Lin CT, Prasad M, Chung CH, Puthal D, El-Sayed H, Sankar S, et al. IoT-based wireless polysomnography intelligent system for sleep monitoring. *IEEE Access.* (2017) 6:405–14. doi: 10.1109/ACCESS.2017.2765702
50. Matar G, Lina JM, Carrier J, Riley A, Kaddoum G. Internet of Things in sleep monitoring: an application for posture recognition using supervised learning. In: *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom).* Munich: IEEE. (2016). p. 1–6.
51. Koydemir HC, Ozcan A. Wearable and implantable sensors for biomedical applications. *Ann Rev Anal Chem.* (2018) 11:127–46. doi: 10.1146/annurev-anchem-061417-125956
52. Radhakrishnan B, Kirubakaran E, Ebenezer V, Belfin R, Ting DI, et al. Remote sleep monitoring and 5G. In:*Secure Communication for 5G and IoT Networks.* Springer. (2022). p. 173–95. doi: 10.1007/978-3-030-79766-9_11
53. Yaeger K, Martini M, Rasouli J, Costa A. Emerging blockchain technology solutions for modern healthcare infrastructure. *J Sci Innov Med.* (2019) 2:1. doi: 10.29024/jsim.7

54. Danker-hopfe H, Anderer P, Zeitlhofer J, Boeck M, Dorn H, Gruber G, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *J Sleep Res.* (2009) 18:74–84. doi: 10.1111/j.1365-2869.2008.00700.x
55. Stephansen JB, Olesen AN, Olsen M, Ambati A, Leary EB, Moore HE, et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat Commun.* (2018) 9:1–15. doi: 10.1038/s41467-018-07229-3
56. Boostani R, Karimzadeh F, Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals. *Comput Methods Programs Biomed.* (2017) 140:77–91. doi: 10.1016/j.cmpb.2016.12.004
57. Memar P, Faradji F. A novel multi-class EEG-based sleep stage classification system. *IEEE Trans Neural Syst Rehabil Eng.* (2017) 26:84–95. doi: 10.1109/TNSRE.2017.2776149
58. Jiang D, Lu Yn, Yu M, Yuanyuan W. Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement. *Exp Syst Appl.* (2019) 121:188–203. doi: 10.1016/j.eswa.2018.12.023
59. Hassan AR, Bhuiyan MIH. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybern Biomed Eng.* (2016) 36:248–55. doi: 10.1016/j.bbe.2015.11.001
60. Hassan AR, Bhuiyan MIH. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomed Signal Process Control.* (2016) 24:1–10. doi: 10.1016/j.bspc.2015.09.002
61. Seifpour S, Niknazar H, Mikaeili M, Nasrabadi AM. A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal. *Exp Syst Appl.* (2018) 104:277–93. doi: 10.1016/j.eswa.2018.03.020
62. Zhou J, Wang G, Liu J, Wu D, Xu W, Wang Z, et al. Automatic sleep stage classification with single channel EEG signal based on two-layer stacked ensemble model. *IEEE Access.* (2020) 8:57283–97. doi: 10.1109/ACCESS.2020.2982434
63. Liu C, Tan B, Fu M, Li J, Wang J, Hou F, et al. Automatic sleep staging with a single-channel EEG based on ensemble empirical mode decomposition. *Phys A Stat Mech Appl.* (2021) 567:125685. doi: 10.1016/j.physa.2020.125685
64. Loh HW, Ooi CP, Vicnesh J, Oh SL, Faust O, Gertych A, et al. Automated detection of sleep stages using deep learning techniques: a systematic review of the last decade (2010–2020). *Appl Sci.* (2020) 10:8963. doi: 10.3390/app10248963
65. Li F, Yan R, Mahini R, Wei L, Wang Z, Mathiak K, et al. End-to-end sleep staging using convolutional neural network in raw single-channel EEG. *Biomed Signal Process Control.* (2021) 63:102203. doi: 10.1016/j.bspc.2020.102203
66. Sors A, Bonnet S, Mirek S, Vercueil L, Payen JF. A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomed Signal Process Control.* (2018) 42:107–14. doi: 10.1016/j.bspc.2017.12.001
67. Seo H, Back S, Lee S, Park D, Kim T, Lee K. Intra-and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG. *Biomed Signal Process Control.* (2020) 61:102037. doi: 10.1016/j.bspc.2020.102037
68. Zhou D, Hu G, Zhang J, Wang J, Yan R, Li F, et al. SingleChannelNet: a model for automatic sleep stage classification with raw single-channel EEG. *bioRxiv.* (2021) p. 2020–09.
69. Chen K, Zhang C, Ma J, Wang G, Zhang J. Sleep staging from single-channel EEG with multi-scale feature and contextual information. *Sleep Breath.* (2019) 23:1159–67. doi: 10.1007/s11325-019-01789-4
70. Michielli N, Acharya UR, Molinari F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput Biol Med.* (2019) 106:71–81. doi: 10.1016/j.compbiomed.2019.01.013
71. Albahri OS, Albahri AS, Mohammed K, Zaidan A, Zaidan B, Hashim M, et al. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. *J Med Syst.* (2018) 42:1–27. doi: 10.1007/s10916-018-0943-4
72. Ray PP, Dash D, Kumar N. Sensors for internet of medical things: State-of-the-art, security and privacy issues, challenges and future directions. *Comput Commun.* (2020) 160:111–31. doi: 10.1016/j.comcom.2020.05.029
73. Korhonen I, Parkka J, Van Gils M. Health monitoring in the home of the future. *IEEE Eng Med Biol Mag.* (2003) 22:66–73. doi: 10.1109/MEMB.2003.1213628
74. Aazam M, Huh EN. Fog computing and smart gateway based communication for cloud of things. In: *2014 International Conference on Future Internet of Things and Cloud.* Barcelona: IEEE (2014). p. 464–70.
75. Sangat P, Indrawan-Santiago M, Taniar D. Sensor data management in the cloud: data storage, data ingestion, and data retrieval. *Concurr Comput Pract Exp.* (2018) 30:e4354. doi: 10.1002/cpe.4354
76. Yacchirema DC, Sarabia-Jácome D, Palau CE, Esteve M. A smart system for sleep monitoring by integrating IoT with big data analytics. *IEEE Access.* (2018) 6:35988–6001. doi: 10.1109/ACCESS.2018.2849822
77. Akmal M, Zubair S, Alquhayz H. Classification analysis of tensor-based recovered missing EEG data. *IEEE Access.* (2021) 9:41745–56. doi: 10.1109/ACCESS.2021.3063382

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Public Health

OPINION
published: 26 April 2022
doi: 10.3389/fpubh.2022.850805

Assistive Methodologies for Parkinson's Disease Tremor Management—A Health Opinion

V. Dineshkumar[1], D. Raveena Judie Dolly[1]*, D. J. Jagannath[1]* and J. Dinesh Peter[2]

[1] Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India, [2] Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

Keywords: Parkinson's disease, assistive devices, tremor, postural instability, non-invasive

OPEN ACCESS

Edited by:
Steven Fernandes,
Creighton University, United States

Reviewed by:
Vijaya Aanand A. S.,
Tata Consultancy Services, India
Xavier Anand,
Geisinger Health System,
United States

Correspondence:
D. Raveena Judie Dolly
dollydinesh@karunya.edu
D. J. Jagannath
jagan@karunya.edu

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *08 January 2022*
Accepted: *11 March 2022*
Published: *26 April 2022*

Citation:
Dineshkumar V, Dolly DRJ, Jagannath DJ and Peter JD (2022) Assistive Methodologies for Parkinson's Disease Tremor Management—A Health Opinion. Front. Public Health 10:850805. doi: 10.3389/fpubh.2022.850805

INTRODUCTION

Parkinson's disease makes lives challenging every day due to the evolving and progressive motor symptoms such as tremors, slow movements, postural instability, and stiffness. These physical symptoms can then in turn affect the thoughts, leading to a state of depression. Tremor is an involuntary, unintended, periodic movement of the muscle of one or more parts of the body and can affect the head, legs, or arms, but predominantly affects the hands. This paper provides an overview on controlling the Parkinson's tremor in the hand through assistive methodologies. Non-invasive low cost assistive devices are considered to reduce the hand tremor caused by Parkinson's disease. Artificial Intelligence tools offer insights to evaluate speech disorders of Parkinson's patients. It can also identify them based on facial expressions.

PARKINSON'S—IN THE PERSPECTIVE OF HEALTH

Parkinson's is a neurodegenerative disorder (1, 6, 15) that occurs due to the death of dopaminergic neurons (15). Electrophysiology is a way to examine the patients in the way of past events and physical exams. Our paper aims to provide a detailed survey on Parkinson's tremors and the ways that they can be detected (1), controlled (3, 13, 17), and analyzed (5, 17). The assessment of Parkinson's disease is based on the clinical interview, the physical examination, and structured instruments (15). Drawbacks to the use of clinical ratings include the reliance on real-time human vision to quantify small differences in motion and significant inter-rater variability due to inherent subjectivity in scoring the procedures. Tremor is an involuntary, unintended, periodic movement of the muscle of one or more parts of the body that can affect the head, legs, arms, and predominantly the hands. Parkinson's tremor can be detected with the help of the active particular muscles in the hand at the time of movement (8, 22). To detect the tremor, the subject should sit in a comfortable place and rest their hands (1). The hand tremor in Parkinson's disease is a periodic signal which has a frequency range (1–3, 6) where the number of oscillation/per time can be noted. The frequency can be calculated manually in the time domain by the total number of cycles/per second. But it is not an easy task in the case of random signals with many frequencies. To study the tremor in the human body, knowledge about the natural frequency is very important. Based on the physical properties, each part of the body will oscillate based on its frequency. The oscillation is like a mechanical (1, 5, 17) component for tremors detected a change in their frequency when the mass is loaded. The energy comes from the irregular rate of the motor units. Tremor can be produced in one or more structures in the CNS; it causes an oscillation and transmits with the motor system.

At that point, the tremor is at the origin, with no change in frequency, but the mass is present. A comparison of frequency in the limbs with several oscillations in the generation of tremors is considered. If a single oscillation is produced, all the limbs generate the same frequency. If the frequency is varied, there may be independent oscillators. Those oscillators play an important role in diagnosis. The frequency range of the Parkinson's tremor is between 4 and 7 Hz as a "rest tremor" (1). To detect hand tremors, accelerometer (1–3, 9, 15, 20), Gyrosocope (1, 4, 10, 17), EMG (8, 22) techniques, and IMU (13, 16, 17) motion sesnors are used.

ASSISTIVE DEVICES—FOR HEALTH ENHANCEMENT

Numerous devices have been proposed to improve lives in the medical arena. Certain devices are designed to help clinicians with accurate diagnosis. Certain other devices provide support to the patients. Devices with low cost may be useful for people on lower incomes. This paper highlights an assistive devices for Parkinson's disease patients experiencing tremor in the hands.

Tuned Vibration Absorber

Hashem et al. (2), proposed a tuned vibration absorber is proposed to suppress the vibrations in the human arm experimental model with two degrees of freedom. From the perspective of dynamics, the degree of freedom is actuated by more than two muscles. The DOF model consists of two pairs of springs to replace the muscles. The aim of this technique is to reduce the tremor, with the help of the human arm model theoretically and numerically; one pair of elbow muscles and the other parallel to the shoulder muscle is considered. In the vibration control approach, a spring-mass damper with an oscillatory system is taken to extract or absorb the vibrations. The vibration absorber is made up of another combination of mass-spring dampers which is added to reduce the amplitude of the vibration. The PD tremor has a frequency range of around 2–12 Hz. It involves a broadband vibration control issue; to eliminate this, they implement the vibration absorber. To create a physical prototype, TVA has proof mass, tuning structure, and body. The proof/absorb mass is used to dissipate the vibrational energy. The tuning structure consists of a beam spring and guild slider. The body with the case is attached to the forearm. The frequency response of the two joints determines the correlation in controlled and uncontrolled cases.

Intelligent Glove

Kazi et al. (3) designed an intelligent glove and a piezoelectric actuator are proposed to control unintentional trembling. The IV Training Arm Tremor model (2) is intended to collect the data to induce vibration in the human forearm. The rig holds the hand model in a horizontal axis to match the postural tremor. The unbalanced masses with DC motors (2) are used for the exciting source to enhance the postural tremor behavior. The assessment of the rig and real human hand tremor can be recorded and measured. Accelerometer (1) and laser displacement sensor are used to measure the displacement and acceleration of the hand tremor. The amplitude of vibration can be described in terms of acceleration and displacement in the time and frequency domain, resulting in the suppression of the glove (12, 14, 18). The frequency (2) of acceleration signal, displacement signal (2), and piezoelectric frequency gives an excellent way to reduce the tremor. The advantage of this technique is the glove with IV training arm tremor model which will be able to reduce the tremor.

Signal Sensors

Deep Brain stimulation is a surgical procedure for Parkinson's disease. There has been no accurate monitoring system using this simulation effect till now. A sensor module was devised by Dai and D'Angelo (4) that combines the accelerometer (1, 3, 15) and gyroscope (1) using MEMS technology. The sensor module is placed at the tip of the finger and sends the measured data to the computer through USB. The assessment is done with the help of sensor data and some adaptive algorithms to categorize their severity level with a standard rating scale through linear regression model and lists the scale (UPDRS) values (15, 21) in Graphical User Interface. This method is considered to be more advantageous as it is purely non-invasive. By processing, a spectral analysis and statistical analysis on the sensor data for tremor quantification has been performed.

Adaptive Tremor Cancellation

Pathak et al. (5) suggested reducing the tremor and stabilizing the hand by active cancellation technology. The advantage of this technique is that the device is compact, non-invasive, and lightweight. The ACT consists of a power supply unit, sensor, and motion generating platform. An accelerometer (1, 15) is integrated into the spoon to measure the direction of the spoon in x and y directions. The two DC motors (2, 3) are connected with mechanical burdens and coupled with a spoon in both directions. The peak amplitude of displacement (2) can be analyzed by the signal extraction method.

Cantilever Vibration Control

Srivani Padma et al. (6) proposed a measurement device based on the cantilever vibration method. Due to tremors, strain is varied and this variation creates vibrations on the cantilever and is attained in the Fiber Bragg Grating sensor to measure the vibration from a hand tremor. This device is placed on the backhand and the patient holds it. Because of strain variation, the vibration data can be recorded with the help of a sensor, and FFT is applied to the recorded data to get a frequency response (2, 3). The FBG sensor has a fast response, low fatigue, and self-regard from electromagnetic interference which is considered to be advantageous. This device is used for entry-level diagnostic purposes.

Wearable Motion Tracker

Papini et al. (7) proposed a programmable hand tremor concept to suppress the hand tremor through simulation. This tool is based on the wrist-haptic interface with comparable space, and the wrist is attached to the novel end effector spherical joint. The user's wrist exerts the controlled forces, while the frequency (2, 3) and amplitude can be correlated with the people having tremors.This method proves to be ideal since it is non-invasive. Using an optical motion tracking system, a tracking-based algorithm has been developed, with the help of a tremor signal from the patient. The user stands in front of the table (2) and the device is attached to the user's forearm to explore the objects. This device can be controlled by comparing with trajectories of the tremor-affected people. In this device, they produced a low-frequency motion from the users and focus to improve the functionality of the system.

Adaptive Control System

Xu et al. (8) designed a closed-loop system to examine the tremor from Parkinson's disease. This closed-loop system consists of a DSP module, front module, and TENS module. In the front module, surface EMG (1, 22) applies to the flexor digitorum superficial muscles and amplifies. Conversion of analog into digital through sampling and analysis of the tremor signal using the DSP module and TENS module is discussed. The closed-loop system can be operated in two modes: auto mode and testing mode. The testing mode is used to modify the parameters manually such as amplitude, frequency, and pulse width for the TENS. In auto mode, the DSP module processes the sampling of the surface EMG (22) and detects the presence of tremor through spectrum analysis. If a tremor happens, the DSP gives the trigger signal to TENS to simulate the pulses from the testing mode parameters. To detect tremors due to Parkinson's disease, biphasic current pulses of EMG signal are utilized to suppress the tremors on the lateral surface of the hand based on a predefined threshold, which is considered to be the highlight of the proposed method. Experiments were performed on healthy and tremor patients to validate this closed-loop system.

Wearable Wrist Watch

Jeon et al. (9) proposed a low-power, wearable device that is that measures the tremor signal from Parkinson's patients using a wristwatch type device that consists of a triaxial gyroscope (1) that has (+/-) 2,000 dps and a triaxial accelerometer (1, 3, 15) which has (+/-) 16 g along all the axes. The measurements adopted were displacement (2, 3, 5) and acceleration at the same time for tremor analysis. The device is placed in the wrist and middle fingers of both hands. If a severe tremor symptom occurs, the device could be placed on other fingers. The signals can be recorded when the patient is at a rest position and video is also recorded for further evaluation.

Sensors Activated Wearable Glove

Turkistani et al. (10) proposed a glove (3) with inbuilt vibrator sensors that produce oscillations and minimize hand tremors. The glove also consists of accelerometer (1, 15)-gyroscope (1, 4)-based MEMS (4) motors that can be placed on each finger and interface with the microcontroller which is an easy way to connect the vibration motors and output pins of the Arduino board. The data represents the vibration changes in the x,y,z directions. The vibration motors play a major role to reduce the tremors in the fingers. In this device, the reduction of the tremor in up to 40%.

IoT Interfaced Control

Sachindra Ragul et al. (11) developed an IoT-based device to control and perform frequency analysis of the hand tremors. A flex sensor and an accelerometer (1, 15) are placed in the glove (3, 10) and it is used to measure the amplitude and frequency of the hand tremors. The control module consists of a microcontroller (10) with a Wifi module that is used to upload the tremor data to the cloud and is monitored by the doctors. The Wifi communication can be improved and maintained by particle photons. Most tremors happen in the hand, so by placing a highspeed rotating disc on the hand, the hand tremor is stabilized. The methodology seems to be of very low cost and low power compared with the EMG technique (8, 22).

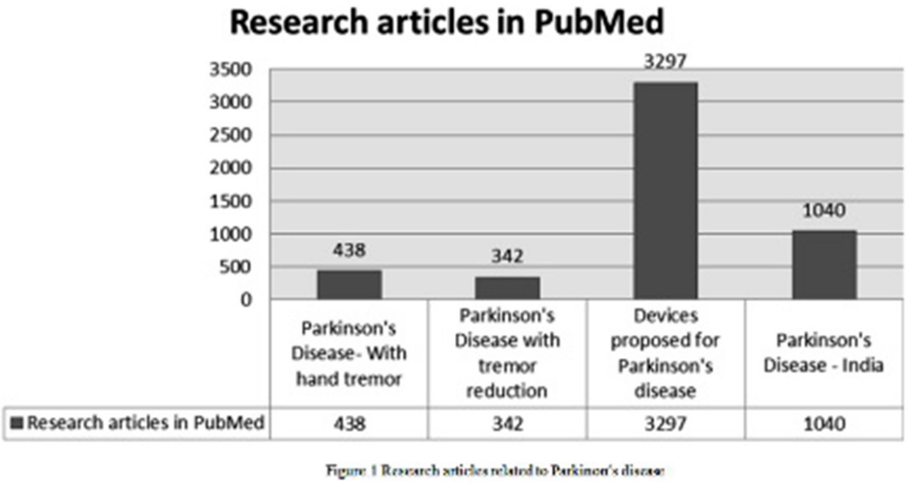


FIGURE 1 | Research articles related to Parkinson's disease.

Role of Artificial Intelligence

Artificial Intelligence offers a way to many medical miracles. It assists humans in every possible way in the present era. AI is a boon to mankind and will serve as a platform for many assistive devices in the field of medicine. To analyze a Parkinson's patient based on facial expressions, speech disorder is a field of research that would definitely provide a new way of life and hope to people. Diagnosis and treatment with the help of AI is developing quickly in the medical field.

DISCUSSION

Human beings face many challenging diseases day to day. Research and technology has become vital to identifying these diseases and diagnosign, curing, and alleviating health conditions. Parkinson's disease make lives difficult every day due to the progressing motor symptoms such as tremor, slow movements, imbalance, and stiffness. These symptoms in turn affect the thoughts, leading to a state of depression. The survey indicates that many research proposals and devices are evolving to manage the effects that occur due to Parkinson's disease. The research articles to diagnose and cure PD from 1957 proposed under various categories out of 79,766 articles are depicted in **Figure 1**.

This indicates that research has been continually in process for early diagnosis and proper treatment of PD.

Tremor is an involuntary, unintended, periodic movement of the muscle of one or more parts of the body which can affect the head, legs, or arms, but predominantly the hands. The tremor (12, 19) caused due to PD begins in a limb and progresses toward the hand and fingers. This is generally termed as a pill—rolling tremor where the movement rubs the thumb and forefinger back-and-forth. It occurs when the hand is at rest. Risk factors include age, heredity, exposure to toxins, and gender.

The vibrations caused due to tremors of Parkinson's disease patients are hoped to be managed and stabilized. The hand movements can be assisted with the help of a device. The assistive device needs to be very comfortable to wear, portable, lightweight, and stabilize the hand when tremors occur. With the assistive device, the patient having tremors can perform their activities without any external support.

AUTHOR CONTRIBUTIONS

VD and DD devised the work, the main conceptual ideas, the proof outline, and worked out almost all of the technical details. DJ and JP worked on the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This article was based on the work supported and funded by the Indian Council of Medical Research (ICMR) RFC Number: (P-10) ITR/Ad-hoc/47/2020-21.

ACKNOWLEDGMENTS

Authors express their earnest thankfulness to the Karunya Institute of Technology and Sciences for the research environment and infrastructure that has been extended.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.850805/full#supplementary-material

REFERENCES

1. Felipe V, Panagiotis K, Shabbir M, Dietrich H. How to do an electrophysiological study of tremor. *Clin Neurophysiol Practice.* (2019) 4:134–42. doi: 10.1016/j.cnp.2019.06.002
2. Hashem SM, Farid Golnaraghi M. Tuned vibration absorber for suppression of rest tremor in Parkinson's diseases. *Med Biol Eng Comput.* (2004) 42:61–70. doi: 10.1007/BF02351012
3. Kazi S, Azizan A, Zain ZM, Mailah M. Performance evaluation of smart glove applied to experimental rig to control human hand tremor for Parkinson's disease. In: *International Conference on Applications of Computer Engineering*, Johor Bahru. (2010).
4. Dai H, D'Angelo LT. Quantitative assessment of tremor during deep brain stimulation using a wearable glove system. In: *IEEE International Workshop of Internet-of-Things Networking and Control (IoT-NC)*, Munich. (2013). doi: 10.1109/IoT-NC.2013.6694054
5. Pathak A, Redmond JA, Allen M, Chou KL. A non-invasive handheld assistive device to accommodate essential tremor: a pilot study. *Mov Disord.* (2014) 29:838–42. doi: 10.1002/mds.25796
6. Srivani Padma G, Umesh S, Asokan U, Srinivas T. Parkinsonian hand tremor measurement device based on fiber Bragg grating sensor. In: *IEEE, International Conference on Smart Sensors and Systems (IC-SSS)*, Bengaluru. (2015). doi: 10.1109/SMARTSENS.2015.7873611
7. Papini GPR, Fontana M. Desktop haptic interface for simulation of hand-tremor. In: *IEEE Transactions on Haptics*, Trento. (2015).
8. Xu FL, Hao MZ, Xu SQ, Hu ZX, Xiao Q, Lan N. Development of a closed-loop system for tremor suppression in patients with Parkinson's disease. *Annu Int Conf IEEE Eng Med Biol Soc.* (2016) 2016:7591063. doi: 10.1109/EMBC.2016.7591063
9. Jeon H, Lee W, Park H, Jeon B. Automatic classification of tremor severity in Parkinson's disease using a wearable device. *Sensors.* (2017) 17:2067. doi: 10.3390/s17092067
10. Turkistani A. *Development of an Effective Portable and Flexible Glove for Hand Tremor Suppression, Corpus ID: 41097451, Medicine.* (2017). Available online at: https://scholarworks.wmich.edu (accessed December 21, 2021).
11. Sachindra Ragul A, Zeeshan M, Prem J, Prasanth S, Preethi Packiashree S. Analysis and control of hand tremor using IoT. *Int J Modern Trends Sci Technol.* (2017) 3:62–7.

12. Zhou Y, Naish MD, Jenkinse ME, Trejosabd AL. Design and validation of a novel mechatronic transmission system for wearable tremor suppression device. *Rob Auton Syst.* (2017) 91:38–48. doi: 10.1016/j.robot.2016.12.009
13. Abbasi M, Afsharfard A, Arasteh R, Safaie J. Design of a noninvasive and smart hand tremor attenuation system with active control: a simulation study. *Med Biol Eng Comput.* (2018) 56:1315–24. doi: 10.1007/s11517-017-1769-9
14. Thilagavathi K, Gayatri SG, Aayisha Rumaana MN, Abhirami V. Smart glove to monitor Parkinson's patients. *Int J Innov Technol Expl Eng.* (2018) 8:393–396.
15. McKay GN, Harrigan TP. A low-cost quantitative continuous measurement of movements in the extremities of people with Parkinson's disease. *Sciencedirect Methods.* (2019) 6:169–89. doi: 10.1016/j.mex.2018.12.017
16. Kavindya P, Kulasekera A, Suresh D, Chathuranga A, Senanayake B. Evaluation of hand tremor frequency among patients in Sri Lanka using a soft glove. In: *Moratuwa Engineering Research Conference (MERC),* Moratuwa. (2020). doi: 10.1109/MERCon50084.2020.9185382
17. Thomas D, Paul A, Kumar M, Shelji S, Varghese P, Shilpa PS. *Gyro Glove: Stabilizing the Lives of Those with the Hand Tremor, Volume 10*. Ernakulam: IJESC. (2020).
18. Kavindya P, Wanasinghe AT, Vithanage IA. Towards a soft hand tremor suppression device for primary care. In: *Conference: IROS 2020 Workshop on Robot-Assisted Training for Primary Care* (2020).
19. Yang H, Shen Y, Zhuang W. A smart wearable ring device for sensing hand tremor of Parkinson's Patients. *Comput Model Eng Sci.* (2021) 126:1217–38. doi: 10.32604/cmes.2021.014558
20. Channa A, Ifrim RC, Popescu D, Popescu N. A-wear bracelet for detection of hand tremor and bradykinesia in Parkinson's patients. *Sensors.* (2021) 21:981. doi: 10.3390/s21030981
21. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS- UPDRS): scale presentation and clinic metric testing results. *Mov Disord.* (2008) 23:2129–70. doi: 10.1002/mds.22340
22. Delrobaei M, Memar S, Pieterman M, Stratton TW, McIsaac K, Jog M. Towards remote monitoring of Parkinson's disease tremor using wearable motion capture systems. *J Neurol Sci.* (2018) 384:38–45. doi: 10.1016/j.jns.2017.11.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fpubh.2022.969268

OPEN ACCESS

EDITED BY
Zhenxing Zhang,
University of Illinois at
Urbana-Champaign, United States

REVIEWED BY
Durai Raj Vincent P M,
VIT University, India
Venkatesan Rajinikanth,
St. Joseph's College of
Engineering, India

*CORRESPONDENCE
Javeria Amin
Javeria.amin@uow.edu.pk

SPECIALTY SECTION
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

RECEIVED 14 June 2022
ACCEPTED 03 August 2022
PUBLISHED 06 September 2022

CITATION
Amin J, Sharif M, Mallah GA and Fernandes SL (2022) An optimized features selection approach based on Manta Ray Foraging Optimization (MRFO) method for parasite malaria classification.
Front. Public Health 10:969268.
doi: 10.3389/fpubh.2022.969268

An optimized features selection approach based on Manta Ray Foraging Optimization (MRFO) method for parasite malaria classification

Javeria Amin[1]*, Muhammad Sharif[2], Ghulam Ali Mallah[3] and Steven L. Fernandes[4]

[1]Department of Computer Science, University of Wah, Wah Cantt, Pakistan, [2]Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan, [3]Department of Computer Science, Shah Abdul Latif University, Khairpur, Pakistan, [4]Department of Computer Science, Design and Journalism, Creighton University, Omaha, NE, United States

Malaria is a serious and lethal disease that has been reported by the World Health Organization (WHO), with an estimated 219 million new cases and 435,000 deaths globally. The most frequent malaria detection method relies mainly on the specialists who examine the samples under a microscope. Therefore, a computerized malaria diagnosis system is required. In this article, malaria cell segmentation and classification methods are proposed. The malaria cells are segmented using a color-based k-mean clustering approach on the selected number of clusters. After segmentation, deep features are extracted using pre-trained models such as efficient-net-b0 and shuffle-net, and the best features are selected using the Manta-Ray Foraging Optimization (MRFO) method. Two experiments are performed for classification using 10-fold cross-validation, the first experiment is based on the best features selected from the pre-trained models individually, while the second experiment is performed based on the selection of best features from the fusion of extracted features using both pre-trained models. The proposed method provided an accuracy of 99.2% for classification using the linear kernel of the SVM classifier. An empirical study demonstrates that the fused features vector results are better as compared to the individual best-selected features vector and the existing latest methods published so far.

KEYWORDS

clusters, malaria, K-mean, MRFO, features

Introduction

Approximately 200 million people die from the mosquito-borne disease malaria every year. It claimed the lives of 405,000 people in 2018, accounting for almost 67% of all children under the age of five (1). It is carried on by the Plasmodium virus, which has four distinct species that can infect people. The disease has been labeled endemic across 38 different parts of the globe. Initial malaria symptoms are fever, chills, vomiting/headache might be potentially moderate and hard to diagnose. The non-treatment of malaria produces sickness and leads to death (2). Malaria is generally linked to poverty, and it is the most common disease in developing countries. The gold standard for malaria diagnosis is microscopic sliding inspection; however, an alternative is a polymerase chain reaction (PCR) test (3). The examination through the sliding microscope method is commonly used for malaria diagnosis. The sensitivity of this method has relied on the pathologist's expertise. The manual slides examination is a tedious and time-consuming task that leads to misdiagnosis. This problem is tackled through a low-cost computer-aided diagnosis system (4). The red blood cells are segmented by a neural network that provides an accuracy of 93.72% (5). The convolutional neural network has been used for malaria detection with an accuracy of 75.39% (6). The deep learning algorithm has been embedded on mobile devices that might be utilized as an application for online malaria cell detection. The proposed technique consists of two steps, where the minimum global screening method is employed for the screening of malaria slides that are subsequently used with CNN with scratch for infected cell classification which provided prediction accuracy of 93.46% (7). The supervised/unsupervised learning models are widely used for the analysis of malaria cells. Several techniques have been developed to detect malaria, but there is still a void in this field, in which larger variation appears among the malaria cells. Furthermore, pertinent feature extraction, as well as the selection, is a challenge for accurate malaria cell segmentation and classification (8). The microscopic images of malaria have noise, poor contrast, illumination, stain, low quality, variations in intensity, size, and irregularity within the region of interest. That's why accurate detection of malaria is a challenging task (9). Therefore, the microscopic malaria images are preprocessed to improve their quality for accurate detection of malaria. After that, hand-crafted features are fused with the deep features and the best features are selected by using the MRFO method, which provides good malaria classification results. Therefore, in this research, a novel model has been proposed for segmentation and malaria cell classification. The foremost contribution steps are described as:

1. Color-based K-mean clustering is applied to three selected clusters to segment the malaria cells
2. Deep extracted features from segmented images are serially fused and then selected active features using MRFO for malaria cell classification.

Article organization is defined as Section Related works discusses the related work; Section Proposed methodology elaborates on the proposed method steps and results are manifested in Section Results and discussion and finally, the conclusion is written in Section Conclusion.

Related works

In the literature, several optimizations, clustering, and classification techniques are widely used for the analysis of malaria cells, some of which are discussed in this section (8, 10–14). The classification techniques used for the diagnosis of the malaria cells, in which AdaBoost (15), Naïve Bayes Tree (16), SVM (17), DT (18), and Linear Discriminant (19), classifiers are involved. Custom convolutional neural models are widely used for the analysis of malaria cells and provide an accuracy of 97.37%. Furthermore, transfer learning models are also commonly utilized for malaria cell classification such as VGG16, AlexNet, DenseNet121, ResNet50, and Xception with an accuracy of 91.50–95.9%. In these transfer learning models, VGG16 and ResNet50 provided competent results compared to the others. Custom CNN models provided improved results compared to transfer learning and they gives an accuracy of 97.37% (20). The Dense Attentive Circular Network is used for malaria cell detection and achieved results are compared to the transfer learning models such as DPN92 and DenseNet121.This method provided an accuracy of 97.47 and 87.88% on DenseNet121 and DPN-92, respectively (21). The pre-trained networks such as ResNet50, AlexNet, and VGG19 are used for malaria classification and provide an accuracy of 93.88, 96.33, and 93.72%, respectively. Features are extracted from pre-trained VGG-16 and input to SVM for discrimination between infected/uninfected cells of malaria with 93.1% accuracy (22). Custom CNN (23–31) and pre-trained efficientnet-b0 model are used for features extraction and they provided accuracy of 97.74 and 98.82%, respectively (20). DCNN model is used for the classification of blood smear images with a 94.79% classification accuracy (32). The original malaria cell images are preprocessed based on the L*a*b* and then extracted for deep features through Dense-Net-169 and Dense-Net 53 for deep feature analysis. Furthermore, the best features are selected using the whale optimization method for malaria cell classification with 99.67% accuracy (33). Malaria cells are classified using deep-sweep software with >0.95 ROC (34). Features are extracted from transfer learning models which are dense-net-201, dense-net-121, Resnet-101, Resnet-50, VGG-16, and VGG-19 for features extraction and input to SVM, NB, and KNN classifiers for malaria cell classification (35). LeNet, GoogLeNet, and AlexNet models are used for feature extraction with 94% accuracy for malaria cell classification (36). Mask-RCNN is used for malaria cell segmentation with a 94.57 correct rate (37). The faster-RCNN is used for model training with a single multi-shot detector

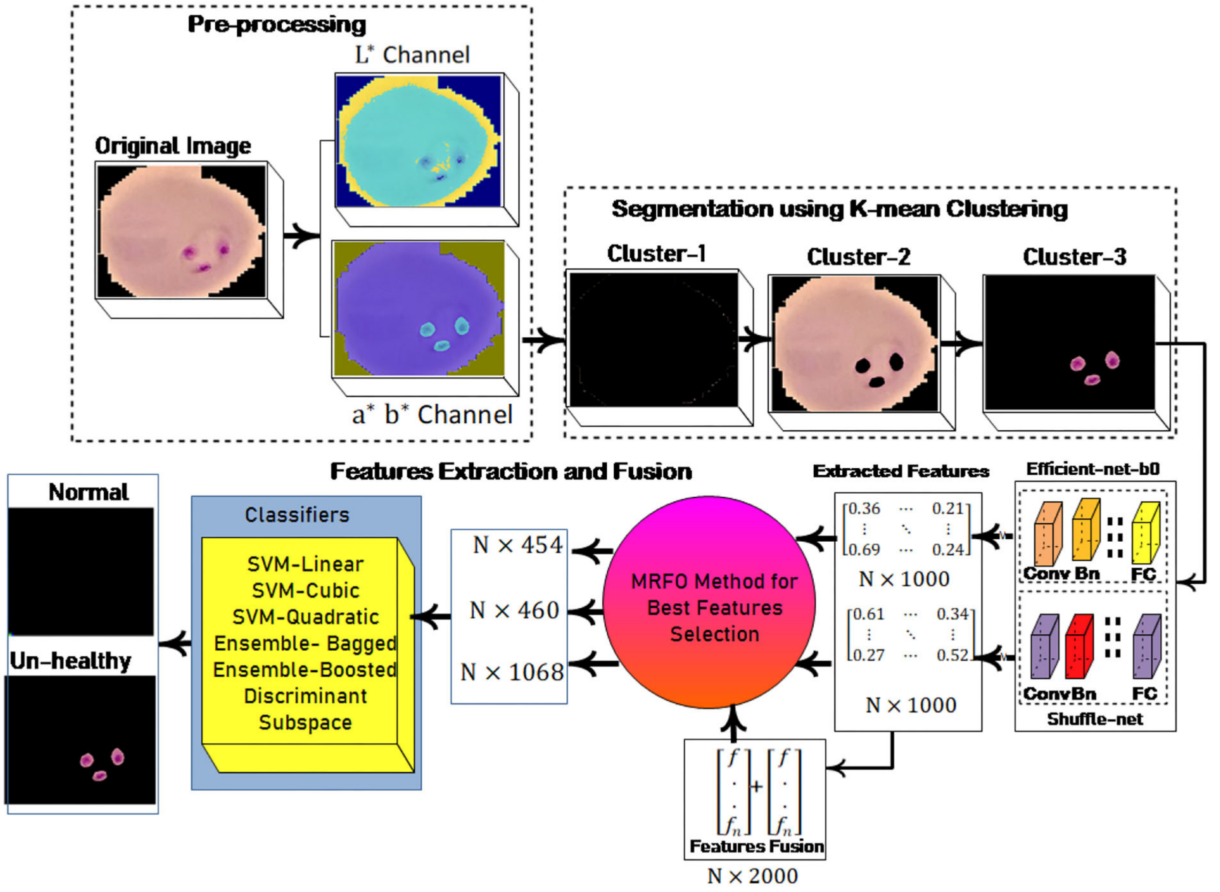


FIGURE 1
Proposed architecture for malaria classification.

SSD for localization of the malaria cells with a 0.94 prediction rate (38).

Proposed methodology

In this research, a method is proposed for the segmentation and classification of malaria cells. The three-dimensional segmentation of the malaria cells is performed using K-mean clustering. After segmentation, the segmented images are fed to the proposed classification model. In this model, features are extracted from pre-trained models such as efficient-net-b0 and shuffle-net (39) with a dimension of $N \times 1{,}000$. In which best $N \times 454$ features from the shuffle-net model and $N \times 460$ features from efficient-net-b0 (40) model. Furthermore, extracted features from both models are fused serially to create a fused feature vector (FV) with a length of $N \times 2{,}000$. The best $N \times 1{,}068$ features are selected out of $N \times 2{,}000$ using the proposed MRFO method and passed to benchmark classifiers to classify the infected and un-infected malaria cells. The proposed method steps are illustrated in Figure 1, a normal image is black because it has no infected region.

Segmentation of the malaria cells using K-mean clustering

The RGB input images φ are converted into the $L^*a^*b^*$ color space for contrast adjustment. The mathematical representation is defined as:

$$\begin{aligned} \varphi L^* &= 116\, h\left(\frac{Y}{Y_\partial}\right) - 16 \\ \varphi a^* &= 500\, [h\left(\frac{X}{X_\partial}\right) - h\left(\frac{Y}{Y_\partial}\right)] \\ \varphi b^* &= 200\, [h\left(\frac{Y}{Y_\partial}\right) - h\left(\frac{Z}{Z_\partial}\right)] \end{aligned} \qquad (1)$$

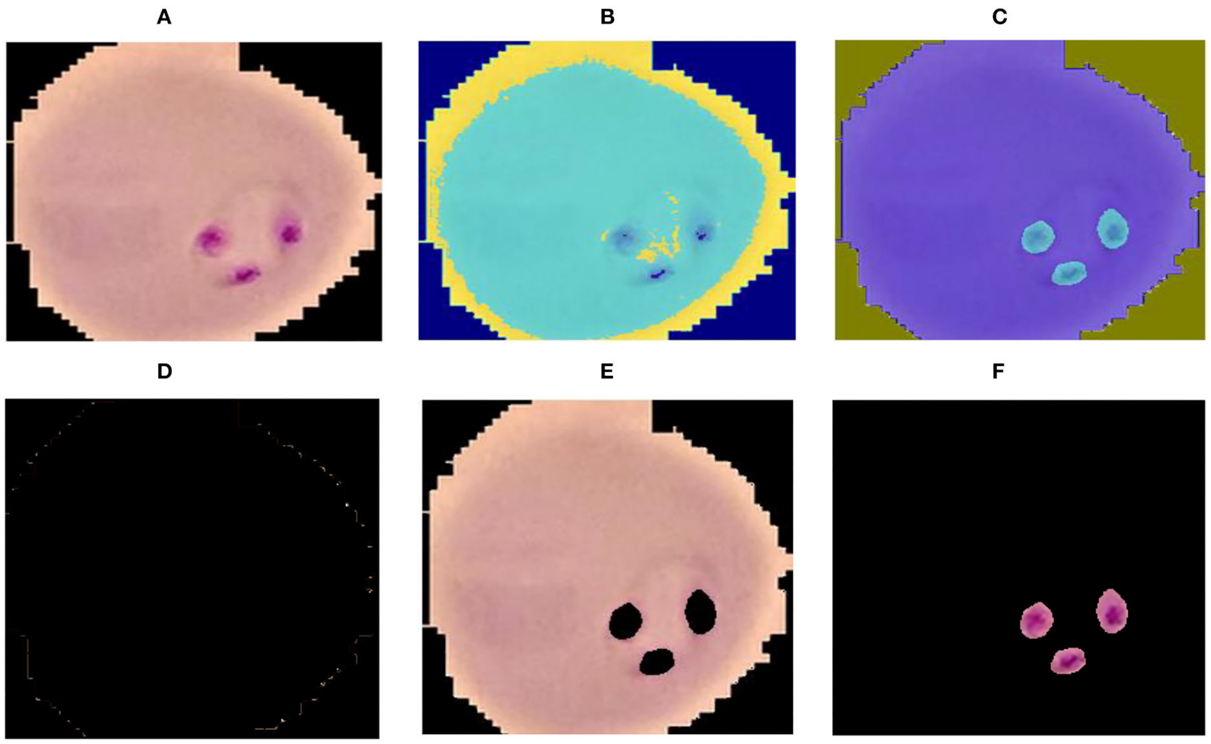


FIGURE 2
Segmentation results **(A)** input image **(B)** luminance channel **(C)** a and b channel **(D)** cluster-1 **(E)** cluster-2 and **(F)** cluster-3.

Here $X_\partial,\ Y_\partial,\ Z_\partial$ denotes the white-tristimulus

$$h(f) = \begin{cases} \sqrt[3]{\mathrm{f}} & f > 0.008 \\ 7.7f + \frac{16}{116} & f \leq 0.008 \end{cases}$$

The K-mean clustering is applied to the luminance channel for segmentation. In K-mean clustering, we uniformly selected the observation centroid c_1 from the input data X (41). The distance is computed to the observation to the c_1 that is represented $d(x_m,\ cj)$. Then centroid c2 is selected randomly from the X data point with the probability is defined as follows:

$$\frac{d^2(x_m,\ c1)}{\sum_{j=1}^{n} d^2(x_j,\ c1)} \tag{2}$$

The distance from each observation to the centroid is computed and allocated observation to the nearest centroid is mathematically expressed as:

$$\frac{d^2(x_m,\ c1)}{\sum_{\{h;x_h\ \varepsilon cp\}} d^2(x_m,\ c1)} \tag{3}$$

where cp denotes the closet centroid across each observation and x_m relate to the cp. In the proposed method malaria-infected region is segmented on the selected $k = 3$ values. The proposed segmentation results are shown in Figure 2.

Malaria cells classification

The malaria cells are classified using the proposed convolutional neural network-based transfer learning model such as efficientnet-b0 and shuffle-net. The pre-trained efficientnet-b0 consists of 290 layers such as convolutional (65), batch-normalization (49), sigmoid (65), element-wise multiplication (65), convolution group (15), average global pooling (16), addition (9), fully connected (FC), addition (15), classification, softmax, and global average pool (16). The shuffle-net consists of 172 layers such as input (1), convolution (49), (49) batch-normalization, ReLU (33), max-pooling (01), average pooling (02), 16 shuffling channels, 1 fully connected (FC), 1 softmax, 1 classification, 15 addition, 01 average global pool, and 2 depth concatenation. This research extracted features from the MatMul FC layer of efficient-netb0 and node-202 FC layer of shuffle-net. The extracted features are fused serially and fed to the MRFO method for the selection of optimum features. The proposed feature extraction and selection process are presented in Figure 3.

Best features selection using MRFO method

Manta Ray Foraging Optimization (MRFO) method is used for best feature selection. MRFO consists of three main steps

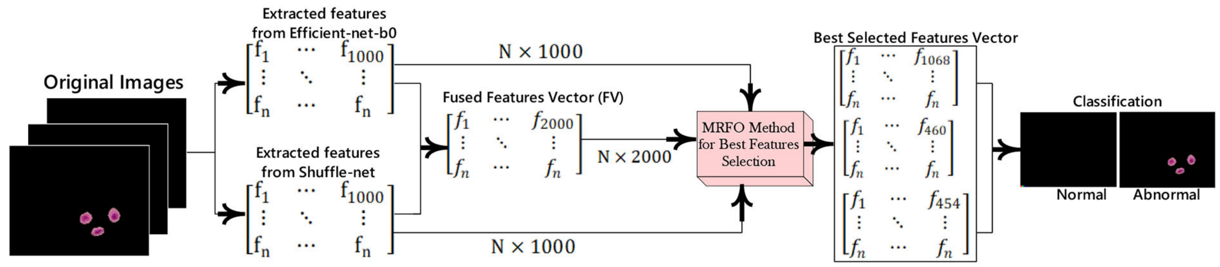


FIGURE 3
Best features extraction and selection process for malaria cell classification.

TABLE 1 Selection of the best features using MRFO.

Lower bound	Upper bound	Threshold	Somersault	Total iterations	Total solutions	Classification error rate
0	***1***	***0.5***	***2***	***100***	***10***	***0.008***
		0.6	3		12	0.009
		0.4	1		09	0.020
		0.7	4		11	0.031
		0.3	5		08	0.091

such as chain fore-aging, cyclone fore-aging, and somersault fore-aging (42–44).

Chain fore-aging

Manta rays might see the direction of the plankton in MRFO and move toward it. The better the position is, the greater will be the concentration of plankton. Although the optimal answer is unknown, MRFO believes that the optimal solution discovered so far has been the plankton that comprises a greater concentration of manta rays that desire to reach and consume it. Manta rays construct a foraging chain by queuing up head to toe. Everyone else moves toward the meal and the individuals in front of it, except for first. So that, everyone gets updated by optimal answers found thus far and explanations ahead of it in each iteration. The mathematical representation of the chain foraging is defined as:

$$\begin{aligned} & x_i{}^d\,(\omega+1) \\ &= \begin{cases} x_i{}^d\,(\omega) + \partial.\left(x^d{}_{Best}\,(\omega) - x_i{}^d\,(\omega)\right) + \alpha.(x^d{}_{Best}\,(\omega) - x_i{}^d\,(\omega)) & i=1 \\ x_i{}^d\,(\omega) + \partial.\left(x^d{}_{i-1}\,(\omega) - x_i{}^d\,(\omega)\right) + \alpha.\left(x^d{}_{Best}\,(\omega) - x_i{}^d\,(\omega)\right) & i=2,\ldots N \end{cases} \quad (4) \\ \alpha &= 2.\partial.\sqrt{\left|log\,(\partial)\right|} \end{aligned}$$

Here ∂ is a random value in the range of [0 1], α represents the learning rate, d denotes the dimension and $x_i{}^d\,(\omega)$ is the i individual at t time.

Cyclone foraging

When a group of the manta rays spots plankton in the intense water, they form an extensive foraging chain and travel in a spiral toward the meal. WOA uses a spiral foraging method that is comparable to this. According to the foraging technique of manta ray swarms, every manta ray moves toward the one ahead of it as well as spiraling toward the meal. The best answer identified through the cyclone foraging technique has better exploitation of the area because all individuals do the search at random with the food to their reference place. This trend is also employed to significantly enhance research. By allocating an unplanned place in the whole research area to the source point, we may need every participant to seek a different place.

$$\begin{aligned} & x_i{}^d\,(\omega+1) \\ &= \begin{cases} x_i{}^d\,(\omega) + \partial.\left(x^d{}_{Best}\,(\omega) - x_i{}^d\,(\omega)\right) + \beta.(x^d{}_{Best}\,(\omega) - x_i{}^d\,(\omega)) & i=1 \\ x_i{}^d\,(\omega) + \partial.\left(x^d{}_{i-1}\,(\omega) - x_i{}^d\,(\omega)\right) + \beta.\left(x^d{}_{Best}\,(\omega) - x_i{}^d\,(\omega)\right) & i=2,\ldots N \end{cases} \quad (5) \end{aligned}$$

$\beta = 2e^{\partial_1 \frac{T-t+1}{T}}.\ \sin(2\pi r_1)$ where β represents the coefficient of weights and T is the total iterations.

Somersault fore-aging

This technique is mainly involved with research and permits MRFO to organize global research. The meal position is considered a hinge in these behaviors. Every person moves to and fro around the hinge, somersaulting to the given assignment. Consequently, they always adjust their placements to the greatest possible position found until now.

$$x_i^{\,d}(\omega+1) = x_i^{\,d}(\omega) + S\,.\left(\partial_2\,.\,x^d{}_{Best} - \partial_3\,.\,x_i^{\,d}(\omega)\right), \quad i = 1\ldots..N \qquad (6)$$

Here's denote the range of the somersault factor, $S = 2$, ∂_2, ∂_3 are random two numbers in [0.1] range. In this research parameters of MRFO are selected for best features selection as presented in Table 1.

Table 1 describes the parameters of MRFO, in which 0.5 threshold,two Somersault, and 100 iterations with 10 total solutions are used for model training due to less error rate these are selected for further processing. This experimental model is converge after the 60 epochs on a 0.128 fitness value as presented in Figure 4.

This method is applied to individual feature vector and the fusion of both feature vectors. In this experiment, we achieve the dimension of $N \times 454$ features from shuffle net and $N \times 460$ features from efficient-net-b0 features. Finally, achieved $N \times 1{,}068$ best-selected features after the fusion of both feature vectors. The selected feature vectors are fed to the ensemble and SVM classifiers for malaria cell classification.

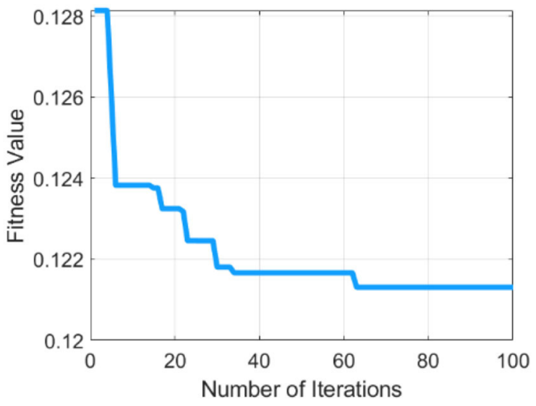


FIGURE 4
Graphical representation of the MRFO method.

Results and discussion

The proposed method results are evaluated on a publically available malaria benchmark dataset. This dataset contains 2,750 of two classes in which 1,375 infected and 1,375 un-infected images are included (45). The proposed method of experimentation is implemented on the Core-i7 window operating system, Nvidia 2070-RTX GPU.

Experiment#1: Malaria classification

The input microscopic malaria cells are segmented using color-based k-mean clustering. In this experiment malaria, cells are classified into two classes using precision (Pi), accuracy (Accy), recall (Rec), and F1-score (F1e) as presented in Tables 2, 3. The proposed method classification results using efficient-net-b0 are shown in Table 2.

In Table 2, features are extracted from the efficientnet-b0 and the best features are selected using MRFO that are fed to the SVM and ensemble classifiers. In this experiment, we achieved a maximum accuracy of 94.68% using a linear classifier as compared to others. Ten-fold cross-validation is used to calculate the categorization results using the best-selected features vector dimension of $N \times 454$ from the shuffle-net as presented in Table 3.

Mean predicted scores are also computed to authenticate the proposed method performance as shown in Table 4; Figure 5.

In Table 3, a significant test is performed in which the ROC–AUC value is computed on bagged, boosted, linear, cubic

TABLE 2 Classification of malaria cells using efficient-net-b0 features.

Classifier	Infected	Un-infected	Accy	Pi	Rec	F1e
Linear	✓		94.68%	0.95	0.95	0.95
		✓	94.68%	0.95	0.95	0.95
Quadratic	✓		93.87%	0.93	0.94	0.94
		✓	93.87%	0.94	0.93	0.94
Cubic	✓		92.22%	0.92	0.92	0.92
		✓	92.22%	0.92	0.92	0.92
Boosted	✓		91.52%	0.92	0.92	0.92
		✓	91.52%	0.90	0.91	0.91
Bagged	✓		91.31%	0.93	0.92	0.92
		✓	91.31%	0.90	0.91	0.90
Subspace discriminant	✓		92.62%	0.95	0.94	0.94
		✓	92.62%	0.90	0.91	0.90

TABLE 3 ROC-AUC values on benchmark classifiers using efficient-net-b0.

Boosted	Bagged	Subspace discriminant + boosted	Subspace discriminant + bagged	Linear	Cubic	Quadratic + linear	Quadratic + cubic	ROC-AUC
✓								0.737
	✓							0.729
		✓						0.726
			✓					0.727
				✓				0.738
					✓			0.726
						✓		0.758
							✓	0.760

TABLE 4 Classification results using shuffle-net.

Classifier	Infected	Un-infected	Accy	Pi	Rec	F1e
Linear	✓		93.67%	0.95	0.93	0.94
		✓	93.67%	0.92	0.94	0.93
Quadratic	✓		94.87%	0.96	0.95	0.96
		✓	94.87%	0.93	0.95	0.94
Cubic	✓		95.16%	0.96	0.95	0.96
		✓	95.16%	0.94	0.95	0.94
Boosted	✓		93.68%	0.94	0.95	0.95
		✓	93.68%	0.93	0.92	0.93
Bagged	✓		94.63%	0.95	0.96	0.95
		✓	94.63%	0.94	0.93	0.93
Subspace discriminant	✓		94.76%	0.94	0.97	0.96
		✓	94.76%	0.95	0.92	0.93

classifiers, and fusion of subspace discriminant + boosted, subspace discriminant + bagged, quadratic + linear, quadratic + cubic classifiers. This experiment achieved the highest ROC-AUC of 0.760 using the fusion of quadratic + cubic classifiers. The classification results based on shuffle-net features are presented in Table 4.

The classification results are mentioned in Table 4, in which we achieved 95.16 accuracies on the SVM cubic classifier, which is far better compared to others. Furthermore, classification results are also computed in terms of ROC–AUC, in which classifiers are used individually as well as the fusion of the boosted, bagged, linear, and cubic kernels with the subspace discriminant and quadratic, respectively. The quantitative analysis of the features vector obtained from the shuffle net is presented in Figure 6; Table 5.

Figure 6 shows the mean predicted scores on the benchmark classifiers. The quantitative computed results are mentioned in Table 5.

In the classification of malaria cells, we achieved a ROC–AUC of 0.865 using a bagged and subspace discriminant + bagged classifier. The classification results are computed using the fused features vector as presented in Table 6.

In the results in Table 6, after applying the fused features vector, we obtained an accuracy of 99.2% on the linear classifier and 97.64% on the quadratic kernel. The fused features vector provides excellent results compared to the individual features vector that authenticates the proposed method novelty. The mean of the correct prediction scores on the fused features vector is presented in Table 7.

Mean values of the correct prediction scores are presented in Table 7, in which the highest ROC-AUC of 0.952 using bagged and subspace discriminant + bagged classifier as compared to others. In Figure 7, a combination of the ROC curve of the selected classifier is plotted as well as the mean predicted probability of the benchmark kernels such as linear, cubic, quadratic, boosted, bagged, and discriminant is plotted.

Figure 7 shows the computed classification results of the individual and fusion of the bags with discriminate, linear with quadratic, and quadratic with cubic. The proposed method results are compared to the current existing methods as presented in Table 8.

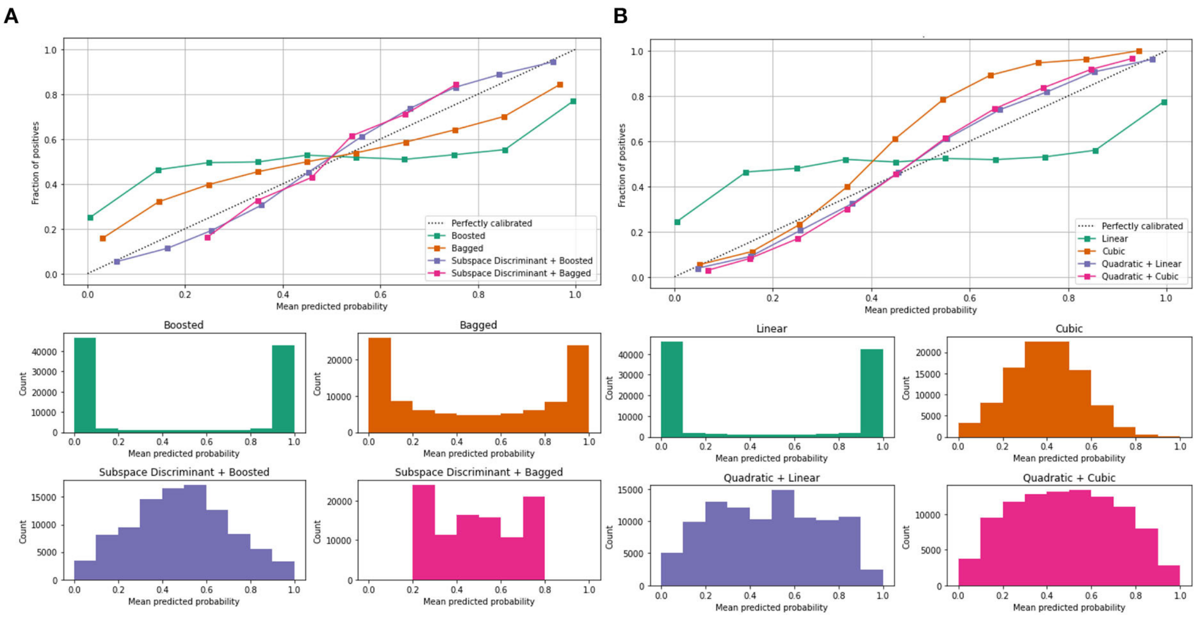


FIGURE 5
Mean predicted scores on the benchmark classifiers **(A)** ensemble **(B)** SVM. The ROC-AUC values are computed on benchmark classifiers as presented in Table 3.

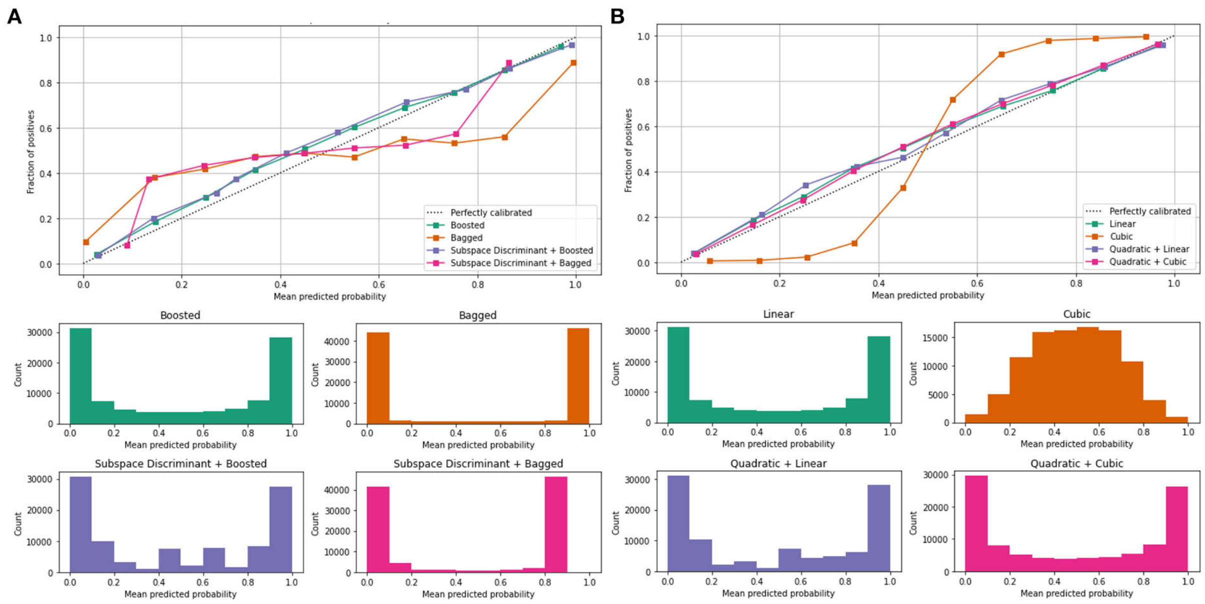


FIGURE 6
Mean predicted scores using shuffle-net features vector **(A)** ensemble and **(B)** SVM.

In Table 8, the proposed classification method results are contrasted with the most recent works published on the same benchmark dataset. In which the pre-trained ResNet-18 model is used for malaria classification it gives an accuracy of 98.68% (46). The classifiers SVM, neural network, and XG-boost are used for malaria classification with the accuracy of 94, 90, and 80%, respectively (47). The DAG–CNN model is used for parasite malaria classification and it gives an

TABLE 5 Quantitative results for malaria classification using shuffle-net features.

Boosted	Bagged	Subspace discriminant + boosted	Subspace discriminant + bagged	Linear	Cubic	Quadratic + linear	Quadratic + cubic	ROC-AUC
✓								0.863
	✓							0.865
		✓						0.862
			✓					0.865
				✓				0.863
					✓			0.863
						✓		0.863
							✓	0.862

TABLE 6 Classification of malaria cells using fused features vector.

Classifier	Infected	Un-infected	Accy	Pi	Rec	F1e
Linear	✓		99.2%	0.99	0.99	0.99
		✓	99.2%	0.99	0.99	0.99
Quadratic	✓		97.64%	0.98	0.97	0.98
		✓	97.64%	0.97	0.98	0.98
Cubic	✓		95.97%	0.97	0.95	0.96
		✓	95.97%	0.95	0.96	0.96
Boosted	✓		96.17%	0.97	0.96	0.96
		✓	96.17%	0.96	0.97	0.96
Bagged	✓		96.39%	0.97	0.96	0.96
		✓	96.39%	0.96	0.97	0.96
Subspace discriminant	✓		97.49%	1.00	0.95	0.97
		✓	97.49%	0.96	1.00	0.98

TABLE 7 Classification of malaria cells in terms of mean predicted scores using fused features vector.

Boosted	Bagged	Subspace discriminant + boosted	Subspace discriminant + bagged	Linear	Cubic	Quadratic + linear	Quadratic + cubic	ROC-AUC
✓								0.951
	✓							0.952
		✓						0.951
			✓					0.952
				✓				0.951
					✓			0.951
						✓		0.951
							✓	0.951

accuracy of 94.79% (32). NC–GCN model is used for feature extraction and malaria cells are classified with 94.17% accuracy (48). The pre-trained inception-v3 and VGG-19 models are used for feature extraction. The average achieved accuracy is 96%.

Compared to the most recent techniques used in this research a method is proposed to investigate the deep features analysis using efficient-net-b0 and shuffle-net. These features are used for classification in two ways. First, the MRFO method is applied to the extracted features of both models and the best-selected features are passed to the classifiers. In the second way, extracted features are serially fused and optimum features are selected using the MRFO method. In this experiment, we observed that the fused

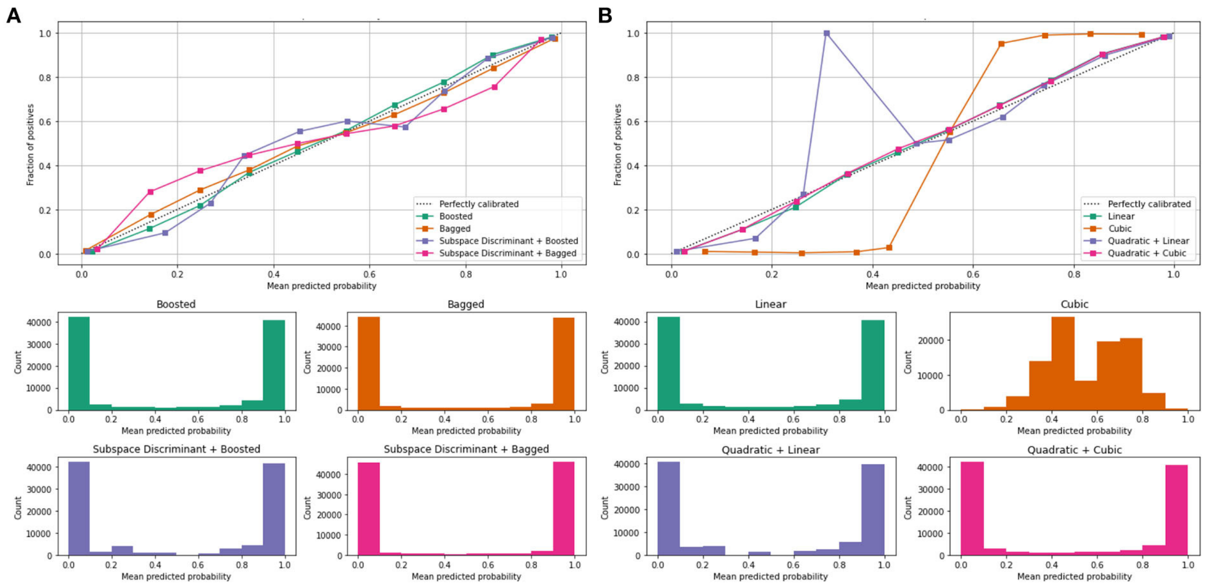


FIGURE 7
Mean predicted scores using fused features vector **(A)** boosted, bagged, and discriminant kernels of ensemble classifier and **(B)** linear, cubic, and quadratic kernels of SVM.

TABLE 8 Proposed classification method results comparison.

Ref#	Year	Accuracy
(46)	2022	97.68%
(47)	2022	94.00% using SVM, 90.00% XG-Boost, 80.00% and neural networks classifier
(32)	2022	94.79%
(48)	2022	94.17%
(49)	2022	96.00%
Proposed Method		99.20%

features vector performs significant improvement in malaria cell classification.

Conclusion

Parasite malaria segmentation and classification is an intricate task due to the large variation and illumination in microscopic malaria images. Therefore, in this research, a method is proposed in which the quality of the input images is improved by applying a pre-processing method. The RGB microscopic malaria images are converted into CIELab color space. The luminance channel is selected for further processing to improve the image contrast. The best-segmented results of the malaria cells are achieved in the third cluster. The segmented images are fed to the proposed classification model. The proposed method more accurately classifies the malaria cells due to features fusion, and optimum features selected by the MRFO method. The proposed method provides an accuracy of 95.16% on the individual best-selected features vector and 99.2% on the fused best-selected features vector. Furthermore, to authenticate the performance of the proposed classification method ROC-AUC values are computed using individual classifiers such as linear, cubic, bagged, boosted, and the fusion of cubic + quadratic, and linear + quadratic, bagged + subspace discrimination, boosted + subspace + discriminant classifiers. In this experiment highest, the 0.95 ROC-AUC was achieved using a bagged + subspace discriminant classifier. In comparison to the most recent approaches, the experimental results show that the fused features vector produces the best outcomes.

Data availability statement

The original contributions presented in the study are included in the article/supplementary files, further inquiries can be directed to the corresponding author/s.

Author contributions

JA performed writing draft, conceptualization, and implementation. MS contributed as the part of result validation team and writing the conclusion of the paper. SF and GM contributed to the data curation, investigation, and literature reviews. SF contributed to the analysis and editing the original draft. GM contributed to the resources and project administration. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. World Health Organization. *World Malaria Report 2021 (2012).* Available online at: https://www.who.int/malaria/publications/world--malaria--report$-$2017/report/en/ (accessed April 4, 2022).

2. Griffith KS, Lewis LS, Mali S, Parise ME. Treatment of malaria in the United States: a systematic review. *JAMA.* (2007) 297:2264–77. doi: 10.1001/jama.297.20.2264

3. Siwal N, Singh US, Dash M, Kar S, Rani S, Rawal C, et al. Malaria diagnosis by PCR revealed differential distribution of mono and mixed species infections by *Plasmodium falciparum* and *P vivax* in India. *PLoS ONE.* (2018) 13:e0193046. doi: 10.1371/journal.pone.0193046

4. Murphy SC, Shott JP, Parikh S, Etter P, Prescott WR, Stewart VA. Malaria diagnostics in clinical trials. *Am J Trop Med Hyg.* (2013) 89:824. doi: 10.4269/ajtmh.12-0675

5. Delgado–Ortet M, Molina A, Alférez S, Rodellar J, Merino A, A. deep learning approach for segmentation of red blood cell images and malaria detection. *Entropy.* (2020) 22:657. doi: 10.3390/e22060657

6. Huertas–Tato J, Martín A, Fierrez J. Camacho D. Fusing CNNs and statistical indicators to improve image classification. *Inform Fusion.* (2022) 79:174–87. doi: 10.1016/j.inffus.2021.09.012

7. Yang F, Poostchi M, Yu H, Zhou Z, Silamut K, Yu J, et al. Deep learning for smartphone–based malaria parasite detection in thick blood smears. *IEEE J Biomed Health Inform.* (2019) 24:1427–38. doi: 10.1109/JBHI.2019.2939121

8. Das DK, Mukherjee R, Chakraborty C. Computational microscopic imaging for malaria parasite detection: a systematic review. *J microsc.* (2015) 260:1–19. doi: 10.1111/jmi.12270

9. Amin J. Segmentation and classification of diabetic retinopathy. (2019) 2:9–19.

10. Setyawan D, Wardoyo R, Wibowo ME, Murhandarwati EE, Jamilah J. Malaria classification using convolutional neural network: a review. In: *2021 Sixth International Conference on Informatics and Computing (ICIC).* (2021). 1–9. doi: 10.1109/ICIC54025.2021.9632998

11. Amin J, Anjum MA, Sharif A, Raza M, Kadry S, Nam Y. *Malaria Parasite Detection Using a Quantum–Convolutional Network.* (2022) 70:6023–6039. doi: 10.32604/cmc.2022.019115

12. Elavarasan D, Vincent PM DR, Srinivasan K, Chang CY, A. hybrid CFS filter and RF–RFE wrapper–based feature extraction for enhanced agricultural crop yield prediction modeling. *Agriculture.* (2020) 10:400. doi: 10.3390/agriculture10090400

13. Mahendran N, PM DR, A. deep learning framework with an embedded–based feature selection approach for the early detection of the Alzheimer's disease. *Comput Biolo Med.* (2022) 141:1050–56. doi: 10.1016/j.compbiomed.2021.105056

14. Dey S, Nath P, Biswas S, Nath S, Ganguly A. Malaria detection through digital microscopic imaging using Deep Greedy Network with transfer learning. *J Med Imaging.* (2021) 8:054502. doi: 10.1117/1.JMI.8.5.0 54502

15. Zhang Z, Ong LS, Fang K, Matthew A, Dauwels J, Dao M, et al. Image classification of unlabeled malaria parasites in red blood cells. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2016) 3981–4. doi: 10.1109/EMBC.2016.7591599

16. Aliyu A, Prasad R, Fonkam M, A. framework for predicting malaria using naïve Bayes classifier. *Int J Telemed Appl.* (2018) 3:78–93. doi: 10.1504/IJTMCP.2018.093623

17. Pinkaew A, Limpiti T, Trirat A. Automated classification of malaria parasite species on thick blood film using support vector machine. In: *2015 8th Biomedical Engineering International Conference (BMEiCON).* (2015) 1–5. doi: 10.1109/BMEiCON.2015.7399524

18. Arowolo MO, Adebiyi M, Adebiyi A, Okesola O. PCA model for RNA–Seq malaria vector data classification using KNN and decision tree algorithm. In: *2020 international conference in mathematics, computer engineering and computer science (ICMCECS).* (2020) 1–8. doi: 10.1109/ICMCECS47690.2020.240881

19. Jimoh RG, Abisoye OA, Uthman MM. Ensemble feed–forward neural network and support vector machine for prediction of multiclass malaria infection. *Int J Inf Commun Technol.* (2022) 21:117–48. doi: 10.32890/jict2022.21.1.6

20. Marques G, Ferreras A. de la Torre–Diez I. An ensemble–based approach for automated medical diagnosis of malaria using EfficientNet. *Multimed Tools Appl.* (2022) 29:1–18. doi: 10.1007/s11042-022-12624-6

21. Shi B, Zhou X, Qin Z, Sun L, Xu Y. Corn ear quality recognition based on DCGAN data enhancement and transfer learning. In: *The 4th International Conference on Electronics, Communications and Control Engineering.* (2021). 62–68. doi: 10.1145/3462676.3462686

22. Rajaraman S, Antani SK, Poostchi M, Silamut K, Hossain MA, Maude RJ, et al. Pre–trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *Peer J.* (2018) 6:e4568. doi: 10.7717/peerj.4568

23. Amin J, Anjum MA, Gul N, Sharif M. A secure two–qubit quantum model for segmentation and classification of brain tumor using MRI images based on blockchain. *Neural Comput Appl.* (2022). 1–14. doi: 10.1007/s00521-022-07388-x

24. Amin J, Anjum MA, Malik M. Fused information of DeepLabv3+ and transfer learning model for semantic segmentation and rich features selection using equilibrium optimizer (EO) for classification of NPDR lesions. *Knowl Based Syst.* (2022)249:108881. doi: 10.1016/j.knosys.2022.108881

25. Amin J, Anjum MA, Sharif M, Jabeen S, Kadry S, Moreno Ger P, et al. new model for brain tumor detection using ensemble transfer learning and quantum variational classifier. *Comput Intell Neurosci.* (2022). doi: 10.1155/2022/3236305

26. Amin J, Anjum MA, Sharif M, Kadry S, Nadeem A, Ahmad SF. Liver tumor localization based on YOLOv3 and 3D–semantic segmentation using deep neural networks. *Diagnostics.* (2022) 12:823. doi: 10.3390/diagnostics12040823

27. Amin J, Sharif M, Fernandes SL, Wang SH, Saba T, Khan AR. Breast microscopic cancer segmentation and classification using unique 4-qubit-quantum model. *Microsc Res Tech.* (2022) 85:1926–36. doi: 10.1002/jemt.24054

28. Amin J, Sharif M, Anjum MA, Nam Y, Kadry S, Taniar D. Diagnosis of COVID−19 infection using three–dimensional semantic segmentation and classification of computed tomography images. *Comput Mater Contin.* (2021) 68:2451–67. doi: 10.32604/cmc.2021.014199

29. Amin J, Sharif M, Anjum MA, Siddiqa A, Kadry S, Nam Y, et al. *3d semantic deep learning networks for leukemia detection.* (2021). doi: 10.32604/cmc.2021.015249

30. Amin J, Anjum MA, Sharif A, Sharif MI. A modified classical–quantum model for diabetic foot ulcer classification. Intelligent Decision Technologies. *Intell Decis. Technol.* (2022) 16:1–6. doi: 10.3233/IDT-210017

31. Sadaf D, Amin J, Sharif M, Yasmin M. Detection of diabetic foot ulcer using machine/deep learning. *Med Image Anal.* (2000) 101–123. doi: 10.1201/9781003230540-7

32. Oyewola DO, Dada EG, Misra S, Damaševičius R. A novel data augmentation convolutional neural network for detecting malaria parasite in blood smear images. *Appl Artif Intell.* (2022) 1–22. doi: 10.1080/08839514.2022.2033473

33. Imran T, Sharif M, Tariq U, Zhang YD, Nam Y, Nam Y, et al. Malaria blood smear classification using deep learning and best features selection. *Comput Mater Contin.* (2021) 71:1–15. doi: 10.32604/cmc.2022.018946

34. Deelder W, Benavente ED, Phelan J, Manko E, Campino S, Palla L, et al. Using deep learning to identify recent positive selection in malaria parasite sequence data. *Malar J.* (2021) 20:1–9. doi: 10.1186/s12936-021-03788-x

35. Abubakar A, Ajuji M, Yahya IU. DeepFMD: computational analysis for malaria detection in blood–smear images using deep–learning features. *Appl Syst Innov.* (2021) 4:82. doi: 10.3390/asi4040082

36. Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy VV, et al. Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In*: 2017 IEEE EMBS international conference on biomedical & health informatics (BHI).* (2017). 101–4. doi: 10.1109/BHI.2017.7897215

37. Loh DR, Yong WX, Yapeter J, Subburaj K, Chandramohanadas R, A. deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using Mask R–CNN. *Comput Med Imaging Graph.* (2021) 88:101845. doi: 10.1016/j.compmedimag.2020.101845

38. Nakasi R, Mwebaze E, Zawedde A, Tusubira J, Akera B, Maiga G, et al. new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre–trained deep learning models SN. *Appl Sci.* (2020) 2:1–7. doi: 10.1007/s42452-020-3000-0

39. Zhang X, Zhou X, Lin M, Sun J. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2018). 6848–56. doi: 10.1109/CVPR.2018.00716

40. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning.* (2019). 6105–14.

41. Arthur D, Vassilvitskii S. k–means++: The advantages of careful seeding. In: *Proceedings of the 18th Annual Acm–Siam Symposiumon Discrete Algorithms.* (2006). 1027–35.

42. Zhao W, Zhang Z, Wang L. Manta ray foraging optimization: an effective bio–inspired optimizer for engineering applications. *Eng Appl Artif Intell.* (2020) 87:103300. doi: 10.1016/j.engappai.2019.103300

43. Got A, Zouache D, Moussaoui A, MOMRFO. multi–objective Manta ray foraging optimizer for handling engineering design problems. *Knowl Based Syst.* (2022) 237:107880. doi: 10.1016/j.knosys.2021.107880

44. El–Shorbagy MA, Omar HA, Fetouh T. Hybridization of manta–ray foraging optimization algorithm with pseudo parameter–based genetic algorithm for dealing optimization problems and unit commitment problem. *Mathematics.* (2022) 10:2179. doi: 10.3390/math10132179

45. Narayanan BN, Ali R, Hardie RC. Performance analysis of machine learning and deep learning architectures for malaria detection on cell images. In: *Applications of Machine Learnin.* (2019) 111390W. doi: 10.1117/12.2524681

46. Loddo, Fadda C, Di Ruberto C. An empirical evaluation of convolutional networks for malaria diagnosis. *J Imaging.* (2022) 8:66. doi: 10.3390/jimaging8030066

47. Alharbi AH, Lin M, Ashwini B, Jabarulla MY, Shah MA. Detection of peripheral malarial parasites in blood smears using deep learning models. *Comput Intell Neurosci.* (2022). doi: 10.1155/2022/3922763

48. Meng X, Ha Y, Tian J. Neighbor correlated graph convolutional network for multi–stage malaria parasite recognition. *Multimed Tools Appl.* (2022) 81:11393–414. doi: 10.1007/s11042-022-12098-6

49. Magotra V, Rohil MK. Malaria diagnosis using a lightweight deep convolutional neural network. *Int J Telemed Appl.* (2022) 2022:1–8. doi: 10.1155/2022/4176982

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact