



OPEN ACCESS

EDITED BY

Kathiravan Srinivasan,
Vellore Institute of Technology, India

REVIEWED BY

Mala Rajendran,
Mepco Schlenk Engineering College, India
Rakesh Chandra Joshi,
Dr. A.P.J. Abdul Kalam Technical University,
India

*CORRESPONDENCE

Mohammed Mansour
✉ mmansour755@gmail.com

RECEIVED 21 July 2023

ACCEPTED 03 October 2023

PUBLISHED 19 October 2023

CITATION

Mahmud S, Donmez TB, Mansour M, Kutlu M
and Freeman C (2023) Anemia detection
through non-invasive analysis of lip mucosa
images. *Front. Big Data* 6:1241899.
doi: 10.3389/fdata.2023.1241899

COPYRIGHT

© 2023 Mahmud, Donmez, Mansour, Kutlu and
Freeman. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Anemia detection through non-invasive analysis of lip mucosa images

Shekhar Mahmud¹, Turker Berk Donmez²,
Mohammed Mansour^{3*}, Mustafa Kutlu³ and Chris Freeman⁴

¹Department of Systems Engineering, Military Technological College, Muscat, Oman, ²Department of Biomedical Engineering, Sakarya University of Applied Sciences, Serdivan, Sakarya, Türkiye, ³Department of Mechatronics Engineering, Sakarya University of Applied Sciences, Serdivan, Sakarya, Türkiye, ⁴Electronics and Computer Sciences, University of Southampton, Southampton, United Kingdom

This paper aims to detect anemia using images of the lip mucosa, where the skin tissue is thin, and to confirm the feasibility of detecting anemia noninvasively and in the home environment using machine learning (ML). Data were collected from 138 patients, including 100 women and 38 men. Six ML algorithms: artificial neural network (ANN), decision tree (DT), k-nearest neighbors (KNN), logistic regression (LR), naive bayes (NB), and support vector machine (SVM) which are widely used in medical applications, were used to classify the collected data. Two different data types were obtained from participants' images (RGB red color values and HSV saturation values) as features, with age, sex, and hemoglobin levels utilized to perform classification. The ML algorithm was used to analyze and classify images of the lip mucosa quickly and accurately, potentially increasing the efficiency of anemia screening programs. The accuracy, precision, recall, and F-measure were evaluated to assess how well ML models performed in predicting anemia. The results showed that NB reported the highest accuracy (96%) among the other ML models used. DT, KNN and ANN reported an accuracies of (93%), while LR and SVM had an accuracy of (79%) and (75%) respectively. This research suggests that employing ML approaches to identify anemia will help classify the diagnosis, which will then help to create efficient preventive measures. Compared to blood tests, this noninvasive procedure is more practical and accessible to patients. Furthermore, ML algorithms may be created and trained to assess lip mucosa photos at a minimal cost, making it an affordable screening method in regions with a shortage of healthcare resources.

KEYWORDS

anemia, machine learning, classification, support vector machine (SVM), decision tree

Introduction

Anemia is characterized by a reduction of hemoglobin-containing red blood cells in the blood. The criteria for anemia, as determined by the World Health Organization (WHO), are defined as a hemoglobin level in the blood below 13 g/dL in men, 12 g/dL in women, and below 11 g/dL in pregnant women (Conrad, 2011). The most common causes of anemia are a decrease in red blood cell production or an increase in red blood cell destruction and loss, which is higher than normal (Brown, 1991; Aapro et al., 2002; Martinsson et al., 2014). Additionally, the production of malformed red blood cells in some hereditary blood diseases can also cause anemia. This results in a decrease in the average red blood cell count in the blood. The gold standard for detecting anemia is by taking intravenous blood

from a venous vein and analyzing this blood by hemogram (Prefumo et al., 2019; An et al., 2021; Milovanovic et al., 2022). However, invasive procedures, particularly in pregnant and pediatric groups, are painful and difficult to coordinate (Bashiri et al., 2003). The subject must also go to a clinic to receive the relevant procedure. In light of the coronavirus pandemic that began in 2019, performing these procedures in medicine and conventional follow-up mechanisms are no longer feasible. Non-invasive anemia follow-up may provide benefits in terms of patient comfort. Although anemia has many clinical side effects, it typically progresses with pallor of the skin. Therefore, it is more easily diagnosed in areas where the skin is thin, such as the conjunctiva, lips, tongue, and oral mucosa were Wang et al. employed a smartphone application that uses the camera and various illumination sources to noninvasively check blood hemoglobin content, Tamir et al. identified anemia by examining the pallor of the eye's front conjunctiva, by examining the color and metadata of smartphone photographs taken on the fingernail bed, Mannio et al. estimate hemoglobin levels and Selfie Anemia, a non-invasive hemoglobin estimate smartphone app that operates under regulated lighting conditions, was created by Noriega et al. (Wang et al., 2016; Tamir et al., 2017; Mannino et al., 2018; Rojas et al., 2019). Attempts have been made to detect anemia non-invasively, with a focus on the conjunctiva analysis using various methods (Dimauro et al., 2020; Rahman et al., 2020; Suner et al., 2021). However, it is extremely difficult for patients to detect anemia from a conjunctiva image captured with a simple phone camera to be used for communication with the IoT.

The diagnosis of anemia, in non-invasive methods using images of the conjunctiva, palm and nail bed, has its advantages. It also faces some limitations. These limitations include challenges related to accuracy due to variations in skin color, the impact of factors, the presence of medical conditions, and diversity among patients. Additionally, there are concerns regarding data quality, privacy issues costs involved, approval processes, and the need for clinical validation. It is important to remember that any diagnostic approach like this should complement judgment and other tests to ensure reliable results, in real life healthcare settings. To address these limitations effectively requires research, thorough validation procedures, and careful consideration during implementation. Using lip mucosa images for diagnosis offers unique benefits. This method is non-invasive meaning it doesn't cause any pain or inconvenience to patients. It's easily accessible and suitable for affordable screening programs making it patient friendly. By analyzing the lip mucosa, we can avoid the discomfort and possible infection risks associated with blood tests. This approach is particularly appealing to people who have a fear of needles or those who live in healthcare settings, as it promotes acceptance and participation. Further highlighting its potential as a useful and effective diagnostic approach is its adaptability for continuous monitoring and application in certain groups, such as pediatric patients. To guarantee that lip mucosa analysis is helpful in detecting illnesses like anemia, however, it needs to be extensively confirmed via scientific research and clinical trials.

In general, machine learning (ML) refers to computer techniques that automatically find approaches and parameters to arrive at the best solution to a problem, as opposed to

being pre-programmed by a person to propose a predetermined solution. This learning process is classified as a branch of Artificial Intelligence (AI), which simulates a component of human intellect and can be used for intelligent goals. A crucial component of the ML methodology is the technique used to carry out classification, regression, clustering, or prescriptive modeling. These techniques can be separated into supervised and unsupervised strategies. This research is the first to use lip mucous images for anemia detection in the literature. The aim of this paper is to detect anemia using images of the lip mucous, where the skin tissue is thin, and to confirm the feasibility of non-invasive anemia detection in the home environment. This will be achieved by developing classical ML algorithms trained using the collected patient data and invasive blood values. Six widely used ML algorithms in medical applications, namely artificial neural network (ANN), decision tree (DT), k-nearest neighbors (KNN), logistic regression (LR), naive bayes (NB), and support vector machine (SVM) (Kambeitz et al., 2015; Arbabshirani et al., 2017), were employed to classify the collected data. Subsequently, important statistical metrics were used to evaluate the performance of the algorithms. Participant's data of age between 16 and 76 was used and just confirmed anaemia cases as well were chosen. Using the lips mucous images, for diagnosis. Two extracted features from lip mucous images (RGB and HSV), age, sex and haemoglobin level, the experiment was done to predict anaemia using machine learning algorithms.

The accuracy of anemia diagnosis based on lip mucosa analysis can be significantly impacted by a variety of factors related to lip appearance and conditions, including lip pallor caused by vitamin B12 deficiency, dark spots, uneven skin tone, and abnormalities like scaly or thick lips, lip sores, or leukoderma. These elements may alter the hue or create distortions in the pictures of the lips, which may cause misunderstandings. The development of reliable algorithms that take into consideration these lip-related alterations and distinguish between true anemia-related color changes and those resulting from other lip disorders is essential for effective anemia identification. Furthermore, to reduce the possibility of misdiagnosis and guarantee accurate findings, clinical judgment and a comprehensive patient evaluation should be used in conjunction with computerized diagnosis. The examination of the lip mucosa for the diagnosis of anemia has obstacles from smoking and the use of lipstick, which requires identifying possible problems with changes in lip color, changes in texture and the variability of data due to these variables. The data set must be supplemented with photos of smokers and lipstick users, and machine learning algorithms must be modified to account for these particular traits. These steps are necessary to improve accuracy and sensitivity. Even in populations with different lip diseases caused by smoking or lipstick usage, the reliability and usefulness of the diagnostic procedure can be guaranteed by carefully weighing these difficulties and implementing mitigation measures.

The study is important due to its utilization of AI through ML, which is widely used for medical prediction and diagnosis (Kononenko, 2001; Jiang et al., 2017; Christodoulou et al., 2019; Alballa and Al-Turaiqi, 2021). Furthermore, this study offers a non-invasive method that is more convenient and accessible to patients than blood tests, especially in areas where these tests are not readily available. Early detection of anemia is crucial

as it can prevent serious health consequences such as fatigue, weakness, and decreased immune function. Moreover, developing and training ML algorithms to analyze lip mucous images is a cost-effective screening method in areas where healthcare resources are limited. Furthermore, the objective analysis of lip mucous images by ML algorithms reduces the risk of human error or bias in the diagnosis of anemia. Lastly, the ML algorithm could be used to quickly and accurately analyze large volumes of lip mucous images, which can potentially increase the efficiency of anemia screening programs once validated. In comparison to previous research and conventional diagnostic techniques, the approach outlined in our work shows potential for greatly improving the accuracy and sensitivity of anemia identification. The use of machine learning algorithms with features selection, the inclusion of different data kinds and demographic data, a sizable and diverse dataset, and thorough evaluation utilizing performance measures are all credited with this increase. Furthermore, lip mucosa analysis's non-invasiveness, affordability, accessibility, and capability for continuous monitoring should result in improved patient compliance and faster anemia identification. Though rigorous validation, comparisons with other diagnostic techniques, and clinical trials are essential stages in proving its superiority in actual clinical practice, they are not sufficient by themselves to support these assertions.

The remaining parts of the article are planned as follows: Section 2 provides a review of anemia detection using non-invasive methods. Section 3 explains the methodology. Section 4 outlines and presents the results. Section 5 presents the discussion. The last section concludes the article and suggests directions for future research.

Related work

In the field of biomedical data classification, numerous studies have been published in the last decade, providing a foundation for the current research. This section will thus focus on the details of previously applied non-invasive methods for anemia detection, including conjunctiva and fingertip analysis, as well as the ML models employed in these studies.

Suner et al. aimed to detect anemia in patients in the emergency room at a training and research hospital (Suner et al., 2021). In the first stage, images of both conjunctiva of 142 patients were taken with a smartphone. From each image, a region of interest was selected that targeted the palpebral conjunctiva. Image-based parameters were extracted and used in step-by-step regression analyses to develop a predictive model of predicted hemoglobin (HbC). In phase 2, a validation model was created with data from 202 new emergency room patients. The final model, based on all 344 patients, was tested for accuracy of anemia and transfusion thresholds.

Rojas et al. designed an application called selfienemia to estimate hemoglobin levels under controlled lighting conditions (Rojas et al., 2019). After taking a photo and processing it in the app, a colorimetric analysis is performed using a mathematical model from the cloud service. A special camera was used outside the application for better control of external conditions in this

prototype. Sixty-four tongue images and 64 conjunctival images were taken, and the results of the application were compared with traditional large blood count (CBC), which is considered the gold standard test for diagnosing anemia. In the analysis of tongue images, the results were 91.89% sensitive and 85.18% specific, and in the analysis of palpebral conjunctiva, the results were 91.89% sensitive and 70.34% specific.

Sevani et al. aimed to support the process of detecting anemia using conjunctival pallor with a smartphone camera (Sevani et al., 2018). They applied the K-Means clustering method to analyze the pixels of conjunctival images represented by digital characters in RGB formats. They compared the test results obtained from this application with laboratory results, demonstrating that the method provided an accuracy of 90%. Mannino et al. estimated hemoglobin levels by analyzing the color and metadata of nail bed photos taken with a smartphone (Mannino et al., 2018). In their study of 100 people, they were able to test anemia with a sensitivity of 97%.

Hasan et al. utilized video images of fingertip captured with a smartphone camera under a flashlight and trained ANN to predict HGB levels non-invasively (Hasan et al., 2018). Red, green, and blue pixel densities were calculated in 100 blocks of fields in each frame in 10 seconds (300 frames) of recorded video by 75 adults, and this method was applied to all 300 frames. ANN was then used to develop a derived model for predicting hemoglobin levels. They found that there was a correlation of 0.93 between the model and gold standard hemoglobin levels in their sample of patients aged 20 to 56 years.

Tamir et al. developed an android application to capture a photo of the anterior eye conjunctiva with a smartphone camera in suitable lighting conditions with appropriate resolution (Tamir et al., 2017). These images were then processed to obtain spectra of the conjunctival color and RGB components, which were compared to a threshold to determine anemia. In their study of 19 subjects whose hemoglobin levels were known, they compared the values of 15 people with the laboratory results and correctly identified them at a rate of 78.9%.

Li et al. proposed a novel method with dynamic spectrum, using a spectrograph with a computer to scan the transmission spectrum of fingertip (Li et al., 2017). An average prediction correlation coefficient (R) of 0.8399 is achieved in their experiment. Using a light source, Wang et al. performed chromatic analysis of images taken of the patient's finger, using an application called HemaApp (Wang et al., 2016). When they analytically evaluated 31 patients between the ages of 6 and 77, they obtained a correlation of 0.82 with the blood test. HemaApp showed a sensitivity of 85.7% and a specificity of 76.5% in anemia screening.

Most previous methods for detecting anemia have utilized data such as conjunctival images, human nails, and fingertip. ML has been used in some of these methods to develop both invasive and noninvasive approaches for detecting anemia. However, there is still a need to improve the performance of these methods by incorporating new data and employing more straightforward approaches. This study is particularly important as it is the first to utilize lip mucous images for predicting anemia using ML. The study's use of ML techniques, which has been widely employed in the medical field for prediction and diagnosis (Kononenko, 2001; Jiang et al., 2017; Christodoulou et al., 2019; Alballa and Al-Turaiki, 2021), is what provides it with its significance.

TABLE 1 Participants demographic values.

| Variable | | Frequency | | Percentage of anemia patients |
|----------|--------|-----------|--------|-------------------------------|
| | | Healthy | Anemia | |
| Gender | Female | 77 | 23 | 72.46 |
| | Male | 32 | 6 | 27.54 |
| Age | 18–30 | 53 | | 38.4 |
| | 30–50 | 36 | | 26 |
| | 50 | 49 | | 35.6 |

Methodology

The classification problem aims to detect anemia using a dataset of collected lip images. The study began by building a dataset that includes two types of lip images: healthy and anemic. Data preparation was then performed, followed by the using of ML models for classification, which were evaluated to determine the best model.

Participants

Following the Sakarya University Ethical Approval (E-71522473-05.01.04-74571-458), participants were recruited between November and December 2021. The study group consisted of participants residing in the Pamukova District of Sakarya Province. Data were collected from 138 participants, including 100 women and 38 men by a team of experienced medical doctors. Smokers images have not been included in this study. According to the WHO criteria, 23 women and 6 men were diagnosed with anemia (Organization, 2011). The range of hemoglobin level 129 level of healthy individuals was 121 to 167 grams per liter (g/L). The study group was selected using the convenience sampling method (Patton, 2005). The range of haemoglobin level for male was 133–167 g/L with ages 16–68 years old between and the range of haemoglobin level for female was 121–158 g/L with ages between 11 and 76 years old. The range of anemia patient was 80–130 grams per liter (g/L). Demographic variables such as gender and age ranges are presented in Table 1.

Experimental Setup

The experimental setup shown in Figure 1 was designed to measure the facial features of the participants, it is designed to display only the lip area with the help of an adjustable frame. A Canon camera EOS 2000D was used for collecting the lip mucous images. The camera's flash was used for lighting. Examples of collected lip mucosa images for healthy and non healthy one are shown in Figures 2, 3 respectively.



FIGURE 1 System setup: (1) camera, (2) frame.

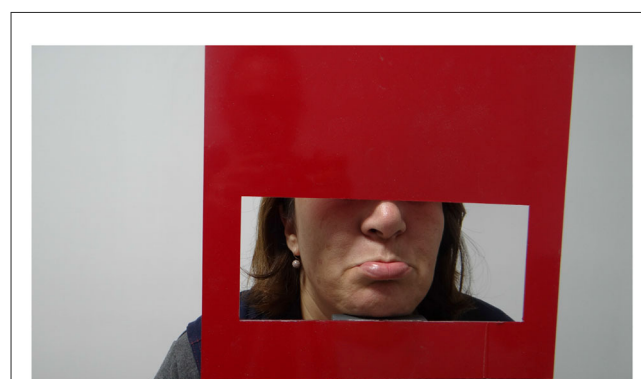


FIGURE 2 Healthy person representation.

Software Setup

A custom Python application was developed to process and analyze the data collected from the participants. Firstly, the lip contour of each participant was obtained using corner detection, thresholding, and framing (see Figures 4, 5 for lips pallor examples). Next, the digital image within the frame was converted to RGB and HSV formats. Finally, classical ML algorithms were employed to perform the classification task. These steps will be discussed in detail in this section.

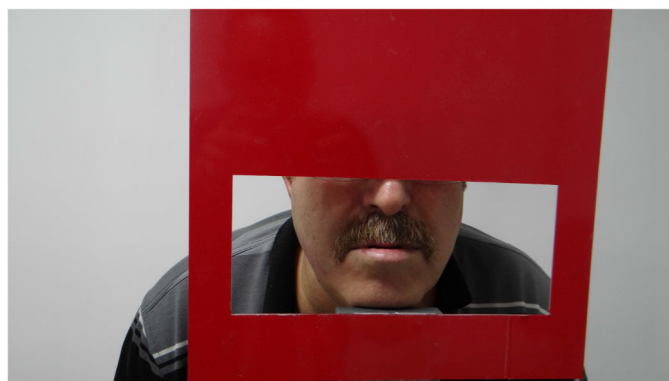


FIGURE 3
Anemia patient.



FIGURE 4
Healthy lips pallor with Hb of 137 grams per liter (g/L).



FIGURE 5
Anemia patient lips pallor with Hb of 116 grams per liter (g/L).

features of the image that were sent to the classification algorithm, thereby increasing the classification success.

Data Normalization

In the absence of normalization adjustments, the larger scale variable will totally dominate ML algorithms' attempts to predict trends. Numerous ML methods demand statistical rescaling of their input variables to prevent sacrificing numerical stability (Trentin, 2015). In order to improve the model's fit to the supplied data, minimum maximum normalization techniques are used (Senvar and Sennaroglu, 2016). According to it, all attributes are equally important in terms of size (Singh and Singh, 2020). The unnormalized 145 data are linearly adjusted by the normalization technique to a defined lower and upper bound (Han et al., 2011). Typically, the dataset is rescaled to lie between 0 and 1 or -1 and 1. In this study, the minimum maximum normalization technique with a [0,1] scale was examined. The five inputs were normalized to the value between [0 and 1]. Equation 1 demonstrates the process used to transform raw data into normalized data, where X stands for the real data, X_{min} represents the lowest value found in the dataset of all X values, and X_{max} reflects the highest X-value found in the dataset. Thus, $X_{normalized}$ displays the normalized X value and ranges from 0 to 1.

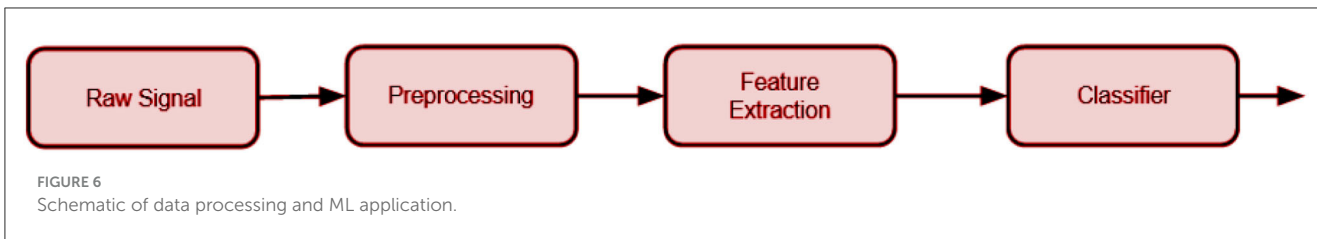
$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Data processing and preparing

Feature extraction and classification of the images and hemoglobin data obtained using the setup are explained in Figure 6. A high-resolution image (4896x2752) of the participants was captured using a camera. The images were collected on the same day as the blood sampling. Thus, the data duration and the number of samples taken from each subject are close but not the same. Two different data types were obtained from participants' images as features: RGB (Red, Green, Blue) red color values and HSV saturation values, along with age, sex, and hemoglobin levels, which were used for classification. Feature extraction identified different

Machine Learning Algorithms

DT is a decision-making method that has a tree structure and Random Forest (RF) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. It improves predictive accuracy with average and reduces over fitting (Gnanapriya et al., 2010; Schonlau and Zou, 2020; Mansour et al., 2023). In this study, the DT were constructed with a maximum depth of 1 for estimating the three joints moment. SVM is a popular supervised ML algorithm used for

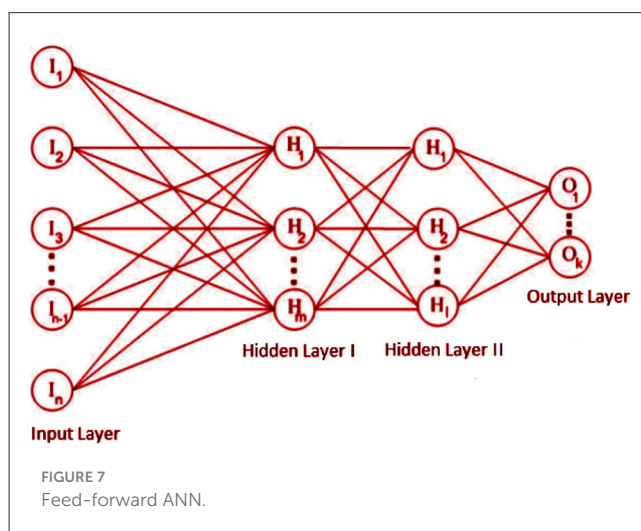


both classification and regression tasks. It is particularly effective in solving complex problems with high-dimensional data. SVM aims to find an optimal hyperplane that separates the data points of different classes, maximizing the margin between the classes.

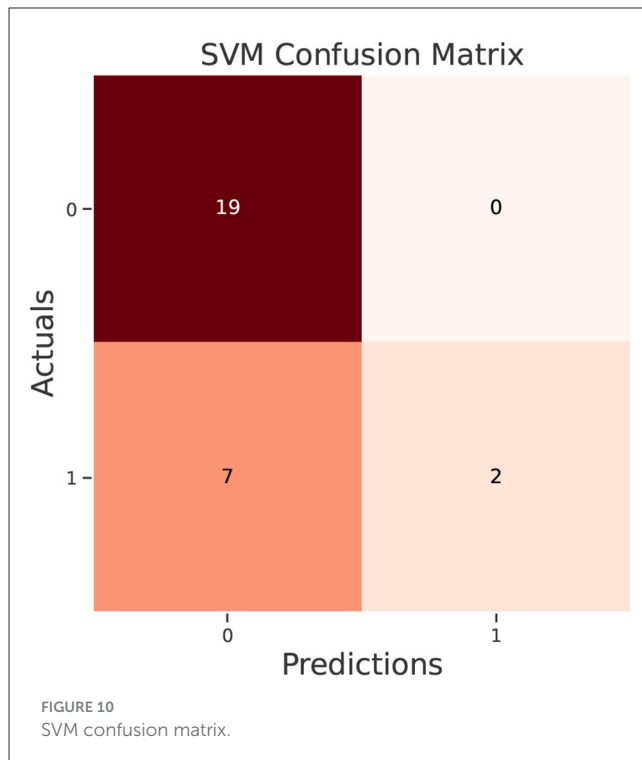
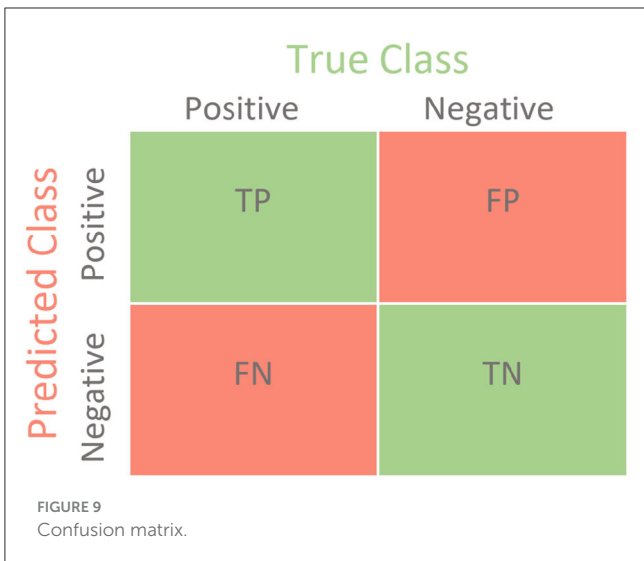
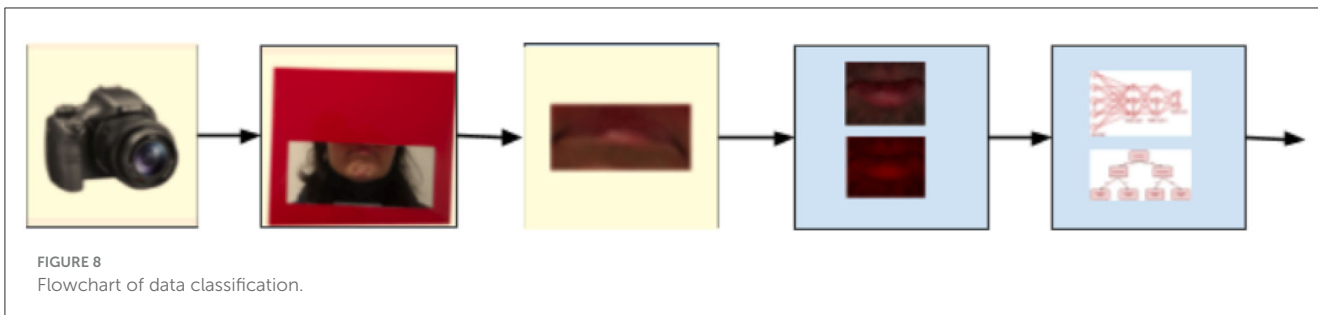
For supervised classification and regression applications, the KNN algorithm is a non-parametric, instance-based learning technique. KNN is adaptable and useful for a variety of tasks because, unlike many other machine learning algorithms, it does not make firm assumptions about the distribution of the underlying data. The foundation of KNN is the idea that data points with comparable properties are more likely to fall into one class or display similar goal values. For supervised classification problems, the NB algorithm is a probabilistic ML technique. Although NB is straightforward and makes the "naive" premise of feature independence, it has been successful in a number of fields. In order to shed light on the NB algorithm's inner workings and demonstrate its adaptability for practical applications. A popular linear model in ML for binary and multi-class classification applications is LR. Despite being straightforward, it acts as a crucial building element for a number of sophisticated procedures.

Compared with previous ML techniques, ANN algorithm is widely applied for biomedical data classification. As lack of large data sets in the healthcare and interconnected complex relationships between the individual biological components have encouraged the scientific research community to integrate ANN models. It adapt within nonlinear boundaries and is efficient to provide better classification. ANN has the ability to learn from continuous data and update the model, which is not present in other ML algorithms such as the decision tree. An example of a back feed-forward neural network is shown in Figure 7. It is a simple classification algorithm where the information is routed from input to output. The backpropagation algorithm was a widely utilized technique for training multiple-layer perceptrons. The multilayered layer perceptron's weight values are adjusted by approaching the intricate intersection of input and output data. The structure of the model in this study was obtained when the input, output, and one hidden layer are 10-5-1 (number of neurons) Input, hidden, and output layers, respectively, used Relu, Relu, and sigmoid activation functions. The ANN model were applied with various hidden layer neurons until it reaches its best results, it was trained from 5 to 10 neurons. The training cycle was 100 and Adam optimizer was used.

The performance of ML algorithms is critical to their usefulness and effectiveness in solving real-world problems (Hossin and Sulaiman, 2015; Santafe et al., 2015; Mansour et al., 2023). High-performing algorithms can make more accurate predictions, process data more efficiently, scale to handle large datasets, and be more interpretable, leading to better decision-making and improved outcomes. Classification models are widely used in ML



to predict outcomes based on a set of characteristics. To evaluate the performance of such models, several commonly used metrics are available. The choice of metric depends on the nature of the problem, class balance, and desired outcome of the model (Prati et al., 2011; Santafe et al., 2015; Tharwat, 2021). Accuracy is a simple metric that measures the proportion of correct predictions out of all predictions made by the model (Folorunso et al., 2022). However, it may not be the best choice when there is class imbalance in the data. Precision measures the proportion of true positives out of all positive predictions made by the model and is useful when the goal is to minimize false positives (Juba and Le, 2019; Miao and Zhu, 2022). Recall, on the other hand, measures the proportion of true positives out of all the actual positive examples in the data, and is useful when the goal is to minimize false negatives (Ali et al., 2022; Miao and Zhu, 2022). F1 score is the harmonic mean of precision and recall and is useful to balance the importance of both (Hossin and Sulaiman, 2015). AUC measures the ability of the model to distinguish between positive and negative examples and is useful when identifying the best threshold to separate positive and negative examples (Tharwat, 2021). Finally, the confusion matrix provides a more detailed view of the performance of the model than any single metric by displaying the number of true positives, true negatives, false positives, and false negatives for the given model (Haghighi et al., 2018; Liang, 2022). Overall, the performance evaluation of a ML model is essential to determine its effectiveness and identify areas for improvement. A combination of metrics should be used to evaluate the model, and the context of the problem should be considered when choosing a metric.



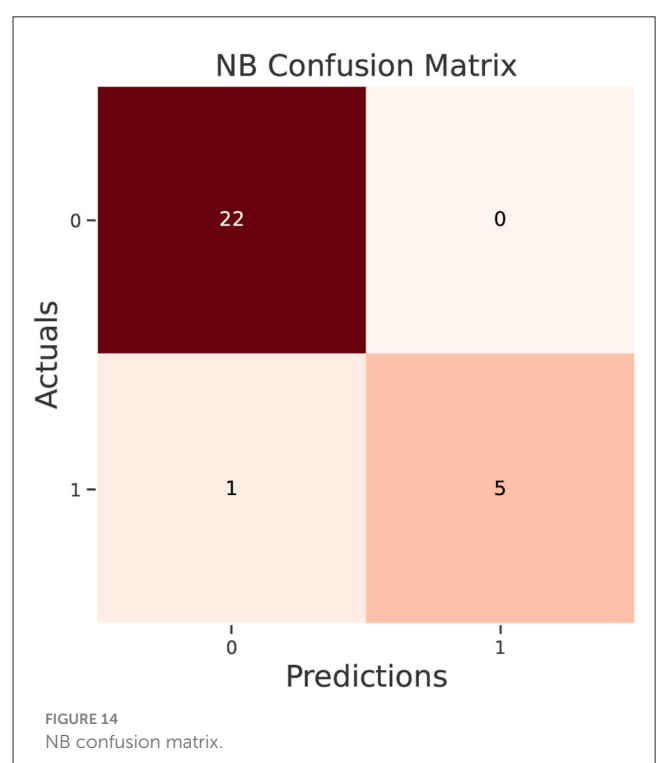
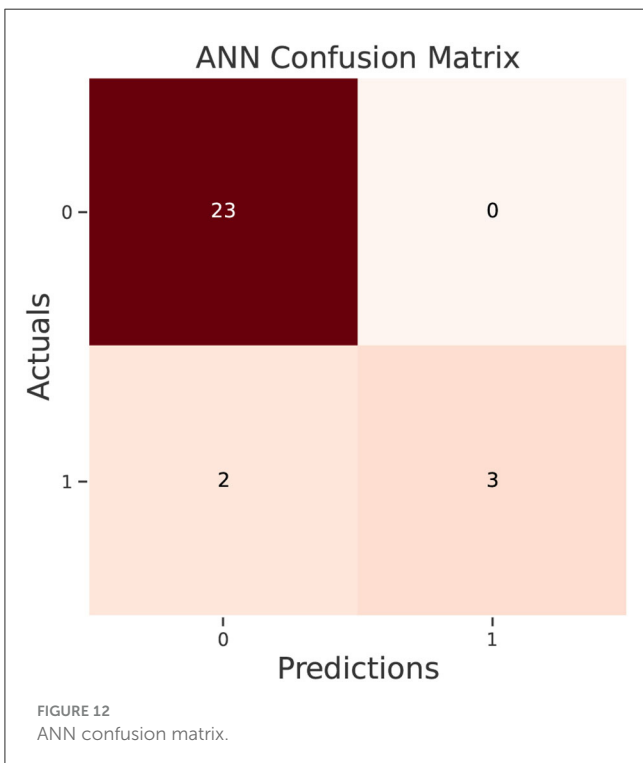
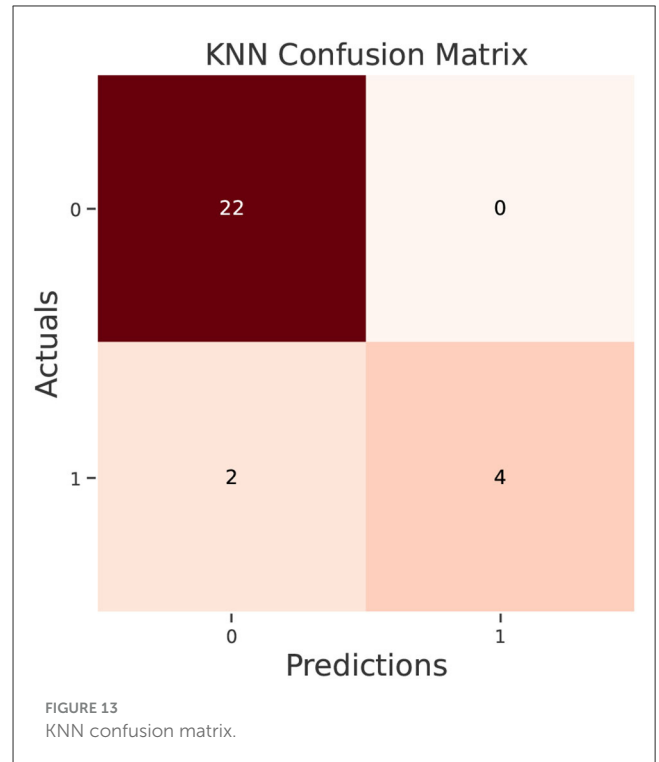
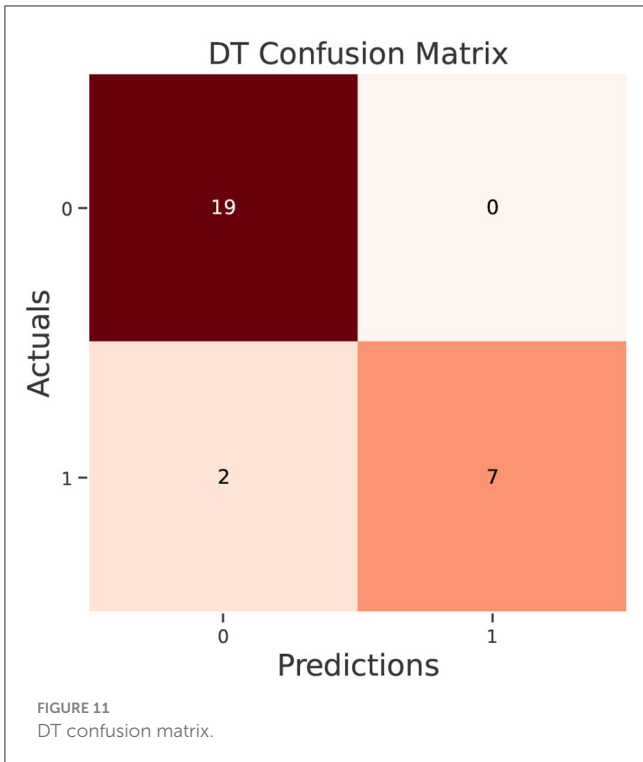
Results and discussion

The experiments were performed using the Python programming language and its Keras and Tensorflow libraries. This research classifies anemia by applying the input characteristics produced with the extraction of RGB (Red value), HSV (Saturation), age, gender, hemoglobin levels. In the classification process, the split data function was applied to separate the data into train, val and test data where 60% (82) of the dataset was randomly assigned as train data, 20% (28) were randomly assigned as validate and 20% (28) as test data. The flow diagram describing the processing and classification is shown in Figure 8, where the process starts with the collection of patient data and images, the cropping of the image, feature extraction and the preparation of the data, and finally classification and prediction of anemia cases and evaluation of ML models. In the classification process, six algorithms were applied; KNN, NB, LR, SVM, DT, and ANN algorithms, which are frequently applied in the anemia detection literature, were used. To evaluate the classification algorithms used in this study; different methods were used. These methods are accuracy, precision, recall, and F1 score.

The test data confusion matrix prediction evaluation for SVM, DT, ANN, KNN, NB and LR are shown in Figures 10–15. Four main points to consider here, True positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). These points clarify the success of the algorithm where TP and TN are the correct

predictions in the both positive and negative classes, and FP and FN are the false predictions respectively (Figure 9). For SVM algorithm (Figure 10), the TP was 19 and the TN was 2 from a total of 19 and 9 respectively. Figure 11 presents the confusion matrix for DT algorithm. For DT, the TP was 19 and the TN was 7 from a total of 19 and 9 respectively. Figure 12 presents the confusion matrix for ANN algorithm. For ANN, the TP was 23 and the TN was 3 from a total of 23 and 5 respectively. Figure 13 presents the confusion matrix for KNN, the TP was 22 and the TN was 4 from a total of 22 and 6 respectively. Figure 14 presents the confusion matrix for NB, the TP was 22 and the TN was 5 from a total of 22 and 6 respectively. Finally, Figure 15 presents the confusion matrix for LR, the TP was 22 and the TN was 0 from a total of 22 and 6 respectively.

The success rates of all algorithms are shown in Tables 2–5 where Tables 2, 3 are the evaluation of both train and validation data. Tables 4, 5 present the macro and weighted average of the test data for both the positive and negative classes. In these tables and based on confusion matrices, four main parameters; accuracy (Equation 2), AUC (area under the ROC curve), precision (Equation 3), recall (Equation 4) and F1 score (Equation 5) which



evaluate the six algorithms applied to classify anemia applying the input characteristic matrix of RGB (red value), HSV (saturation), age, sex, hemoglobin levels.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

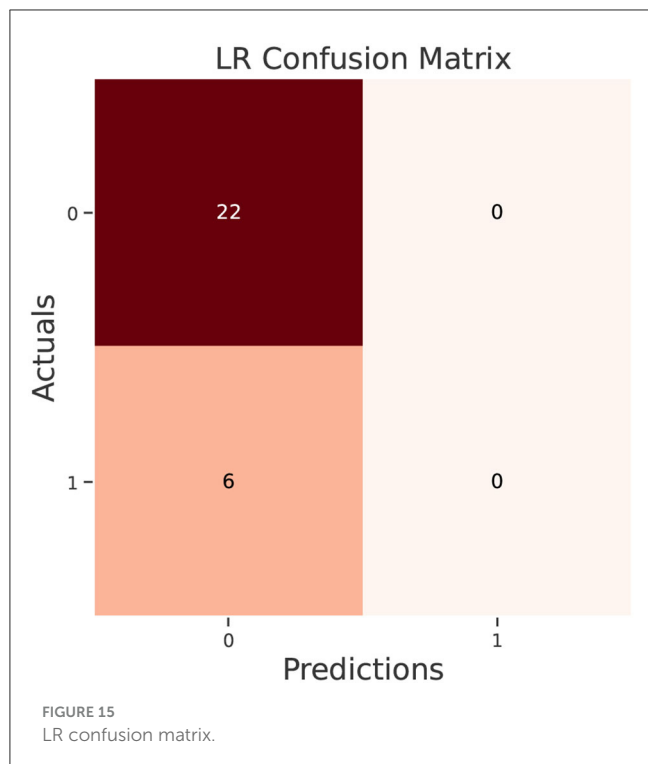


TABLE 2 Train data; accuracy, precision, recall and F1 score.

| Algorithm | Accuracy (%) | Precision | Recall | F1 score |
|-----------|--------------|-----------|--------|----------|
| SVM | 84 | 42 | 50 | 46 |
| LR | 93 | 96 | 82 | 87 |
| DT | 100 | 100 | 100 | 100 |
| ANN | 79 | 40 | 50 | 44 |
| KNN | 90 | 91 | 79 | 83 |
| NB | 93 | 96 | 82 | 87 |

TABLE 3 Val data; accuracy, precision, recall and F1 score.

| Algorithm | Accuracy (%) | Precision | Recall | F1 score |
|-----------|--------------|-----------|--------|----------|
| SVM | 75 | 38 | 50 | 43 |
| LR | 96 | 98 | 92 | 94 |
| DT | 93 | 86 | 96 | 89 |
| ANN | 75 | 38 | 50 | 43 |
| KNN | 89 | 86 | 81 | 83 |
| NB | 96 | 98 | 92 | 94 |

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

When comparing the results obtained from the anemia classification performed using ML techniques, it was found that the disease was diagnosed with an accuracy success rate of 96% using NB, 93% using DT KNN and ANN, LM classifies the data with an accuracy success rate of 79% and SVM classifies the data with an accuracy success rate of 75%. The AUC were 50, 83, 86, 91, 95, and

96 for LR, KNN, SVM, NB, DT, and ANN respectively. Although this value seems high, higher percentages and more data-trained models are required to use it as a medical diagnosis.

Discussion

In this study, extracted data from images of lip mucous were used to train a ML models to identify anemia. The results of this study are in agreement with those of previous investigations that used ML models to predict anemia using both invasive and non-invasive techniques. Accuracy, precision, recall, and F-score were used to assess how well ML models performed in predicting anemia. The models examined demonstrated high and considerable accuracy. For predicting anemia, NB reported the highest accuracy, the highest precision and F1 score and SVM reported the lowest scores.

Limited methods have been found in the literature to detect anemia using ML. For example, the K-means clustering technique was used to carry out conjunctival pallor image-based anemia detection, which demonstrated an accuracy of 90% when comparing the test results acquired from this application with laboratory results (Sevani et al., 2018). ANN were used to assess images of fingertip captured with a smartphone camera to predict HGB levels non-invasively. As a result, a correlation of 0.93 was observed between the model and gold standard hemoglobin levels (Hasan et al., 2018). Convolutional neural networks was techniques was used to carry out conjunctival image-based anemia detection with accuracy of 94% (Magdalena et al., 2022). YOLO v5 was used to detect anemia using conjunctiva image collection with sensitivity of 71% and a specificity of 89% (Rivero-Palacio et al., 2022). AlexNet was used to calculate total hemoglobin concentration by developing frequency-domain multidistance approach, based on a non-contact oximeter, provided data on total hemoglobin with accuracy of 87.50% (Moral and Bal, 2020). Compared to these methods that used ML in general for anemia detection, the current study presents a non-invasive method that uses ML models to detect anemia with higher accuracy, reaching 99% using DT. This new method is simple and can be developed for the detection of real-time anemia. Several methods were found in the literature to detect anemia.

Anemia is a prevalent condition that affects millions of people worldwide, but it often goes undiagnosed until it becomes severe. Early detection of anemia is crucial for effective treatment, which is why the use of ML to detect anemia through lip mucosa image classification could be significant. This method provides a non-invasive and cost-effective alternative to traditional anemia screening methods such as blood tests, which can be invasive and costly, particularly in resource-limited settings. The automation of diagnosis through ML algorithms can reduce the need for expert human intervention and speed up the process of diagnosis and treatment. The potential for automation also makes this method scalable and easily accessible, allowing for widespread implementation of anemia detection tools. Our research suggests that employing ML approaches to detect anemia will aid in classifying the diagnosis, which will then help in the creation of efficient preventive measures. As a result, this research evaluates the predictive capability of several ML algorithms in addition

TABLE 4 Test data macro averages; accuracy, precision, recall and F1 score.

| Algorithm | Accuracy (%) | AUC | Precision | Recall | F1 score |
|-----------|--------------|-----|-----------|--------|----------|
| SVM | 75 | 86 | 87 | 61 | 60 |
| LR | 79 | 50 | 93 | 50 | 44 |
| DT | 93 | 95 | 89 | 95 | 91 |
| ANN | 93 | 96 | 96 | 80 | 85 |
| KNN | 93 | 83 | 96 | 83 | 88 |
| NB | 96 | 91 | 98 | 92 | 94 |

TABLE 5 Test data weighted averages; accuracy, precision, recall and F1 score.

| Algorithm | Accuracy (%) | AUC | Precision | Recall | F1 score |
|-----------|--------------|-----|-----------|--------|----------|
| SVM | 75 | 86 | 87 | 61 | 60 |
| LR | 79 | 50 | 62 | 79 | 69 |
| DT | 93 | 95 | 94 | 93 | 93 |
| ANN | 93 | 96 | 93 | 93 | 92 |
| KNN | 93 | 83 | 93 | 93 | 92 |
| NB | 96 | 91 | 97 | 96 | 96 |

to addressing the integration of cutting-edge technology for the prediction and diagnosis of low hemoglobin levels.

When compared to other human body parts, lip pallor, which is defined by the paleness or loss of natural color in the lips, provides significant advantages as a non-invasive diagnostic site. First of all, there is no need for specialist equipment or invasive procedures to examine the lips because they are quite obvious and accessible. Due to its accessibility, lip pallor is a sensible option for diagnostic techniques, enabling quick and little disruptive patient examinations. A rich vascular network with countless small blood vessels close to the surface is also present on the lips. Due to the vascular richness, variations in blood flow and oxygenation, which frequently show up as changes in lip color, may be quickly identified. The lips are perfect for identifying blood-related diseases like anemia because of the strong relationship between lip color and the circulatory system. This provides real-time information about a patient's health.

Lip pallor analysis is non-invasive, which increases patient comfort and compliance. The procedure is harmless and acceptable for people of all ages, whether it involves straightforward eye exams or more sophisticated imaging techniques. This technique also fits well with ethical standards and cultural norms because it is typically socially acceptable in all cultures to examine one's lips. Such acceptance may increase a patient's willingness to participate in lip-based diagnostic tests. The ability to capture high-resolution photographs of the lips thanks to advancements in imaging technology also makes it possible to analyze color variations and texture changes precisely. This degree of specificity is essential for identifying minute symptoms of diseases like anemia and guarantees that lip pallor analysis will always be a practical and affordable screening technique in medical settings, making it a crucial tool in the field of non-invasive diagnostics.

The contribution of this study to medical research is also significant, as the use of ML in medical research is still in its

early stages. This study could provide a foundation for further research and development of ML based tools for anemia detection and diagnosis, potentially leading to even more accurate and effective diagnostic tools. The use of ML for anemia detection using lip mucosa image classification could have significant implications for healthcare, particularly in resource-limited settings where traditional screening methods may not be readily available. The potential for early detection, non-invasive and cost-effective screening, automation of diagnosis, contribution to medical research, and further development of diagnostic tools make this study a promising avenue for improving healthcare outcomes.

Conclusions

Images of the lip mucosa, which have thin skin tissue, were used in this study to identify anemia. Data from 138 patients, including 100 women and 38 men, were collected. Rgb red color values and hsv saturation values were obtained from participant images and used as features, along with age, sex, and hemoglobin levels, to perform classification. The efficacy of ML models in predicting anemia was tested using accuracy, precision, recall, and F score. The findings indicated that among the ML algorithms utilized, NB achieved the highest accuracy at 96%, while SVM received accuracy ratings of 75%. This research suggests that using ML to recognize anemia will aid in classifying the diagnosis, which would subsequently facilitate the development of effective preventive measures.

This work has established a link between anemia detection and lip mucous images, where the conjunctiva is used in general practices. This newly found method is easy to use, and it was demonstrated for the first time that it can be classified on the lip mucous. Although the accuracy of 93% appears high, more data-trained models and larger percentages are needed to use it as a

diagnostic in medicine. In the future, augmented data will be used for an online classification along with deep neural networks and a mobile application.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Sakarya University Ethical Approval (E-71522473-05.01.04-74571-458). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

TD collecting the lip mucosa images and literature review. MM preparing the software and manuscript writing. MK preparing the images and feature extraction. CF writing the manuscript. SM

References

- Aapro, M. S., Cella, D., and Zagari, M. (2002). Age, anemia, and fatigue. *Semin. Oncol.* 29, 55–59. doi: 10.1016/S0093-7754(02)70175-9
- Albala, N., and Al-Turaiki, I. (2021). Machine learning approaches in covid-19 diagnosis, mortality, and severity risk prediction: a review. *Inform. Med.* 24, 100564. doi: 10.1016/j.imu.2021.100564
- Ali, I., Mughal, N., Khand, Z. H., Ahmed, J., and Mujtaba, G. (2022). Resume classification system using natural language processing and machine learning techniques. *Mehran Univ. Res. J. Eng. Technol.* 41, 65–79. doi: 10.22581/muet1982.2201.07
- An, R., Huang, Y., Man, Y., Valentine, R. W., Kucukal, E., Goreke, U., et al. (2021). Emerging point-of-care technologies for anemia detection. *Lab Chip* 21, 1843–1865. doi: 10.1039/D0LC01235A
- Arbabshirani, M. R., Plis, S., Sui, J., and Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *Neuroimage* 145, 137–165. doi: 10.1016/j.neuroimage.2016.02.079
- Bashiri, A., Burstein, E., Sheiner, E., and Mazor, M. (2003). Anemia during pregnancy and treatment with intravenous iron: review of the literature. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 110, 2–7. doi: 10.1016/S0301-2115(03)00113-1
- Brown, R. G. (1991). Determining the cause of anemia: general approach, with emphasis on microcytic hypochromic anemias. *Postgrad. Med.* 89, 161–170. doi: 10.1080/00325481.1991.11700925
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., and Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* 110, 12–22. doi: 10.1016/j.jclinepi.2019.02.004
- Conrad, M. E. (2011). *Anemia*. Boston, MA: Butterworths.
- Dimauro, G., Caivano, D., Di Pilato, P., Dipalma, A., and Camporeale, M. G. (2020). A systematic mapping study on research in anemia assessment with non-invasive devices. *Appl. Sci.* 10, 4804. doi: 10.3390/app10144804
- Folorunso, S. O., Awotunde, J. B., Adeniyi, E. A., Abiodun, K. M., and Ayo, F. E. (2022). “Heart disease classification using machine learning models,” in *Informatics and Intelligent Applications: First International Conference, ICIIA 2021*. Ota, Nigeria: Springer, 35–49.
- Gnanapriya, S., Suganya, R., Devi, G. S., and Kumar, M. S. (2010). Data mining concepts and techniques. *Data Knowl. Eng.* 2, 256–263.
- Haghighi, S., Jasemi, M., Hessabi, S., and Zolanvari, A. (2018). Pycm: multiclass confusion matrix library in python. *J. Open Sour. Software* 3, 729. doi: 10.21105/joss.00729
- Han, J., Pei, J., and Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Amsterdam: Elsevier.
- Hasan, M. K., Haque, M. M., Adib, R., Tumpa, J. F., Begum, A., Love, R. R., et al. (2018). “Smarthelp: Smartphone-based hemoglobin level prediction using an artificial neural network,” in *AMIA Annual Symposium Proceedings*. Washington, D.C.: American Medical Informatics Association, 535.
- Hossin, M., and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process.* 5, 1. doi: 10.5121/ijdkp.2015.5201
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke Vascular Neurol.* 2, 4. doi: 10.1136/svn-2017-000101
- Juba, B., and Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. *Proc. AAAI Conference Artif. Intell.* 33, 4039–4048. doi: 10.1609/aaai.v33i01.33014039
- Kambeitz, J., Kambeitz-Ilankovic, L., Leucht, S., Wood, S., Davatzikos, C., Malchow, B., et al. (2015). Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology* 40, 1742–1751. doi: 10.1038/npp.2015.22
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artif. Intell. Med.* 23, 89–109. doi: 10.1016/S0933-3657(01)00077-X
- Li, G., Xu, S., Zhou, M., Zhang, Q., and Lin, L. (2017). Noninvasive hemoglobin measurement based on optimizing dynamic spectrum method. *Spectrosc. Lett.* 50, 164–170. doi: 10.1080/00387010.2017.1302481
- Liang, J. (2022). *Confusion Matrix: Machine Learning*. POGIL Activity Clearinghouse. Available online at: <https://pac.pogil.org/index.php/pac/article/view/304>
- Magdalena, R., Saidah, S., Da'wan, I., Ubaidah, S., Fuadah, Y. N., Herman, N., et al. (2022). Convolutional neural network for anemia detection based on conjunctiva palpebral images. *Jurnal Teknik Informatika (Jutif)* 3, 349–354. doi: 10.20884/1.jutif.2022.3.2.197

writing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Mannino, R. G., Myers, D. R., Tyburski, E. A., Caruso, C., Boudreaux, J., Leong, T., et al. (2018). Smartphone app for non-invasive detection of anemia using only patient-sourced photos. *Nat. Commun.* 9, 1–10. doi: 10.1038/s41467-018-07262-2
- Mansour, M., Demirsoy, M. S., and Kutlu, M. C. (2023). Kidney segmentations using cnn models. *J. Smart Syst. Res.* 4, 1–13. doi: 10.58769/joinsr.1175622
- Mansour, M., Serbest, K., Kutlu, M., and Cilli, M. (2023). “Estimation of lower limb joint moments based on the inverse dynamics approach: a comparison of machine learning algorithms for rapid estimation,” in *Medical & Biological Engineering & Computing*, 1–24.
- Martinson, A., Andersson, C., Andell, P., Koul, S., Engström, G., and Smith, J. G. (2014). Anemia in the general population: prevalence, clinical correlates and prognostic impact. *Eur. J. Epidemiol.* 29:489–498. doi: 10.1007/s10654-014-9929-9
- Miao, J., and Zhu, W. (2022). Precision-recall curve (prc) classification trees. *Evol. Intell.* 15, 1545–1569. doi: 10.1007/s12065-021-00565-2
- Milovanovic, T., Dragasevic, S., Nikolic, A. N., Markovic, A. P., Lalosevic, M. S., Popovic, D. D., et al. (2022). Anemia as a problem: Gp approach. *Digest. Dis.* 40, 370–375. doi: 10.1159/000517579
- Moral, O. T., and Bal, U. (2020). “Non-contact total hemoglobin estimation using a deep learning model,” in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. Istanbul: IEEE, 1–6.
- Patton, M. Q. (2005). *Qualitative Research*. Thousand Oaks, CA: Sage.
- Prati, R. C., Batista, G. E., and Monard, M. C. (2011). A survey on graphical methods for classification predictive performance evaluation. *IEEE Trans. Knowl. Data Eng.* 23, 1601–1618. doi: 10.1109/TKDE.2011.59
- Prefumo, F., Fichera, A., Fratelli, N., and Sartori, E. (2019). Fetal anemia: diagnosis and management. *Best Pract. Res. Clin. Obstet. Gynaecol.* 58, 2–14. doi: 10.1016/j.bpobgyn.2019.01.001
- Rahman, M., Nakib, N., Sadik, S. A., Biswas, A., and Tazul, R. B. (2020). *Point of Care Detection of Anemia in Non-Invasive Manner by Using Image Processing and Convolutional Neural Network With Mobile Devices* (PhD thesis). Dhaka: Brac University.
- Rivero-Palacio, M., Alfonso-Morales, W., and Caicedo-Bravo, E. (2022). Anemia detection using a full embedded mobile application with yolo algorithm. *Commun. Comput. Inf. Sci.* 1471, 3–17. doi: 10.1007/978-3-030-91308-3_1
- Rojas, P. M. W., Noriega, L. A. M., and Silva, A. S. (2019). Hemoglobin screening using cloud based mobile photography applications. *Ingeniería y Universidad* 23, 2. doi: 10.11144/Javeriana.iyu23-2.hsuc
- Santafe, G., Inza, I., and Lozano, J. A. (2015). Dealing with the evaluation of supervised classification algorithms. *Artif. Intell. Rev.* 44, 467–508. doi: 10.1007/s10462-015-9433-y
- Schonlau, M., and Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata J.* 20, 3–29. doi: 10.1177/1536867X20909688
- Senvar, O., and Sennaroglu, B. (2016). Comparing performances of clements, box-cox, johnson methods with weibull distributions for assessing process capability. *J. Ind. Eng. Manag.* 9, 634–656. doi: 10.3926/jiem.1703
- Sevani, N., Persulesy, G., and Persulesy, G. B. V. (2018). “Detection anemia based on conjunctiva pallor level using k-means algorithm,” in *IOP Conference Series: Materials Science and Engineering*. Bristol: IOP Publishing.
- Singh, D., and Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* 97, 105524. doi: 10.1016/j.asoc.2019.105524
- Suner, S., Rayner, J., Ozturan, I. U., Hogan, G., Meehan, C. P., Chambers, A. B., et al. (2021). Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera. *PLoS ONE* 16, e0253495. doi: 10.1371/journal.pone.0253495
- Tamir, A., Jahan, C. S., Saif, M. S., Zaman, S. U., Islam, M. M., Khan, A. I., et al. (2017). “Detection of anemia from image of the anterior conjunctiva of the eye by image processing and thresholding,” in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. Dhaka: IEEE, 697–701.
- Tharwat, A. (2021). Classification assessment methods. *Appl. Comp. Informa.* 17, 168–192. doi: 10.1016/j.aci.2018.08.003
- Trentin, E. (2015). Maximum-likelihood normalization of features increases the robustness of neural-based spoken human-computer interaction. *Pattern Recognit. Lett.* 66, 71–80. doi: 10.1016/j.patrec.2015.07.003
- Wang, E. J., Li, W., Hawkins, D., Gernsheimer, T., Norby-Slycord, C., and Patel, S. N. (2016). “Hemaapp: noninvasive blood screening of hemoglobin using smartphone cameras,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 593–604.
- World Health Organization. (2011). *Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity*. Geneva: World Health Organization.