



Artificial Neural Networks for Prediction of Tuberculosis Disease

Muhammad Tahir Khan^{1,2†}, Aman Chandra Kaushik^{2†}, Linxiang Ji³,
Shaikat Iqbal Malik^{1*}, Sajid Ali⁴ and Dong-Qing Wei^{2*}

¹ Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan,

² College of Life Sciences and Biotechnology, The State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai, China, ³ Department of Physics, Thompson Rivers University, Kamloops, BC, Canada, ⁴ Provincial Tuberculosis Reference Laboratory, Hayatabad Medical Complex, Peshawar, Pakistan

Background: The global burden of tuberculosis (TB) and antibiotic resistance is attracting the attention of researchers to develop some novel and rapid diagnostic tools. Although, the conventional methods like culture are considered as the gold standard, they are time consuming in diagnostic procedure, during which there are more chances in the transmission of disease. Further, the Xpert MTB/RIF assay offers a fast diagnostic facility within 2 h, but due to low sensitivity in some sample types may lead to more serious state of the disease. The role of computer technologies is now increasing in the diagnostic procedures. Here, in the current study we have applied the artificial neural network (ANN) that predicted the TB disease based on the TB suspect data.

Methods: We developed an approach for prediction of TB, based on an ANN. The data was collected from the TB suspects, guardians or care takers along with samples, referred by TB units and health centers. All the samples were processed and cultured. Data was trained on 12,636 records of TB patients, collected during the years 2016 and 2017 from the provincial TB reference laboratory, Khyber Pakhtunkhwa, Pakistan. The training and test set of the suspect data were kept as 70 and 30%, respectively, followed by validation and normalization. The ANN takes the TB suspect's information such as gender, age, HIV-status, previous TB history, sample type, and signs and symptoms for TB prediction.

Results: Based on TB patient data, ANN accurately predicted the *Mycobacterium tuberculosis* (MTB) positive or negative with an overall accuracy of >94%. Further, the accuracy of the test and validation were found to be >93%. This increased accuracy of ANN in the detection of TB suspected patients might be useful for early management of disease to adopt some control measures in further transmission and reduce the drug resistance burden.

Conclusion: ANNs algorithms may play an effective role in the early diagnosis of TB disease that might be applied as a supportive tool. Modern computer technologies should be trained in diagnostics for rapid disease management. Delays in TB diagnosis and initiation treatment may allow the emergence of new cases by transmission, causing high drug resistance in countries with a high TB burden.

Keywords: ANN, TB, data, diagnosis, drug resistance

Abbreviations: ANN, artificial neural network; ATH, Ayub Teaching Hospital; ATO, Agency TB officers; CSE, cerebrospinal fluid; DTO, district TB control officers; KPK, Khyber Pakhtunkhwa; MMTH, Mufti Mehmood Teaching Hospital; MTB, *Mycobacterium tuberculosis*; PMDT, Programmatic Management of Drug Resistant TB; PTRL, Provincial tuberculosis control program; TB, tuberculosis.

OPEN ACCESS

Edited by:

Xing Chen,
China University of Mining
and Technology, China

Reviewed by:

Feng Zhu,
Zhejiang University, China
Willy Ssenooba,
Makerere University, Uganda

*Correspondence:

Shaikat Iqbal Malik
drshaukat@cust.edu.pk
Dong-Qing Wei
dqwei@sjtu.edu.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 12 January 2019

Accepted: 14 February 2019

Published: 04 March 2019

Citation:

Khan MT, Kaushik AC, Ji L,
Malik SI, Ali S and Wei D-Q (2019)
Artificial Neural Networks
for Prediction of Tuberculosis Disease.
Front. Microbiol. 10:395.
doi: 10.3389/fmicb.2019.00395

INTRODUCTION

According to the (World Health Organization [WHO], 2018), 1.7 billion people (23%) of the world's population are estimated to have latent TB infection, indicating a risk of developing active TB during their lifetime (World Health Organization [WHO], 2018). Approximately 10.4 million incidences of TB occurred worldwide, including 5 million (56%) men, 3.5 million (34%) women and 1 million (10%) among children (WHO, 2017). Due to an increase in the world's population, the health care units are continuously struggling to improve the standard and reduce the transmission and cost. Methods commonly used to diagnose TB include, GeneXpert assay, sputum-smear microscopy and chest radiography (Dheda et al., 2017; Ejeta et al., 2018). However, diagnosis became more complicated when the infectious agent spread to other parts of the body – this is referred to as extra pulmonary TB. All these diagnostic methods possess some limitations. Culture method is considered the gold standard for detection of the causative agent of TB, *Mycobacterium tuberculosis* (MTB) but it is time consuming in diagnosis and the chances of contamination are high (Crowle et al., 1991; Osman et al., 2010; Asgharzadeh et al., 2015). Some common issues reported from other diagnostics methods include performance issues, sputum samples from children (pediatric cases), live MTB, highly skilled medical staff for high throughput tools and high cost (Dheda et al., 2017; Pandey et al., 2017). Delay in diagnosis may lead to drug resistance, multidrug resistance (MDR), where an isolate shows resistance to two first line drugs, rifampicin and isoniazid, and extensive drug resistance (XDR) which include MDR and also show resistance to fluoroquinolones and at least one of the injectable drug (Seung et al., 2015).

In health sciences, wet lab tests can be time consuming and the chances of contamination could further lead the disease to an irreversible state. Although the Xpert MTB/RIF assay offers a fast diagnostic facility within 2 h, but due to low sensitivity in some sample types and cost may lead to more serious state of disease (Pandey et al., 2017).

In the last few decades, the researchers have collected an extensive amount of biological data in genomics, proteomics, and in some other fields of biology during the gene and protein expression analysis. To extract some meaningful information and interpret the results, high throughput computational algorithms have been developed (Fojnica et al., 2016; Dande and Samant, 2018). In bioinformatics, data mining is a process of extracting useful information deep inside of large datasets (Sebban et al., 2002; Zheng et al., 2008; Li et al., 2013). These techniques also involve artificial intelligence, statistics, machine learning, and visualization (Li et al., 2013; Dande and Samant, 2018). Such techniques are applied to expose and analyze the hidden information inside the data or sometimes also called Intelligent Data Analysis (IDA), for better prediction of results. This knowledge discovery obtained from health data has some major objectives, including diagnosis in health sciences and simulations (Mello et al., 2006; Guillet and Hamilton, 2007; Chang et al., 2012).

Traditionally implemented diagnostic methods for tuberculosis patients can be minimized with data mining approaches.

National and International laboratory researchers are currently involved in developing new diagnostics, and their evaluation helps to introduce more rapid and accurate methods in the diagnosis of TB along with the evaluations of alternative algorithms for TB reference laboratories (Parsons et al., 2011).

Artificial Neural Networks (ANNs) are operated by using algorithms to interpret non-linear data, independent of sequential pattern. The networks consist of a number of smaller units called neurons, organized between the input of data and the output of results into many layers. The ANN perform and behave like biological neurons, and this behavior may be learned through a backpropagation process. In this process, the precise output of a data set is previously known as input into the network. The least mean square difference of the entire data set is minimized by the continuous comparison of output of the ANN to the known output. A good level precision is adjusted by performing complex tasks without many computing resources (Drew and Monson, 2000).

The main advantage usually provided by ANN is their capability to extract hidden linear and non-linear relationships, even in the high dimensional and complex data sets (Zhang et al., 2010). In order to ease clinical decision makers, some more rapid evaluation techniques with low costs and good precision may further support in the diagnosis of TB to give optimum time for therapy, especially in TB high burden countries. Modern methods in data mining along with some traditional methods like regression have proven to be useful for comparison of prediction power of different models. The objective behind the current study is to support physicians in diagnosis, using predictive models as a diagnostic method for TB. Here, this investigation presents the data mining methods, i.e., classification, decision tree algorithm on the TB suspect data sets with selected attributes of patients to predict the presence or absence of TB disease. This information can be applied to develop less expensive diagnostic methods, dropping drug effects, data modeling, management of health care information systems, public health, and also patient's future prediction.

MATERIALS AND METHODS

Ethics Statement

This study proposal was approved and permitted by Institutional Committee (Ref 30/CUST2017) and incharge and molecular biologist, Provincial Tuberculosis Research Lab (PTRL) (Ref No. 1-06-17) where individual patient names and sensitive information were removed and neither of these have been linked with an individual TB suspect. Further, the study was also conducted according to the WHO Standards and Operational Guidance for Ethics Review of World Health Organization [WHO] (2011). Annex-3(IV)-B (13).

Data Mining of TB Patient Data

Data in the current research was retrieved from TB control program at PTRL, Hayatabad Medical Complex Peshawar. All follow up and diagnostic patients have been included and the data of patients has been collected from guardians or care takers.

TABLE 1 | Characteristics of TB suspects/patients received from KPK TB units.

Characteristics	No.	Characteristics	No.	Characteristics	No.
Gender		Hiv status		Sample type	
Female	6445	Hiv +ve	534	Sputum	11089
Male	6043	Hiv –ve	5	Pleural fluid	343
*Other	148	Unknown	12097	*BAL	312
Total	12636	Total	12636	Gastric aspirate	224
				*CSF	172
				Pus	124
History		Age group		Ascetic fluid	117
Diagnosis	10416	0.1–14	1072	Tissue	104
Follow up	2220	15–24	3474	Urine	69
Total	12636	25–34	2438	Pericardial fluid	43
		35–44	1678	Synovial fluid	16
Disease type		45–54	1690	Lymph node	11
Extra pulmonary	1323	55–64	1336	Total	12636
Pulmonary	11313	65–74	661	Culture	
Total	12636	75–84	222	MTB	1809
		85–94	55	No growth	10827
		95–104	8	Total	12636
		100–105	2		
		Total	12636		

*CSF, cerebrospinal fluid; BAL, bronchoalveolar lavage; Other, she male.

The data include location, age, gender, sample type, history, HIV status. The data set contain information's from 36 different TB units of Khyber Pakhtunkhwa (KPK). The characteristics of data is given **Table 1, 2**.

Dataset Development

Suspects referred by TB units and health care centers during the years 2016 and 2017, were included. Samples were processed and

TABLE 2 | Number of TB suspects received from different units of KPK province.

S.No.	Health center	Cases	S.No.	Health center	Cases
1	*ATO Khyber agency	71	20	DTO Sawabi	32
2	*CMH	296	21	DTO Shangla	4
3	*DTO Bajawar agency	11	22	DTO Swat	6
4	DTO Bannu	8	23	Hayatabad medical complex	1903
5	DTO Buner	8	24	Khyber teaching hospital	699
6	DTO Charsadda	54	25	Kuwait hospital	21
7	DTO Chitral	5	26	Leady reading hospital	2
8	DTO D.I. Khan	8	27	Mardan medical complex	8
9	DTO Dir Lower	7	28	*PMDT ATH	627
10	DTO Dir Upper	13	29	PMDT leady reading hospital	5333
11	DTO Hangu	39	30	PMDT *MMTH	1255
12	DTO Kohat	8	31	PMDT Swat	855
13	DTO Kohistan	1	32	Private	2
14	DTO Lakki	2	33	Peshawar reference lab	111
15	DTO Malakand	4	34	Rehman medical institute	66
16	DTO Mansehra	1	35	DTO Takht Nusrati	4
17	DTO Mardan	17	36	TB control Ganj	1
18	DTO Noshera	37			
19	DTO Peshawar	1117	Total		12636

*ATH, Ayub Teaching Hospital; ATO, Agency TB officers; CMH, Combined Military Hospital; DTO, district TB control officers; MMTH, Multi Mehmood Teaching Hospital; PMDT, Programmatic Management of Drug Resistant TB.

cultured according the previous study (Kent and Kubica, 1985; Khan et al., 2018) and MTB negative and positive was confirmed after the culture result. Further confirmation was carried, using BD MGIT MTBc identification test (TBc ID, Ref: 245159, Becton, Dickinson), a rapid chromatographic immunoassay which detects the MTB complex antigen MPT64 secreted during culture (Arora et al., 2015). The dataset was validated by MATLAB software (Attaway, 2013). Total 12,636 inputs were used in order achieve a good output efficiency, where 70% were used as a training dataset and the remaining 30% were used as testing (Drew and Monson, 2000; Kulkarni et al., 2017). The validation observed for test dataset was about 93.71%.

Artificial Neural Networks Approach

Artificial neural networks are nature inspired algorithms (Fojnica et al., 2016) that include input layer node, hidden layer node, and output node. Every node in a layer has one parallel node in the layer following it, thereby consequently building the stacking. Back-propagation learning algorithm is based on gradient descent search algorithms to fiddle with the correlation weight (Sollich and Krogh, 1996; Kaushik and Sahi, 2018). The output of every neuron was the aggregation of information of neurons of the prior stage multiplied by parallel weights with biased value. Input value was transformed into output with respect to activated functions shown in **Figure 1**.

Step 1 – Normalization of MTB Dataset.

TB patients dataset was normalized according to proposed study (Kaushik and Sahi, 2018).

$$V_{\text{new}} = (V_{\text{old}} - \text{Min}V) / (\text{Max}V - \text{Min}V) * (D_{\text{max}} - D_{\text{min}}) + D_{\text{min}}$$

where V_{new} represent new assessment post-normalization, V_{old} is the assessment before normalization, $\text{Min}V$ is the variable's minimum assessment, $\text{Max}V$ is the variable's maximum estimation. D_{max} and D_{min} are the maximum estimation succeeding normalization and the minimum assessment subsequent to normalization, respectively.

Step 2 – Input the data for training, the interrelated values of input and output execute for training using feed forward back propagation neural network algorithms.

Step 3 – Set Network constraint.

Step 4 – Calculate the neurons of output, every neuron output signal calculated using

$$\text{net}_j = \sum_{i=1 \sim m} w_{ji} x_i + b_j$$

where net_j and w_{ji} are output neurons and connection weight neurons, respectively, while x_i and b_j are the input signal, and bias neurons. The sigmoid function or logistic function, also called the sigmoidal curve (Seggern, 2016), was used for net_j and every neuron of ten hidden layers.

Step 5 – Signal of output layers' calculation using,

$$\text{net}_k = \text{TV}_k + \delta_k^L$$

where TV_k is target value of output neurons and δ_k^L is the error of neuron.

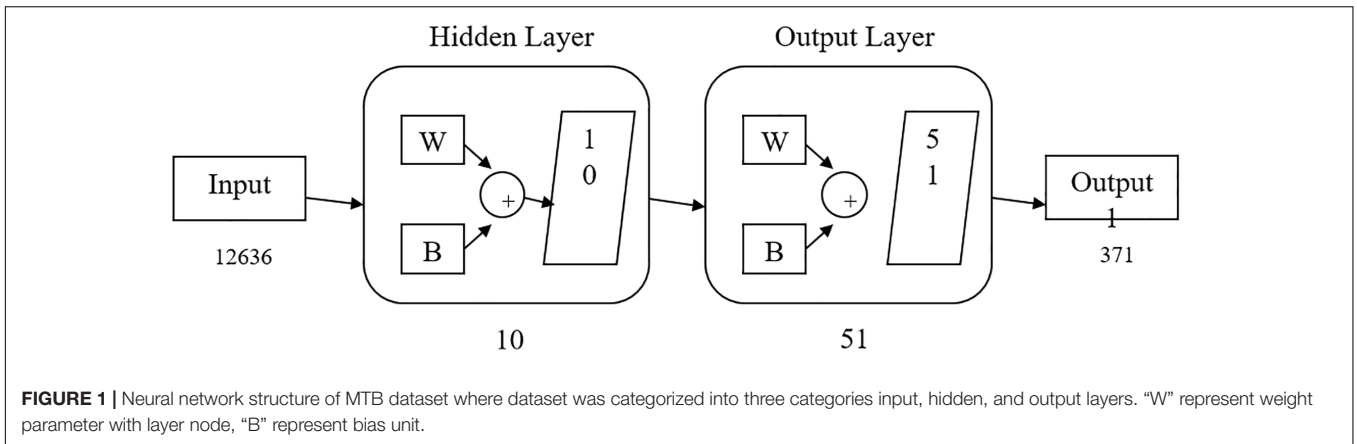


FIGURE 1 | Neural network structure of MTB dataset where dataset was categorized into three categories input, hidden, and output layers. “W” represent weight parameter with layer node, “B” represent bias unit.

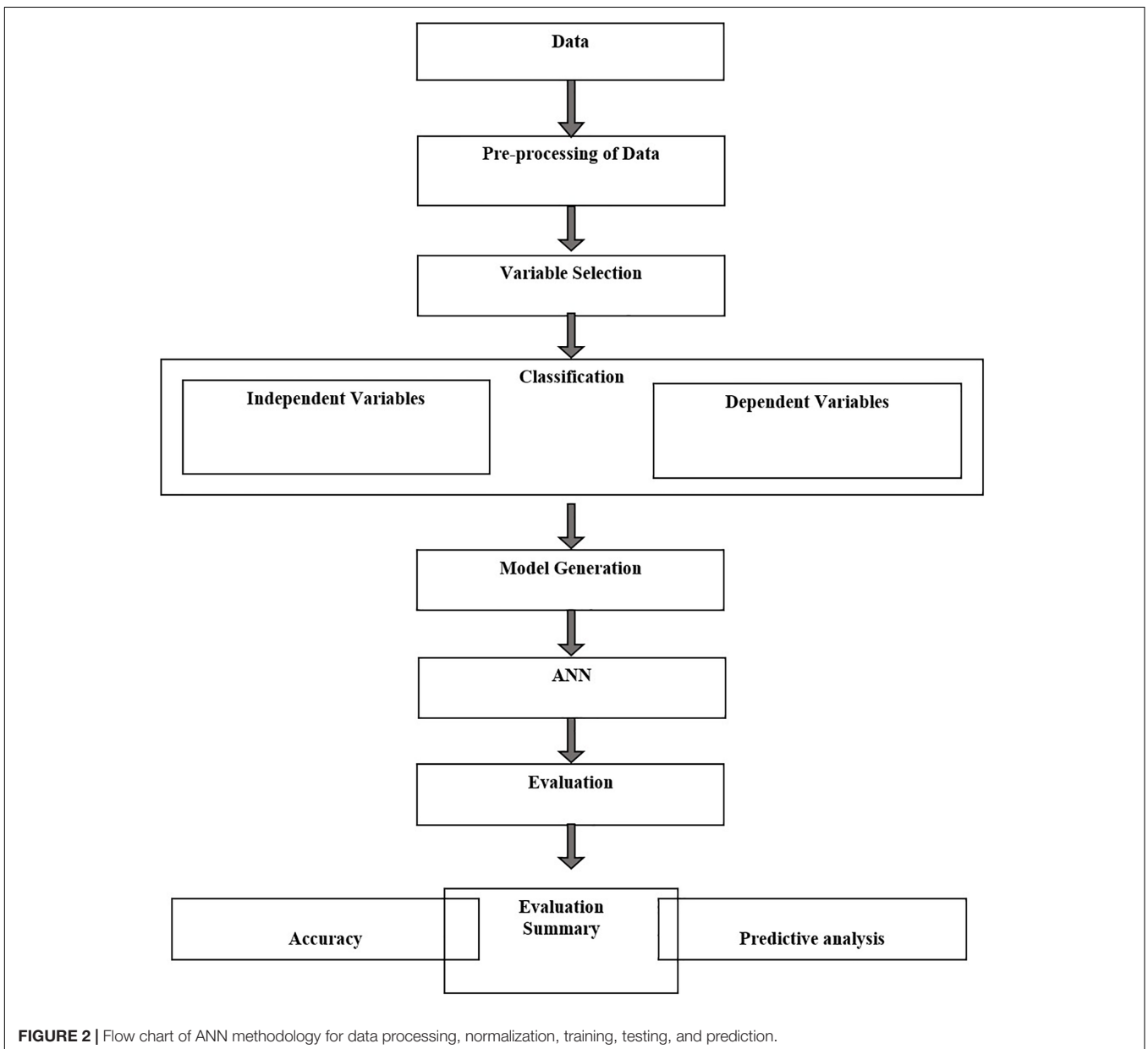


FIGURE 2 | Flow chart of ANN methodology for data processing, normalization, training, testing, and prediction.

Step 6 – Compute the error of neuron k and step 3 and step 6 were repetitive until network was congregated, and the error was computed using,

$$SSE = \sum_{i=1 \sim n} (T_i - Y_i)^2$$

where T_i is actual assessment and Y_i is estimated assessment.

A step by step flowchart methodology has been given in **Figures 1, 2**.

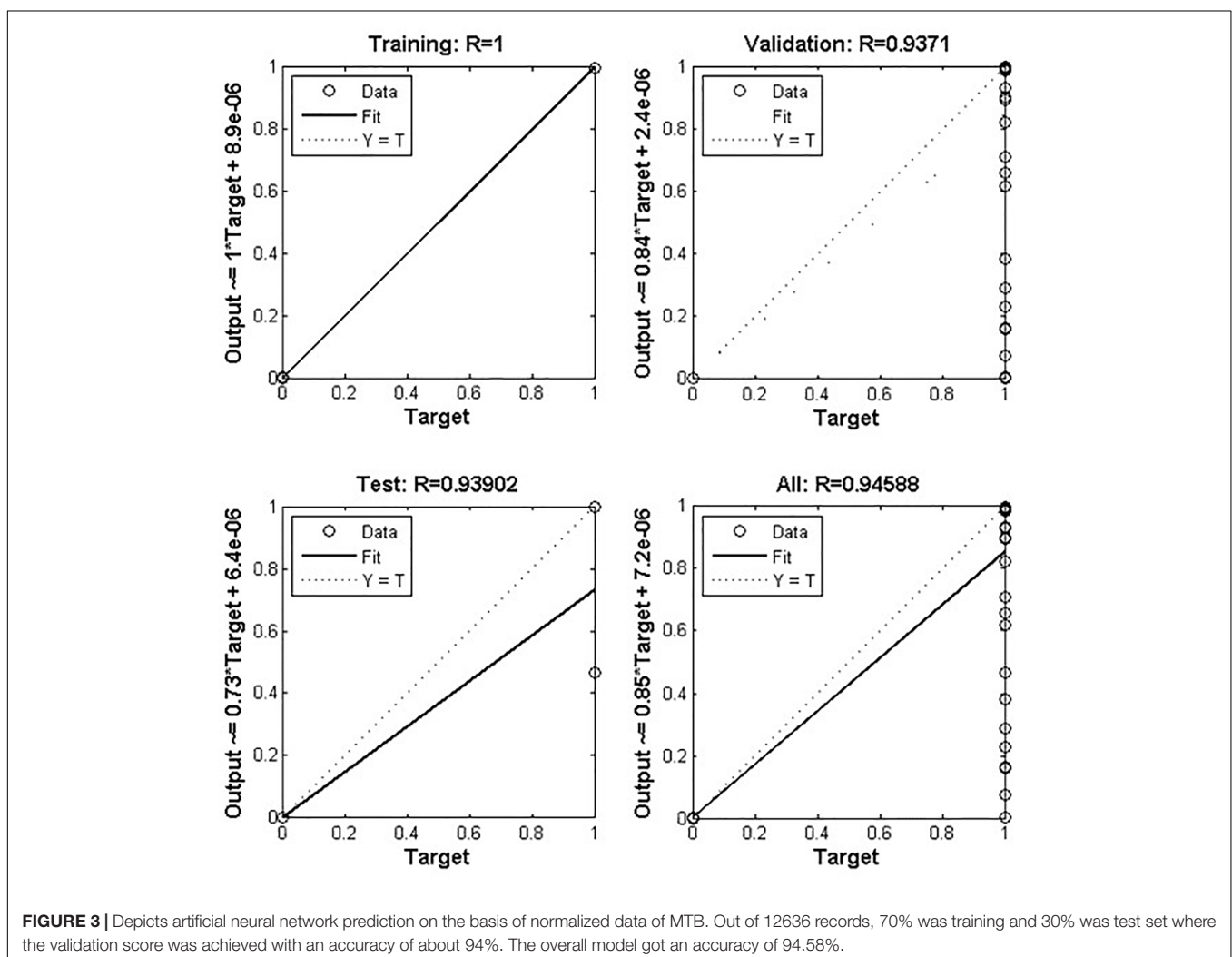
RESULTS

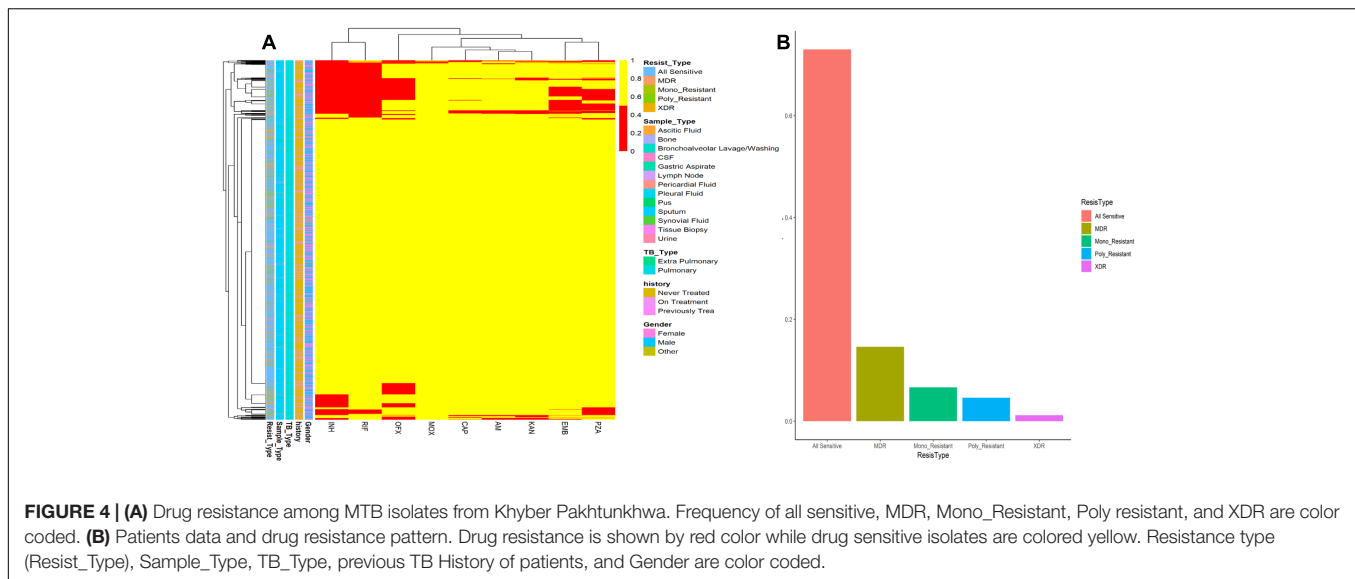
The drug resistance and patient’s characteristics has been shown in **Figure 4**. MDR are very high among the population of KPK followed by other first line drug resistance. Although XDR are very few, they are very hard to treat and often take years to recover and the chances are very rare. Mono-resistant (resistant to any single drug) and poly resistant (resistant to any two or more drugs other than MDR and XDR) have been found in significant numbers. This high prevalence of drug resistance may be due to

the delay type of diagnosis of some gold standard methods like culture. Owing to the current situation, we applied ANN on the patient records to find the accuracy of prediction.

Based on the ANN, the data used in the current study has 12,636 records (**Table 1**), where 70 and 30% were used as training and test sets, respectively. The accuracies of test and validation to predict TB based on patient data, were found at 93.90 and 93.71%, respectively. ANN-based, this algorithm accurately predicted that a suspect may have TB or not and generated the output through the hidden layer implementation. The hidden layers, learning parameters of ANN were as follows.

Number of input layer units	12636
Number of hidden layers	10
Number of first hidden layer units	10
Number of second hidden layer units	10
Number of output layer units	1
Momentum rate	0.88
Learning rate	0.70
Error after learning	0.000050
Learning cycle	30,000





The Architecture of ANN

1. Initialize the weight and parameters μ ($\mu = 0.01$)
2. Compute the sum of the squared errors overall input $F(w) = e^T e$
Where weight of network $w = [w_1, w_2, w_3, \dots, w_n]$ and e is the error vector for the network.
3. Solve to obtain the increment of weight $\Delta w = [J^T J + \mu I]^{-1} J^T e$
Where J is jacobian matrix, μ is learning rate neither μ is multiplied by decay rate β ($0 < \beta < 1$).
4. Using $w + \Delta w$
 $F(w) < F(w)$ then (go back to Step 2)
 $W = w + \Delta w$
 $\mu = \mu \cdot \beta$ ($\beta = 0.1$) (go back to Step 2)
ELSE
 $\mu = \mu / \beta$ (go back to Step 2)
END IF

The approach has been found efficient and possesses robust application for TB disease prediction with prediction accuracy. Using this approach, users can predict the active TB positively or negatively based on the patient's data after clinical sign symptoms including, cough that lasts 3 weeks or longer and pain in the chest coughing up blood or sputum (mucus from deep inside the lungs). Other symptoms of TB disease may include; weakness or fatigue, weight loss, no appetite, chills, fever, sweating at night (CDC Tuberculosis (TB), 2019).

User input data include; age, gender (male, female, other), sample type [bone, bone marrow, bronchoalveolar lavage, cerebrospinal fluid (CSF), gastric aspirate, lung biopsy, lymph nodes, pericardial fluid, pleural fluid, synovial fluid, pus, tissue biopsy, urine, sputum], history (follow up, diagnostic), HIV status (positive, negative, unknown). The model includes 70% training and 30% test set of the entire data set (12636 records) where the validation score was achieved with an accuracy of 94%.

The approach was written in MATLAB script where prediction accuracy was achieved as >94% based on ANN (Figure 3),

dependent on dataset, where dataset input and hidden layer were categorized on two basic parameters W (weight) and B (bias unit), which contain ten sub-models and generate single output based on dataset accuracy. Users can predict their TB risk after entering their data, history, and the appearance of signs and symptoms.

DISCUSSION

Tuberculosis is a challenging disease; in spite of advanced technologies, the diagnosis is often difficult because of the nature of the disease (Dhedha et al., 2017; WHO, 2017; World Health Organization [WHO], 2018). Clinical diagnosis requires standardization, where immunodiagnostic tests may help to improve sensitivity, but not in latent TB and some lack specificity (Newton et al., 2008; Elhassan et al., 2016). Xpert MTB/RIF have been saving our time to detect MTB, but decades old technologies like culture still remained the standard. Today, the battle against TB still poses one of the primary diagnostic problems in the pediatric laboratory (Dunn et al., 2016). Delay in notification and a weak coordination among TB management might be a cause to unnecessary diagnosis and treatment initiation (Yagui et al., 2006; Htun et al., 2018). Although the Xpert MTB/RIF assay offers fast diagnostic facility within 2 h, in some sample types like lymph node tissue biopsy (extrapulmonary TB) the overall sensitivity to rule out the TB is suboptimal (Creswell et al., 2014; Pandey et al., 2017; Tadesse et al., 2018). Performance was found to vary according to specimen type and acid-fast bacilli smear status. Further, the gold standard for MTB drug susceptibility testing is still culture on solid media, taking weeks to months to grow (Lin and Desmond, 2014; Dookie et al., 2018; Koch et al., 2018). Treatment is often empirical and initiated after looking at factors like past medical or social history, or the prevalence of drug resistance in that locality. These may delay the initiation of proper TB treatment that lead to drug resistance (Schaberg et al., 1996; Melchionda et al., 2013; Dookie et al., 2018; Khan et al., 2018). The high prevalence of drug resistance in TB high burden

countries may delay the initiation of appropriate treatment due to culture of MTB *in vitro* which is a time consuming method. These issues should be addressed through more studies especially in TB high burden regions. We found an increased drug resistance in the recent years' data (Figure 4) calculation, pointing toward advanced computational technologies to integrate for diagnosis in MTB prediction (Khan et al., 2018).

Modern neural networks have attained a great significance and importance in the recognition of images (Drew and Monson, 2000; Krizhevsky et al., 2012; Esteva et al., 2017; Kaushik and Sahi, 2018), speech recognition (Hinton et al., 2012), and processing of natural language (Socher et al., 2011). Medicinal researchers have started to apply these tactics in personalized clinical care. Diabetic retinopathy has also been identified through the approaches of deep neural networks (Gulshan et al., 2016) and classifying cancers of skin (Esteva et al., 2017). The applications of such approaches have also been found to be successful in computational biology and bioinformatics such as in inferring target gene expression (Chen et al., 2016), predicting RNA-binding protein sites (Zhang et al., 2016), and in identification and prediction of biomarkers for human chronological age (Putin et al., 2016). To reduce cost and time wastage, various data mining approaches may be helpful in diagnosis and on time initiation of TB therapy.

According to the WHO global TB report, 2018, India, Indonesia, China, Philippines, and Pakistan are the top five countries with 56% TB prevalence of the world. Timely TB diagnosis to reduce transmission and initiation of treatment to improve the outcomes for TB patients is essential, especially in high burden countries (Yagui et al., 2006; Dheda et al., 2017).

Classification and clustering algorithms are working efficiently with good precision in the prediction of the tuberculosis diagnosis. Presence of MTB and patient's data may support such model up to large extents. When handling high-dimensional classification problems, different modeling approaches may be used. Earlier works have applied multivariate logistic regression (Wisnivesky et al., 2005; Solari et al., 2008), classification trees (Mello et al., 2006; Aguiar et al., 2012) and ANN (Aguiar et al., 2013; Dande and Samant, 2018) for predicting smear-negative TB.

CONCLUSION

Artificial neural networks may be applied as a diagnostic tool for TB prediction and supportive in expanding the role for computer technologies in diagnostics for a rapid management of

TB. Therefore, this high correlation (>94% accuracy) with the experimental result of MTB detection may help to choose optimal therapeutic regimens, especially in TB high burden countries. Delays in TB diagnosis and initiation of treatment may allow the emergence of new cases by transmission, and is one of the causes of high drug resistance in TB high burden countries.

The approach developed here may offer and support the rapid diagnosis of MTB with further additions such as drug resistance prediction in near future for better TB management.

DATA AVAILABILITY

The datasets in the present study will be provided upon reasonable request to the corresponding author.

AUTHOR CONTRIBUTIONS

Manuscript was designed by MK, DW, SM, and LJ. ANN was written and run by AK. Data analysis and manuscript writing were carried out by MK, SA, SM, and AK. Manuscript was revised by DW, LJ, and AK.

FUNDING

The work was supported by the grants from the Key Research Area grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract Nos. 61832019 and 61503244), the State Key Lab on Microbial Metabolism, and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The simulations in this work were supported by the Center for High Performance Computing, Shanghai Jiao Tong University and also Higher Education Commission Islamabad, Pakistan under IRSIP No: 1-8/HEC/HRD/2017/8392.

ACKNOWLEDGMENTS

The present study was supported by SA and Anwar Sheed Khan, Molecular biologist microbiologist and project director PTRL Peshawar. Deborah L. Devis, University of Adelaide-Shanghai Jiao Tong University Joint Centre for Agriculture and Health, School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Urrbrae, SA, Australia.

REFERENCES

- Aguiar, F., Pinto, J., Torres, R., Seixas, J., and Mello, F. (2013). Artificial neural network (ANN) for prediction of pulmonary tuberculosis in hospitalized patients. *Eur. Respir. J.* 42:4469.
- Aguiar, F. S., Almeida, L. L., Ruffino-Netto, A., Kritski, A. L., Mello, F. C., and Werneck, G. L. (2012). Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients. *BMC Pulm. Med.* 12:40. doi: 10.1186/1471-2466-12-40

- Arora, J., Kumar, G., Verma, A. K., Bhalla, M., Sarin, R., and Myneedu, V. P. (2015). Utility of MPT64 antigen detection for rapid confirmation of *Mycobacterium tuberculosis* complex. *J. Glob. Infect. Dis.* 7, 66–69. doi: 10.4103/0974-777X.154443
- Asgharzadeh, M., Kafil, H. S., Ghorghanlu, S., and Pourostadi, M. (2015). Laboratory cross-contamination of *Mycobacterium tuberculosis* in Northwest of Iran. *Egypt. J. Chest Dis. Tuberc.* 64, 665–669. doi: 10.1016/j.ejcdt.2015.03.018
- Attaway, S. (2013). *Matlab: A Practical Introduction to Programming and Problem Solving*. Amsterdam: Elsevier, doi: 10.1016/B978-0-12-405876-7.00014-16

- CDC Tuberculosis (TB) (2019). *Disease: Symptoms and Risk Factors [Features]* CDC. Available at: <https://www.cdc.gov/features/tbsymptoms/index.html> [Accessed September 22, 2018].
- Chang, J., Arbeláez, P., Switz, N., Reber, C., Tapley, A., Davis, J. L., et al. (2012). Automated tuberculosis diagnosis using fluorescence images from a mobile microscope. *Med. Image Comput. Comput. Assist. Interv.* 15, 345–352.
- Chen, Y., Li, Y., Narayan, R., Subramanian, A., and Xie, X. (2016). Gene expression inference with deep learning. *Bioinformatics* 32, 1832–1839. doi: 10.1093/bioinformatics/btw074
- Creswell, J., Codlin, A. J., Andre, E., Micek, M. A., Bedru, A., Carter, E. J., et al. (2014). Results from early programmatic implementation of Xpert MTB/RIF testing in nine countries. *BMC Infect. Dis.* 14:2. doi: 10.1186/1471-2334-14-12
- Crowle, A. J., Dahl, R., Ross, E., and May, M. H. (1991). Evidence that vesicles containing living, virulent *Mycobacterium tuberculosis* or *Mycobacterium avium* in cultured human macrophages are not acidic. *Infect. Immun.* 59, 1823–1831.
- Dande, P., and Samant, P. (2018). Acquaintance to artificial neural networks and use of artificial intelligence as a diagnostic tool for tuberculosis: a review. *Tuberculosis* 108, 1–9. doi: 10.1016/j.tube.2017.09.006
- Dheda, K., Gumbo, T., Maartens, G., Dooley, K. E., McNerney, R., Murray, M., et al. (2017). The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *Lancet Respir. Med.* 5, 291–360. doi: 10.1016/S2213-2600(17)30079-6
- Dookie, N., Rambaran, S., Padayatchi, N., Mahomed, S., and Naidoo, K. (2018). Evolution of drug resistance in *Mycobacterium tuberculosis*: a review on the molecular determinants of resistance and implications for personalized care. *J. Antimicrob. Chemother.* 73, 1138–1151. doi: 10.1093/jac/dkx506
- Drew, P. J., and Monson, J. R. T. (2000). Artificial neural networks. *Surgery* 127, 3–11. doi: 10.1067/msy.2000.102173
- Dunn, J. J., Starke, J. R., and Revell, P. A. (2016). Laboratory diagnosis of *Mycobacterium tuberculosis* infection and disease in children. *J. Clin. Microbiol.* 54, 1434–1441. doi: 10.1128/JCM.03043-3015
- Ejeta, E., Beyene, G., Bonsa, Z., and Abebe, G. (2018). Xpert MTB/RIF assay for the diagnosis of *Mycobacterium tuberculosis* and Rifampicin resistance in high human immunodeficiency virus setting in gambella regional state, southwest ethiopia. *J. Clin. Tuberc. Mycobact. Dis.* 12, 14–20. doi: 10.1016/j.jctube.2018.06.002
- Elhassan, M. M., Elmekki, M. A., Osman, A. L., and Hamid, M. E. (2016). Challenges in diagnosing tuberculosis in children: a comparative study from Sudan. *Int. J. Infect. Dis.* 43, 25–29. doi: 10.1016/j.ijid.2015.12.006
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056
- Fojnica, A., Osmanović, A., and Badnjević, A. (2016). “Dynamical model of tuberculosis-multiple strain prediction based on artificial neural network” in *Proceedings of the 2016 5th Mediterranean Conference on Embedded Computing (MECO)*, (Piscataway, NJ: IEEE), 290–293. doi: 10.1109/MECO.2016.7525763
- Guillet, F., and Hamilton, H. J. (eds) (2007). *Quality Measures in Data Mining*. Berlin: Springer-Verlag. doi: 10.1007/978-3-540-44918-8
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410. doi: 10.1001/jama.2016.17216
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal. Process. Mag.* 29, 82–97. doi: 10.1109/MSP.2012.2205597
- Htun, Y. M., Khaing, T. M. M., Yin, Y., Myint, Z., Aung, S. T., Hlaing, T. M., et al. (2018). Delay in diagnosis and treatment among adult multidrug resistant tuberculosis patients in yangon regional tuberculosis center, myanmar: a cross-sectional study. *BMC Health Serv. Res.* 18:878. doi: 10.1186/s12913-018-3715-4
- Kaushik, A. C., and Sahi, S. (2018). HOGPred: artificial neural network-based model for orphan GPCRs. *Neural. Comput. Appl.* 29, 985–992. doi: 10.1007/s00521-016-2502-2506
- Kent, P. T., and Kubica, G. P. (1985). *Public Health Mycobacteriology: A Guide for the Level III Laboratory*. Washington, DC: US Department of Health and Human Services.
- Khan, M. T., Malik, S. I., Ali, S., Sheed Khan, A., Nadeem, T., Zeb, M. T., et al. (2018). Prevalence of pyrazinamide resistance in khyber pakhtunkhwa, Pakistan. *Microb. Drug Resist.* 24, 1417–1421. doi: 10.1089/mdr.2017.0234
- Koch, A., Cox, H., and Mizrahi, V. (2018). Drug-resistant tuberculosis: challenges and opportunities for diagnosis and treatment. *Curr. Opin. Pharmacol.* 42, 7–15. doi: 10.1016/j.coph.2018.05.013
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). “Imagenet classification with deep convolutional neural networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 NIPS'12*, (Red Hook, NY: Curran Associates Inc.), 1097–1105.
- Kulkarni, P., Londhe, S., and Deo, M. (2017). Artificial neural networks for construction management: a review. *Soft Comput. Civil Eng.* 1, 70–88. doi: 10.22115/scce.2017.49580
- Li, X., Ng, S.-K., and Wang, J. T. L. (2013). *Biological Data Mining and Its Applications in Healthcare*, 1st Edn. New Jersey, NJ: World Scientific Publishing.
- Lin, S.-Y. G., and Desmond, E. P. (2014). Molecular diagnosis of tuberculosis and drug resistance. *Clin. Lab. Med.* 34, 297–314. doi: 10.1016/j.cl.2014.02.005
- Melchionda, V., Wyatt, H., Capocci, S., Garcia Medina, R., Solamalai, A., Katiri, S., et al. (2013). Amikacin treatment for multidrug resistant tuberculosis: how much monitoring is required? *Eur. Respir. J.* 42, 1148–1150. doi: 10.1183/09031936.00184312
- Mello, F. C. Q., Bastos, L. G. V., Soares, S. L. M., Rezende, V. M., Conde, M. B., Chaisson, R. E., et al. (2006). Predicting smear negative pulmonary tuberculosis with classification trees and logistic regression: a cross-sectional study. *BMC Public Health* 6:43. doi: 10.1186/1471-2458-6-43
- Newton, S. M., Brent, A. J., Anderson, S., Whittaker, E., and Kampmann, B. (2008). Paediatric tuberculosis. *Lancet Infect. Dis.* 8, 498–510. doi: 10.1016/S1473-3099(08)70182-70188
- Osman, M. K., Mashor, M. Y., and Jaafar, H. (2010). “Detection of *mycobacterium tuberculosis* in Ziehl-Neelsen stained tissue images using Zernike moments and hybrid multilayered perceptron network,” in *Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics*, (Piscataway, NY: IEEE), 4049–4055. doi: 10.1109/ICSMC.2010.5642191
- Pandey, S., Congdon, J., McInnes, B., Pop, A., and Coulter, C. (2017). Evaluation of the GeneXpert MTB/RIF assay on extrapulmonary and respiratory samples other than sputum: a low burden country experience. *Pathology* 49, 70–74. doi: 10.1016/j.pathol.2016.10.004
- Parsons, L. M., Somoskövi, Á., Gutierrez, C., Lee, E., Paramasivan, C. N., Abimiku, A., et al. (2011). Laboratory diagnosis of tuberculosis in resource-poor countries: challenges and opportunities. *Clin. Microbiol. Rev.* 24, 314–350. doi: 10.1128/CMR.00059-10
- Putin, E., Mamoshina, P., Aliper, A., Korzinkin, M., Moskalev, A., Kolosov, A., et al. (2016). Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging* 8, 1021–1033. doi: 10.18632/aging.100968
- Schaberg, T., Rebhan, K., and Lode, H. (1996). Risk factors for side-effects of isoniazid, rifampin and pyrazinamide in patients hospitalized for pulmonary tuberculosis. *Eur. Respir. J.* 9, 2026–2030. doi: 10.1183/09031936.96.09102026
- Sebban, M., Mokrousov, I., Rastogi, N., and Sola, C. (2002). A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*. *Bioinformatics* 18, 235–243. doi: 10.1093/bioinformatics/18.2.235
- Seggern, D. H. (2016). *CRC Standard Curves and Surfaces with Mathematics*. Boca Raton, FL: CRC, doi: 10.1201/b16001
- Seung, K. J., Keshavjee, S., and Rich, M. L. (2015). Multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis. *Cold Spring Harb. Perspect. Med.* 5:a017863. doi: 10.1101/cshperspect.a017863
- Socher, R., Lin, C. C. Y., Ng, A. Y., and Manning, C. D. (2011). “Parsing natural scenes and natural language with recursive neural networks,” in *Proceedings of the 28th International Conference on International Conference on Machine Learning ICML'11*, (Washington, DC: Omni press), 129–136.
- Solari, A., Laurà, M., Salsano, E., Radice, D., Pareyson, D., and CMT-Trial Study Group (2008). Reliability of clinical outcome measures in charcot-marie-tooth disease. *Neuromuscul. Disord.* 18, 19–26. doi: 10.1016/j.nmd.2007.09.006
- Sollich, P., and Krogh, A. (1996). “Learning with ensembles: how over-fitting can be useful,” in *Advances in Neural Information Processing Systems*, eds D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Cambridge, MA: MIT Press), 190–196.
- Tadesse, M., Abebe, G., Bekele, A., Bezabih, M., Yilma, D., Apers, L., et al. (2018). Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis:

- a diagnostic evaluation study. *Clin. Microbiol. Infect.* doi: 10.1016/j.cmi.2018.12.018 [Epub ahead of print].
- WHO (2017). *WHO | Global Tuberculosis Report 2017*. Available at: http://www.who.int/tb/publications/global_report/en/ [Accessed November 12, 2017].
- Wisnivesky, J. P., Bonomi, M., Henschke, C., Iannuzzi, M., and McGinn, T. (2005). Radiation therapy for the treatment of unresected stage I-II non-small cell lung cancer. *Chest* 128, 1461–1467. doi: 10.1378/chest.128.3.1461
- World Health Organization [WHO] (2018). *Global Tuberculosis Report 2018*. Geneva: World Health Organization.
- World Health Organization [WHO] (2011). *Standards and operational guidance for ethics review of health-related research with human participants*. Geneva : World Health Organization. Available at: <http://www.who.int/iris/handle/10665/44783>
- Yagui, M., Perales, M. T., Asencios, L., Vergara, L., Suarez, C., Yale, G., et al. (2006). Timely diagnosis of MDR-TB under program conditions: is rapid drug susceptibility testing sufficient? *Int. J. Tuberc. Lung Dis.* 10, 838–843.
- Zhang, S., Zhou, J., Hu, H., Gong, H., Chen, L., Cheng, C., et al. (2016). A deep learning framework for modeling structural features of RNA-binding protein targets. *Nucleic Acids Res.* 44, e32. doi: 10.1093/nar/gkv1025
- Zhang, X., Pan, F., and Wang, W. (2010). “Finding high-order correlations in high-dimensional biological data,” in *Link Mining: Models, Algorithms, and Applications*, eds P. S. Yu, J. Han, and C. Faloutsos (New York, NY: Springer), 505–534. doi: 10.1007/978-1-4419-6515-8_19
- Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., and Minor, W. (2008). Data mining of metal ion environments present in protein structures. *J. Inorg. Biochem.* 102, 1765–1776. doi: 10.1016/j.jinorgbio.2008.05.006

Conflict of Interest Statement: 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.

Copyright © 2019 Khan, Kaushik, Ji, Malik, Ali and Wei. 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.