



An Effective Vision Based Framework for the Identification of Tuberculosis in Chest X-Ray Images

Tejasvi Ghanshala¹, Vikas Tripathi²(✉), and Bhaskar Pant²

¹ University of British Columbia, Vancouver, Canada
tejasviuniversity18@gmail.com

² Graphic Era Deemed to be University, Dehradun, India
vikastripathi.be@gmail.com, pantbhaskar2@gmail.com

Abstract. Tuberculosis is an infection that influences numerous individuals worldwide. While treatment is conceivable, it requires an exact conclusion first. Especially in developing countries there are by and large accessible X-beam machines, yet frequently the radiological aptitude is missing for precisely surveying the pictures. An automated vision based framework that could play out this undertaking rapidly and inexpensively could radically improve the capacity to analyze and at last treat the sickness. In this paper we propose image analysis based framework using various machine learning techniques like SVM, kNN, Random Forest and Neural Network for effective identification of tuberculosis. The proposed framework using neural network was able to classify better than other classifiers to detect Tuberculosis and achieves accuracy of 80.45% .

Keywords: Image processing · Neural Network · Tuberculosis · Chest X-Ray · Machine learning

1 Introduction

Tuberculosis (TB) is a potent killer of over millions of people all around the world. *Mycobacterium tuberculosis* is an etiological agent of TB and according to WHO in 2017, the estimated number of TB incidents was 10.0 million, comprising 3.2 million incidences in women and one million cases in children. Almost 1.6 million deaths were caused from TB including 0.4 million death of people with HIV [1]. At present, TB is the second leading cause of death from an infectious agent worldwide, after the Human Immunodeficiency Virus (HIV) [2]. The diagnostic and cure of tuberculosis is an obstacle that many researchers have been trying to overcome from several decades.

Diagnosis of pulmonary infection needs a combination of radiographic study, appropriate microbiological problems and clinical awareness. Chest X-ray is an inexpensive technique that is used for rapid detection of pulmonary abnormalities [3]. The radiologist faces a challenging situation during diagnosis of the patient with the pulmonary infection because various infections shares similar sign and symptoms.

The study of image processing will be helpful to identify the minors that the radiologist misses during the X-ray investigation. Drastically improving quantitative performances of image processing tools and algorithm is used for the recognition, detection and segmentation of features. Deep learning jointly with language and image learning tasks helps in knowledge-guided transfer learning, in image captioning and visual question answering [4].

In this paper we have showcased computer aided image analysis for easily identifying and discriminating tuberculosis from normal person's chest X-ray images.

Further paper is structured as follows: Sect. 2 consists of the literature work related to our work. In Sect. 3, our framework for the identification of tuberculosis is mentioned. Section 4 gives the result and discussion of our proposed work. And in Sect. 5 we have concluded our work with its future aspects.

2 Literature Review

Investigation of Pulmonary infection has been a diagnostic challenge since decades and the misdiagnosed patients have suffered from the ill treatment course. A case report of Cupples & Blackie and a case study on 105 patients done by the Dr. Barnes at USC School of Medicine, Los Angeles have suggested that clinical and radiographic features of pulmonary infections may sometimes be indistinguishable [5, 6]. A case study on a 22-year-old man was done by Pinto et al., reported that there can be occasional situations when pulmonary tuberculosis with respiratory failure can masquerade as pneumonia, and entertained as pneumonia by the physicians [7]. Similar case report of 16-year-old female was presented by Singh et al. [8]. The girl was suffering from non-productive cough, significant weight loss, joint pains, nocturnal fever and fatigue. The primary treatment for cough and fever was done by physician, and the differential diagnosis for pneumocystis pneumonia, viral/fungal pneumonia or miliary tuberculosis was performed with HIV test. After the demise of the patient, disease was diagnosed as Tuberculosis after seeing the result of Nucleic Acid Amplification [Xpert™ MTB/RIF]. But unfortunately, the misdiagnosis had killed the little angel [8]. The proper, accurate and inexpensive diagnosis is needed to distinguish between the pulmonary infections. The radiologist and physicians primarily rely on Chest X-Ray report for diagnosis of pulmonary infection. Rohmah et al., have presented an image processing-based approach to identify pulmonary tuberculosis. They tried to reduce the waiting time of diagnosis result. They have used the minimum distance classifier as classification method to detect the tuberculosis through chest X-ray images [9]. Poornimadevi et al., implemented an automated approach by using registration-based segmentation methods for detecting tuberculosis [10]. Fatima et al., have also developed an automated tuberculosis detection technique [11]. Parveen & Sathik, have used unsupervised fuzzy c-means classification learning algorithm for detection of pneumonia infection [12]. Sharma et al., have used a novel approach for detecting pneumonia clouds in chest X-rays [13]. Stephen et al., have deployed several data augmentation algorithms and improved the accuracy of CNN model for detecting pneumonia [14]. They have showcased the usage of deep neural network for detection of pneumonia.

Since inception of machine learning algorithms, application are of these algorithms have been grown exponentially. In today’s world almost everywhere we can find application of machine learning like smart city development, medical analysis, business, education etc. In medical field itself various algorithms have been proposed like SVM, kNN, Decision tree and its variations etc. for disease diagnostics. Fatima et al. have presented a comprehensive survey on usage of machine learning algorithms for identification and analysis of various diseases like heart, liver, hepatitis etc. [15]. In [16] authors have presented comparative analysis between the kNN and SVM algorithms. CNN have been efficiently used for activity detection in given video shown in [17].

The main motive of this paper is to provide an effective mechanism to identify TB in given Chest X-Ray images. As show cased on above literature survey, it is still a challenging task to identify disease in given X-Ray images. Further for identification of suitable machine learning algorithm our survey suggest that comparative analysis should be used to show case effectiveness of any specific algorithm for TB analysis.

3 Methodology

Machine learning has a wide variety of applications like cancer detection, helmet detection etc. We proposed a framework that will detect chest tuberculosis using neural network. Our framework is shown in Fig. 1. As shown in Fig. 1, images are fed in framework then features are extracted from images further training and testing is performed.

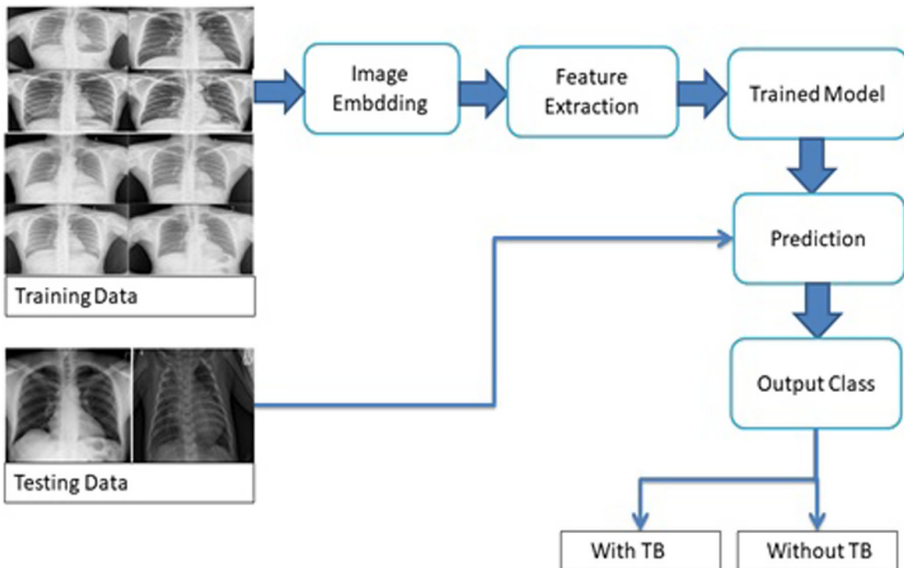


Fig. 1. A framework proposed for identification of Tuberculosis in chest X-Ray.

To get training and testing images, first we divide our database in the ratio of 7:3 and used 70% of the data for training, in training first we applied image embedding on our

database which has two classes with TB and without TB, with TB class depicts the people which are suffering from tuberculosis. Image embedding is used to calculate a feature vector for each image. It calculates features by reading an image and then uploading them to a remote server or evaluates them locally by using different machine models. Further we used these extracted features to train the neural network model which will be further used for prediction.

We have used different algorithms to compare our results. Image embedding has been done using Google's pre-trained Inception V-3 model. It uses CNN for the feature extraction. It takes images as input and passes it to the first convolution layer. The images are then dot product with the feature descriptor and a feature map is obtained. This feature map is further sent to pooling layer; in our case max-pooling has been used. The feature map is thus reduced and a convolution is formed. The process is repeated for the entire convolution layer and then the resultant feature map is converted to 1-D vector. The output is thus sent to ANN and the classification is done. In case of SVM the classification is done with the help of hyperplanes. The aim is to maximize the distance or margin between the supporting hyperplanes and to obtain the optimal decision boundaries. The third algorithm used is kNN in which the classification is based on finding the 'k' nearest neighbors. These k neighbors are calculated and the mode of the class of these k neighbors is chosen as the required class. And the last algorithm used is Random forest. This algorithm is the cluster of multiple decision trees that provide output by calculating the information gain or entropies of parent and child classes. After that the required class is chosen by voting the output of each decision tree.

4 Result and Discussion

To achieve results and deployment of machine learning algorithms, the experiments were run on Machine in which configuration includes Windows machine with Intel(R) Core (TM) i5-6200U CPU @ 2.30 GHz \times 4 with 6 GB RAM. The chest x-ray image dataset is taken from [18] and separated into two different classes. As per benchmark settings used for analysis, we divided the dataset into the ratio of 7:3, 70% of the images are used for training and 30% of the images are used for testing, validation size is kept 20% and we have used batch size of 10. The model was created using three convolution layers with filter size of 3 and number of filters used is 16, 32 and 64 respectively. To extract features for classification, flattening is used to convert multi-dimensional array into 1-D vector with layer size of 128, the model was trained with 1000 iterations and calculated epoch is 28. Some sample images of our dataset are shown in Fig. 2.

Figure 3 and Fig. 4 shows comparative ROC curve for both the classes respectively, both ROC curve contains comparison between different machine learning algorithms. These curve shows that performance of each algorithm is quite stable.

Table 1 includes proportional scrutiny in the form of various statistics like F1 score, recall and precision for both the classes when neural network classification have been deployed. We can see that the recall value of normal people is better than abnormal people means it is detecting normal people with more accuracy than abnormal people but the precision of detecting the abnormal people is more than precision of detecting normal people which is a good sign. Further results can be improved by incorporating

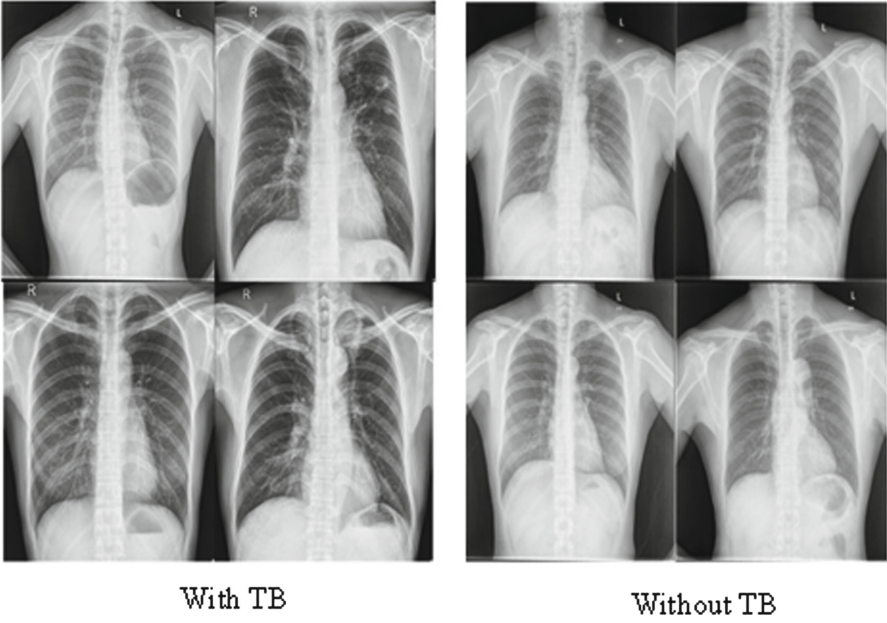


Fig. 2. Sample Chest X- Ray Images of TB and without TB

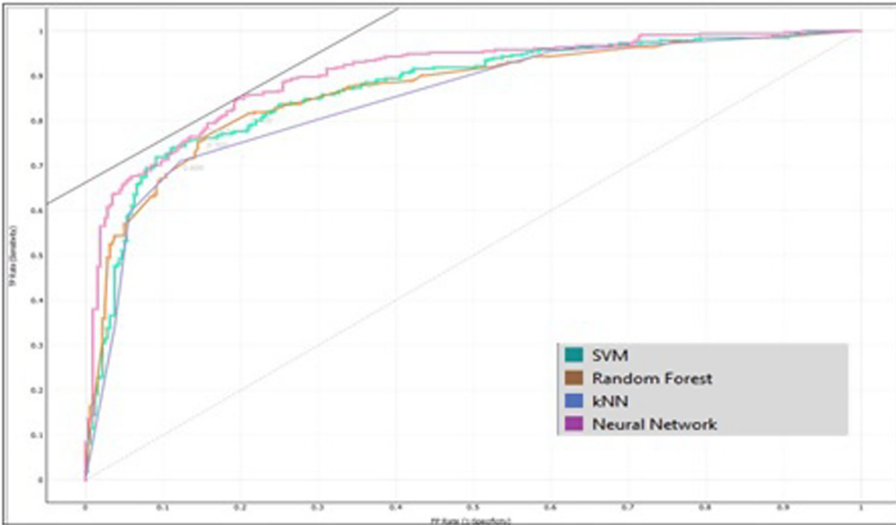


Fig. 3. ROC Curve of various algorithms for class TB

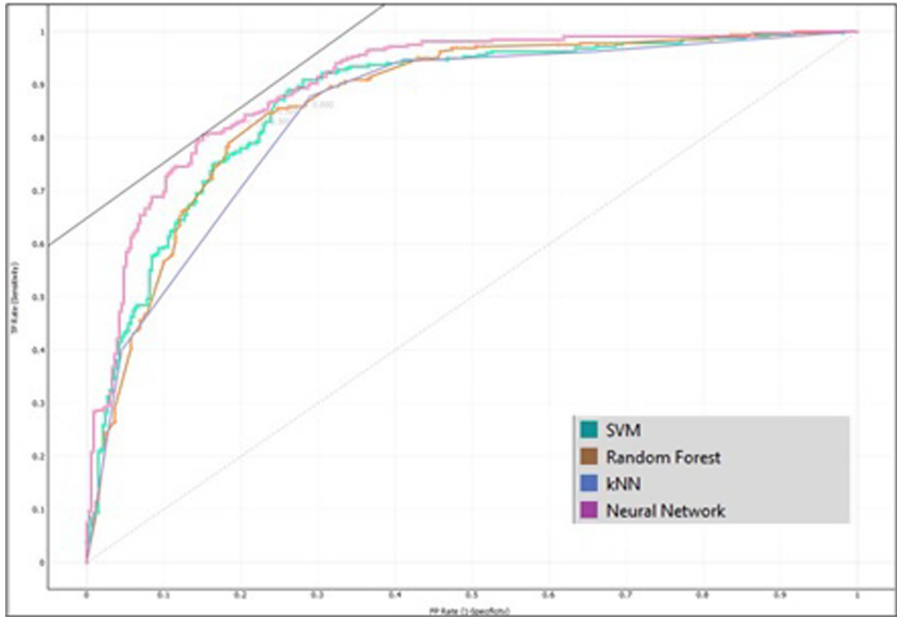


Fig. 4. ROC Curve of various algorithms for class without TB

more images on training and testing. Since images are limited and especially TB images are comparatively different in structure, framework is not able to classify all images of TB accurately however it is still giving efficient results.

Table 1. Comparison between F1-Score, Precision, Recall of different classes.

Class	F1-Score (%)	Precision (%)	Recall (%)
With TB	78.9	87.0	72.0
Without TB	82.0	76.0	89.0
Average	80.45	81.5	80.5

Further to show case the suitability of neural network for analysis of TB and Non TB images, comparisons have been performed with other classification techniques. Table 2 includes proportional scrutiny in the form of recall, precision, AUC, CA and F1-score for different machine learning algorithms. As we can see that Neural Networks provides most efficient results as compared to other techniques. Although results of all techniques are very close to each other however it is clearly visible from table that neural network outperform other techniques in each parameter. Average of all technique show case that deviation of values in different parameters are stable in nature hence framework is effective. Figure 5 and 6 demonstrate stability of framework and effectiveness of neural network. It is clearly visible that neural network outperforms other techniques.

Table 2. Proportional scrutiny of different models.

Model	AUC	CA	F1-score	Precision	Recall
Neural Networks	0.894	0.811	0.811	0.811	0.811
SVM	0.870	0.795	0.790	0.793	0.796
kNN	0.853	0.796	0.795	0.803	0.792
Random Forest	0.851	0.782	0.782	0.782	0.782
Average	0.867	0.796	0.795	0.797	0.795

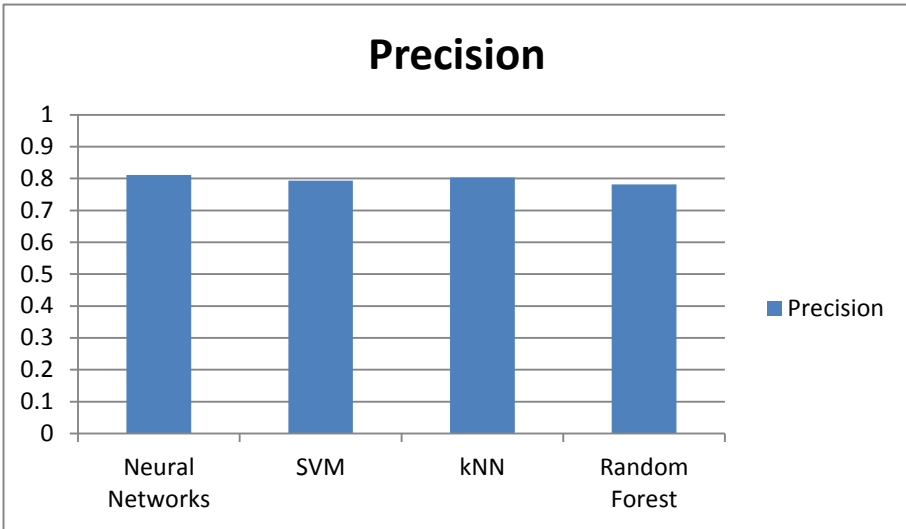


Fig. 5. Precision Statistics of various machine learning algorithms over TB dataset

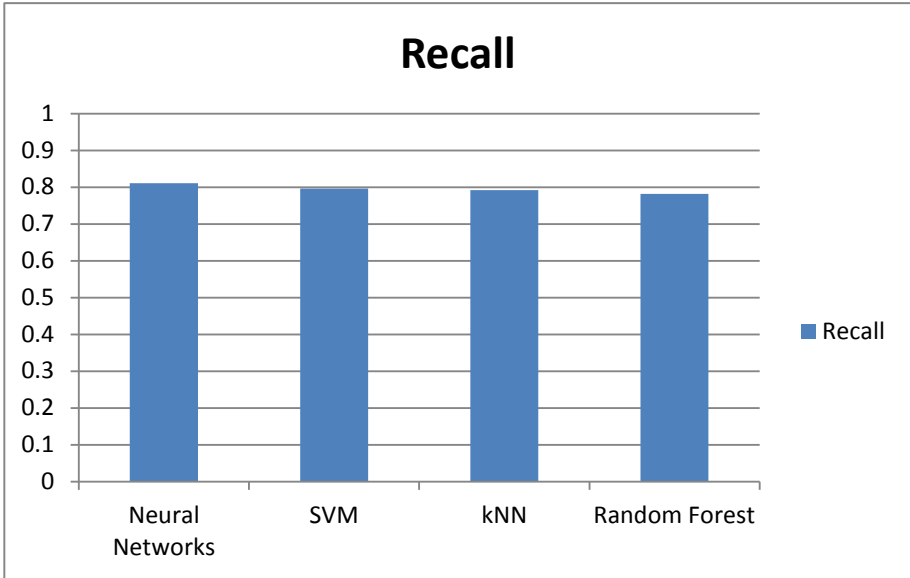


Fig. 6. Recall Statistics of various machine learning algorithms over TB dataset

For more insight of result Table 3 is presented. Table 3 includes true positive, true negative, false negative, false positive values for both the classes. The overall accuracy for detecting normal people is 88.7% and detecting abnormal people is 72.2%. As we can see in table (row 1 with TB) the false positive score is 28, which means that framework has detected x-ray of people having TB as normal people which is not correct similarly row 2 also provide valuable information about effectiveness and limitations of framework. This limitation of efficiency can be corrected by increasing the dataset.

Table 3. Statistics representing the True Positive, True Negative, False Positive, False Negative values along with accuracy and average accuracy.

Class	True Positive	True Negative	False Positive	False Negative	Accuracy
With TB	87	73	28	11	88.7%
Without TB	73	87	11	28	72.2%
Average accuracy					80.45%

5 Conclusion

In the current study we have proposed a novel framework to classify chest X-ray image data into two categories i.e. with TB and without TB. Various classification algorithms have been compared to showcase that neural network is most suited for identification of TB from given X-ray image. The framework is capable to classify all the studied classes with high accuracy. The accuracy can be increased in future by adding more images to the TB class and removing unwanted objects from the chest X-ray images. It is the first study where the investigation between the tuberculosis is performed which is differentiating the X-Rays according to the presence and absence of TB. The discrimination between the pulmonary diseases, pneumonia, tuberculosis, lung cancer is a complex task and our work can be used as a base for this.

References

1. World Health Organization. Global tuberculosis report 2017: World Health Organization (2017). https://www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf
2. Zumla, A., George, A., Sharma, V., Herbert, N.: WHO's 2013 global report on tuberculosis: successes, threats, and opportunities. *Lancet* **382**(9907), 1765–1767 (2013)
3. Franquet, T.: Imaging of pneumonia: trends and algorithms. *Eur. Respir. J.* **18**(1), 196–208 (2001)
4. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106 (2017)
5. Cupples, J.B., Blackie, S.P.: Granulomatous *Pneumocystis carinii* pneumonia mimicking tuberculosis. *Arch. Pathol. Lab. Med.* **113**, 1281–1284 (1989)
6. Barnes, P.F., Steele, M.A., Young, S.M., Vachon, L.A.: Tuberculosis in patients with human immunodeficiency virus infection: how often does it mimic *Pneumocystis carinii* pneumonia. *Chest* **102**(2), 428–432 (1992)
7. Pinto, L.M., Shah, A.C., Shah, K.D., Udwardia, Z.F.: Pulmonary tuberculosis masquerading as community acquired pneumonia. *Respir. Med. CME* **4**(3), 138–140 (2011)
8. Singh, K., Hyatali, S., Giddings, S., Singh, K., Bhagwandass, N.: Miliary tuberculosis presenting with ards and shock: a case report and challenges in current management and diagnosis. *Case reports in critical care* (2017)
9. Rohmah, R.N., Susanto, A., Soesanti, I., Tjokronagoro, M.: Computer Aided Diagnosis for lung tuberculosis identification based on thoracic X-ray. In: 2013 International Conference Information Technology and Electrical Engineering (ICITEE), pp. 73–78 (2013)
10. Poornimadevi, C.S., Sulochana, H.: Automatic detection of pulmonary tuberculosis using image processing techniques. In: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 798–802, March 2016
11. Fatima, S., Shah, S.I.A., Samad, M.Z.: Automated tuberculosis detection and analysis using CXR's images. *Int. J. Comput. Electr. Eng.* **10**(4), 284–290 (2018)
12. Parveen, N., Sathik, M.M.: Detection of pneumonia in chest X-ray images. *J. X-ray Sci. Technol.* **19**(4), 423–428 (2011)
13. Sharma, A., Raju, D., Ranjan, S.: Detection of pneumonia clouds in chest X-ray using image processing approach. In: Nirma University International Conference on Engineering (NUICONE), pp. 1–4, 23 November 2017

14. Stephen, O., Sain, M., Maduh, U.J., Jeong, D.U.: An efficient deep learning approach to pneumonia classification in healthcare. *J. Healthcare Eng.* (2019)
15. Fatima, M., Pasha, M.: Survey of machine learning algorithms for disease diagnostic. *J. Intell. Learn. Syst. Appl.* **9**(01), 1 (2017)
16. Kim, J.I.N.H.O., Kim, B.S., Savarese, S.: Comparing image classification methods: K-nearest-neighbor and support-vector-machines. In: *Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*, vol. 1001, p. 48109–2122 (2002)
17. Bose, C., Sharma, D., Tripathi, V., Singh, A., Pandey, P.: A framework for analyzing the exercise and athletic activities. In: *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 121–124, March 2019
18. Jaeger, S., et al.: Automatic tuberculosis screening using chest radiographs. *IEEE Trans. Med. Imaging* **33**(2), 233–245 (2014). <https://doi.org/10.1109/TMI.2013.2284099>. PMID: 24108713