International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 2495–2502

A Review on Respiratory Disease Detection Using Machine Learning

Harshitha K[1*], Rakesh K R[2], Pavan Kumar S P[3] and Anusha K S[4]
[1,2,3,4]Vidyavardhaka College of Engineering
[1] harshitha.k@vvce.ac.in, [2]rakeshkr@vvce.ac.in, [3]pavanalina@vvce.ac.in, [4]anushaks@vvce.ac.in

Abstract

One of leading causes of death is lung diseases. It refers to disorders that affect the lungs, such as Asthma, lung cancer, pneumonia, influenza, tuberculosis, etc. It is necessary to predict them before crucial stage as these diseases could lead respiratory failure. Recent recognition of machine learning techniques is enabling tools for amassing real-time data and analyzing it for precise prediction of respiratory diseases. In this paper, prediction and diagnosis of lung diseases – Asthma, lung cancer, pneumonia and tuberculosis with several machine learning algorithms and methods are discussed. Numerous ways are found in this literature for detection of diseases. The methodology on datasets from various kinds of sources like medical records, CX-Rays is analyzed in the survey.

Keywords: *lung disease, Asthma, lung cancer, pneumonia, tuberculosis, CX-Rays.*

1. Introduction

The environment where we live is drastically changing, which leads to the compulsion of maintaining the human health. Lung diseases are the precarious reason of death [1]. The risk of this disease is immense in under developed and developing countries. WHO has given a report on premature deaths occurring from household air pollution related diseases stating that around 4 million deaths occur annually. Respiratory diseases refer to the disorders of lung like Asthma, Lung cancer, pneumonia, Tuberculosis, COPD, etc.

Machine learning approaches are extensively used in medical zones. It aids in identifying the hidden patterns in clinical data. Detection of lung disease is one of essential problems and sundry researchers are evolving intelligence in improving the accuracy of predicting machines [2]. Research papers on this provide doctors and further researchers route for detecting lung diseases in early stage with support of deep learning methodology. The early stage prediction can reduced the death rate of elderly citizens of a country. A huge number of chest X-rays, patient history are used as the dataset. Many researchers have done study on Machine Learning Technologies to relate it to disease diagnose. This paper gives a survey on predicting Asthma, Lung Cancer, pneumonia and Tuberculosis[3]

2. Literature Survey

Wasif Akbar, Wei-ping Wu, Muhanmmad Faheem, Muhammad Asim saleem, Noorbakhsh Amiri Golilarz and Amin Ul Haq[4] gave an overview on classifiers for asthma disease estimation. The acuteness of the disease was divided into 4 classes: - Intermittent asthma, mild asthma, moderate asthma and severe asthma. The dataset used

ISSN: 2233-7857IJFGCN

here was taken from Pakistan hospital which has 16000 samples of asthmatic patient's details along with patients suffering from other respiratory diseases. WEKA tool was used for experiment in which a Naïve Bayesian technique with Bayes Theorem was applied. For classification, decision tree code which starts from root node and travels to leaf node was used. To cartel bagging and random selection of figures, random forest was used. This model gave an accuracy of 98.75%.

Pooja M R and Pushpalatha M P[5] proposed a predictive model using Asthma severity indicators. Dataset was collected by University of Innsbruck, Austria. The dataset was collected by surveying the operation of lungs and respiratory tracks of the school children. The social background of the children was also considered. Parameters like presence of allergy, cough, cold, fever, etc. were recorded. The classifiers that were used were K- nearest neighbour and SVM for making study of Asthma parameters. Feature selection was accomplished by ranking features by importance. Both the methods showed a good performance in terms of all different parameters.

Achuth Rao M V, Kausthubha N K, Shivani Yadav, Dipanjan Gope, Uma Maheshwari Krishnaswamy and Prasanta Kumar Ghosh[6] submitted their work on Asthma Severity monitoring based on spirometry evaluations. With the aid of spirometry, cough and wheeze audio signals were measured. The measurement parameters of device were forced vital capacity(FVC) and forced expiratory volume per second(FEV1). For the prediction, support vector regression was deployed. The total data gathered was from 16 persons out of which 12 persons were asthmatic patients. Threshold for FEV1% was set. Feature extraction involved discrete cosine transformation to produce low dimensional sub-bands. The model acquired an accuracy of 77.77%.

Julie L Harvey and Sathish A P Kumar[7] presented a system for predicting development of Asthma using various algorithms which includes linear regression, decision tree, KNN, Naïve Bayes and random forest. The dataset expended in paper was 2016 NSCH. This dataset has 50212 number of observations recorded from children. KNN predicted that 12800 children did not have asthma. The first algorithm that was used was logistic regression by which a correlation plot was determined. Next, Naïve Bayes classification was used which gave an accuracy of 82.7%. Highest accuracy of 90.9% was obtained by random forest classifier.

Joseph Finkelstein and In cheol Jeong[8] introduced a tele monitoring system for the prediction of asthma. For the modelling of this system CART was used and tele monitoring was home-based. Initially the dataset consisted of 6762 record and after the cleaning of dataset, 3470 records were present. Binary classification systems were constructed using seven tree growing process of CART. The prediction period after the event was recorded. The best split was chosen at each node. The value with the maximum capacity of differentiating between two outcomes was selected by the algorithm. The accuracy was 80.9% from this model. The model can be efficiently used in prolonged health disorder.

D Jayaraj and S Sathiamoorthy[9] presented a model for predicting lung cancer using Random forest classifier. CT scans are used for diagnosing the cancer. Computer tools are used for image processing purpose. The dataset used here is taken from LIDC. First the CT images are given as input to the system and after the pre-processing by Gaussian filter; segmentation of images was done by using watershed segmentation algorithm. Later, collection of significant features is done to which the random model classifier is applied. For simulation purpose MATLABR 2014a was used. This presented method gave an accuracy of 89.9%.

ISSN: 2233-7857IJFGCN

Xueyan Mei[10] proposed a technique on prophesying non-small cell lung cancer using random forests and decision tree algorithm. The dataset is textual based and is collected from several hospitals in China. The features included duration of operation, drain age days, length of operation, status of overall survival, chylothorax, etc. This data set includes for about 5123 NSCLC records. After inputting the data, ReliefF algorithm was used to optimize the features. Next, the pulmonary nodule was forecasted by using the classifier. Among Decision tree and Random forest algorithm, random forest showed highest accuracy of 80.25% compared to decision tree with 78.16% accuracy.

DendiGayathri Reddy, Emmidi Naga Hemanth Kumar, Desireddy Lohith Sai Charan Reddy and Monika P[11] proposed a system to predict multiple levels of lung cancer by ensemble method. The dataset was taken from Data World Source, which has 1000 data records. The paper explains about predicting various carcinoma stages using ML concepts. The paper put forth a method which uses an amalgamation of three algorithms – KNN, neural networks and decision trees with bagging. KNN is a data sensitive by nature which uses Euclidean distance. It understands the data. To correct the errors, backpropagation of neural networks is used. Next, CART algorithm was used for classification purpose. To reduce the variance of cost, bagging was used. This integrated model gave an accuracy of 98%.

Janee Alam, Sabrina Alam and Alamgir Hossan [12] submitted their work on prediction of multistage lung cancer detection with SVM classifier. The dataset used here contained 500 lung CT images. The software tool used here was MATLAB to process the image. If the input image contains no affected cell, then probability of the disease is diagnosed. For the purpose of image enhancement, masking is done. For gaining better resolution of image, watershed transform for segmentation is applied. GLCM technique is used for feature extraction. Next, the SVM classifier is applied. This works gave 97% of detection accuracy and 87% of prediction accuracy.

Moataz M Abdelwahab and Shimaa A Abdelrahman[13] introduced a system to predict lung cancer based on genetic mutations. The gene structure is signified by a layer image. The dataset was taken from COSMIC database and sequences of non-muted genes were taken from NCBI database for validation purpose. The genes included EGFR, TP53 and KRAS. The genes are denoted by A, T, C or G and for each of them a pixel value is assigned. These numerical sequences are converted to four layer image representation and are divided into sub-sequences. To represent this image in two dimensions, 2DPCA algorithm is used. Covariance matrix is built from these layers and dominant eigenvectors are chosen. MATLAB R2014 was used to carry out the procedure. Both PCA and 2DPCA were used. 2DPCA gave a highest accuracy of 98.55%.

Mohammed Aledhari, Shelby Joji, Mohamed Hefeida and Fahad Saeed[14] proposed their work on diagnosis of pneumonia using chest radiography. The dataset was taken form ImageNet, NIH which contains high- resolution images. The dataset was pre-processed to 1024*1024 pixels and also the values were cleaned. Next, the dataset was resampled using PIL. The output of this was feed into CNN model that consisted of three layers – VGG16, ResNet-50 and Inception v3. The modelling was performed by Keras with TensorFlow. The model gave an average accuracy of 68%.

Sheikh Rafiul Islam, Santi P Maity, Ajoy Kumar Ray and Mrinal Mandal[15] submitted work on detection of pneumonia based on compressed sensing images. The dataset was obtained from Kaggle and contains 5863 images of CX-rays. Compressed sensing is reconstructed using using deep learning framework. The classification model

ISSN: 2233-7857IJFGCN

uses three fully connected layers which uses ReLU activation function which uses zero-gradient in its domain. In this works, the measurements are projected through Inverse DCT. The model gave an accuracy of 99.80%.

Sing-Ling Jhuo, Mi-Tren Hsieh , Ting- Chien Weng, Mei-Juan Chen, Chieh-Ming Yang and Chia-Hung Yeh[16] proposed a system on prediction of influenza associated with pneumonia. The dataset was taken from Environmental Protection Administration in Taiwan and also from Taiwan Centre of disease control. The model uses forward and backward back propagation algorithm. Multilayer Perceptron algorithm was applied on five different age group patients. It was conclude that based on the climatic factor, elderly people are more likely to get affected by flu. The model gave an accuracy of 77.54% for overall population.

Ferani E Zulvia, R J Kuo and E Roflin[17] presented a system on initial screening method on Tuberculosis. The dataset was collected from five different hospitals in Palembang City, Indonesia. It consists of 374 records with nine features. Out of 374 records, 187 are positive for TB and rest 187 are non-TB. Feature extraction was carried out by PCA. Two-step data transformation was used to convert data into percentage values. Classification purpose included three different algorithms which included-MOGESVM, MOPSOSVM, MODESVM. Out of these three algorithms, MOGESVM was more stable than the other algorithms.

Adnan Fojnica, Ahmed Osmanovic and Almir Badjevic[18] proposed a dynamic model on Tuberculosis based on ANN. The model consisted of 1000 records. The epidemiology is divide into six classes namely: - susceptible, sensitive to drugs, resistant to drug, Infectious –sensitive to drug, treated individuals and infected but not infectious. The training algorithm used was LMA. Training function TRAINLM was used to update the weight and predict values along with LMA. The testing was carried on 1400 patients. The model gave an accuracy of 99.24%.

Sivaramakrishnan Rajaraman and Sameer K Antani[19] proposed a deep learning model to detect Tuberculosis. The dataset was collected from RSNA pneumonia, paediatric pneumonia and Indiana and 80% of it were classified as training set and 20% to test datasets. CNN was deployed for the model. Each CNN block had a normalization layer, non-linear activation and dropout layers. Zero padding was applied to the CNN. There were 64 convolutional filters and the number was increased by factor of two. To reduce variance ensemble learning was used. The accuracy achieved was 94.1%.

The comparison of accuracy of different techniques on various data kinds is shown in Table 1, Table 2 and Table 3.

TABLE 1. A Summary of Accuracies of Lung Diseases Based on Audio Dataset

ISSN: 2233-7857IJFGCN

SL No	Research Paper Title	Dataset Used	Technique Used	Accuracy
1.	Automatic Prediction of Spirometry Readings from Cough and Wheeze for Monitoring of Asthma Severity	Data Collected from 16 patients	Support Vector Machine	77.77%

TABLE 2. A Comparison of Accuracies of Lung Diseases Based on Image Dataset

SL No	Research Paper Title	Dataset Used	Technique Used	Accuracy
1.	Random Forest based Classification Model for Lung Cancer Prediction on computer Tomography Images	CT Scans from LIDC	Watershed segmentation and Algorithm Random Forest	89.9%
2.	Multi-Stage Lung Cancer Detection and Prediction Using Multi-Class SVM Classifier	500 CT images	Watershed segmentation and Support Vector Machine	97%(detection accuracy) and 87%(prediction accuracy)
3.	Four Layer Image Representation for Prediction of Lung Cancer Genetic Mutations Based on 2DPCA	Sequence of non-muted genes from COSMIC dataset	2DPCA algorithm	98.55%
4.	Optimized CNN-based Diagnosis System to Detect the Pneumonia from Chest Radiographs	Chest radiography from Imagenet, NHI	CNN, Keras, Tensorflow	68%
5.	Automatic Detection of Pneumonia on Compressed Sensing Images using Deep Learning	5863 images of x-rays, Kaggle	ReLU activation function	99.80%
6.	Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs	RSNA pneumonia, paediatric pneumonia	CNN	94.1%

TABLE 3. A Comparison of Accuracies of Lung Diseases Based on Medical History

SL No	Research Paper Title	Dataset	Technique Used	Accuracy

ISSN: 2233-7857IJFGCN

		Used		
1.	Machine Learning Classifiers for Asthma Disease prediction: A Practical Illustration	16000 samples of Patient's Data from Hospital, Pakistan	Naive Bayes Theorem	98.75%
2.	A Predictive Model for the effective Prognosis of Asthma using Asthma Severity Indicators.	Data collected from school children, Austria	KNN and SVM	KNN-precision 1.0, Recall 0.99 , F-measure 0.99 and SVM-precision 0.90, recall 1.0, F-measure 0.94.
3.	Machine Learning for Predicting Development of Asthma in Children	50212 observatio--ns from 2016 NSCH	Naïve Bayes and Random Forest	Naïve Bayes-82.7% and Random Forest-90.9%
4.	Using CART for Advanced Prediction of Asthma Based on Telemonitoring Data.	6762 records	CART algorithm	80.9%
5.	Predicting Five-year Overall Survival in Patients with Non-small Cell Lung Cancer by ReliefF Algorithm and Random Forests	5123 NSCLC data of Patients from Hospital, China	Decision Tree and Random Forest	Random Forest 80.25% And Decision Tree 78.166%
6.	Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data Using Ensemble Method	1000 data records, Data World Source	Mixture of KNN, Neural Network and Decision Tree and CART algorithm.	98%
7.	Trend Prediction of Influenza and the Associated Pneumonia in Taiwan using machine Learning	Environme-nt Protection Administra-tion and Taiwan centre of disease.	Multilayer Perceptron Algorithm	77.54%
8.	An Initial Screening Method for Tuberculosis Diseases Using a Multi-objective Gradient Evolution- based Support Vector Machine and C5.0 Decision Tree	Collected from 5 hospitals, Palembay	PCA, MOGESVM,MOPSOSVM,MODESVM	Not Mentioned
9.	Dynamical Model of	Data collected	LMA	99.24%

ISSN: 2233-7857IJFGCN

	Tuberculosis-Multiple Strain Prediction based on Artificial Neural Network	from Hospital		

Conclusion

Machine learning plays a major role in analysing the medical data. This paper illustrates how machine learning can be used on various kinds of medical data and provides an enhanced technique in early prediction of lung diseases like Asthma, lung cancer, pneumonia and Tuberculosis. The work of several researchers and the accuracy given by the various researchers is discussed. This paper also provides a scope for future work in enhancing the accuracy of the disease predicted.

References

[1] Anuradha D Gunasinghe, Achala C Aponso and Harsha Thirimanna, "Early prediction of Lung Diseases, Proceedings of International Conference for Convergence in Technology", 2019.

[2] Ishan Sen, Ikbal Hossain, Md. Faisal Hossan Shakib, Md. Asaduzzaman Imran and Faiz AL Faisal, " In depth Analysis of Lung Disease Prediction using Machine Learning Algorithms", Springer, 2020.

[3] Subrato Bharathi, Prajoy Podder and M Rubaiyat Hossain Mondal,"Hybrid deep learning for detecting lung diseases from X-ray images", Elsevier, 2020.

[4] Wasif Akbar, Wei-ping Wu, Muhanmmad Faheem, Muhammad Asim saleem, Noorbakhsh Amiri Golilarz and Amin Ul Haq, "Machine Learning Classifiers for Asthma Disease prediction: A Practical Illustration", Proceedings of IEEE, 2019.

[5] Pooja M R and Pushpalatha M P, "A Predictive Model for the effective Prognosis of Asthma using Asthma Severity Indicators", Proceedings of ICCCI, 2017.

[6] Achuth Rao M V, Kausthubha N K, Shivani Yadav, Dipanjan Gope, Uma Maheshwari Krishnaswamy and Prasanta Kumar Ghosh, "Automatic Prediction of Spirometry Readings from Cough and Wheeze for Monitoring of Asthma Severity", Proceedings of EUSIPCO, 2017.

[7] Julie L Harvey and Sathish A P Kumar, "Machine Learning for Predicting Development of Asthma in Children", Proceedings of SSCI, 2019.

[8] Joseph Finkelstein and In Cheol Jeong, "Using CART for Advanced Prediction of Asthma Based on Telemonitoring Data", Proceedings of IEEE,2016.

[9] D Jayaraj and S Sathiamoorthy, " Random Forest based Classification Model for Lung Cancer Prediction on computer Tomography Images", Proceedings of ICSSIT, 2019.

[10] Xueyan Mei, "Predicting Five-year Overall Survival in Patients with Non-small Cell Lung Cancer by ReliefF Algorithm and Random Forests", Proceedings of IEEE,2017.

[11] DendiGayathri Reddy, Emmidi Naga Hemanth Kumar, Desireddy Lohith Sai Charan Reddy and Monika, " Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data Using Ensemble Method ",Proceedings of ICAIT, 2019.

[12] Janee Alam, Sabrina Alam and Alamgir Hossan, " Multi-Stage Lung Cancer Detection and Prediction Using Multi-Class SVM Classifier", Proceedings of IEEE.

[13] Moataz M Abdelwahab and Shimaa A Abdelrahman, " Four Layer OImage Representation for Prediction of Lung Cancer Genetic Mutations Based on 2DPCA", Proceedings of IEEE,2017.

ISSN: 2233-7857IJFGCN

[14] Mohammed Aledhari, Shelby Joji, Mohamed Hefeida and Fahad Saeed, "Optimized CNN-based Diagnosis System to Detect the Pneumonia from Chest Radiographs", Proceedings of IEEE,2019.

[15] Sheikh Rafiul Islam, Santi P Maity, Ajoy Kumar Ray and Mrinal Mandal, "Automatic Detection of Pneumonia on Compressed Sensing Images using Deep Learning", Proceedings of IEEE,2019.

[16] Sing-Ling Jhuo, Mi-Tren Hsieh , Ting- Chien Weng, Mei-Juan Chen, Chieh-Ming Yang and Chia-Hung Yeh, "Trend Prediction of Influenza and the Associated Pneumonia in Taiwan using machine Learning", Proceedings of ISPACS, 2019.

ISSN: 2233-7857IJFGCN