

Building a respiratory diseases recognition system Using Data mining techniques

Ahmed Saleh Abdu¹, Mostafa-Sami M. Mostafa², Aya Sedky Adly²

¹ High Insuitute of Engineering and Technology, New Damietta, Egypt

² Faculty of Computers and Artificial Intelligence, Helwan University, Egypt

Abstract— Respiratory diseases are among the most common causes of morbidity and mortality worldwide. Computer-assisted learning tools have been shown to promote self-directed learning and enhance problem-solving and predicting skills. However, few have been developed in the area of respiratory medicine. Thus, we aimed in this study to build a framework to predict the respiratory system diseases by applying Data Mining Techniques (DMT) on clinical information. The framework composed of three main phases which are preprocessing phase, data mining phase and evaluation phase to evaluate the built prediction model. This Approach statistics shows acceptable results.

Index Terms—Data Mining Techniques, Decision Tree, Prediction Model, Respiratory.

I. INTRODUCTION

Respiratory auscultation performed with a conventional stethoscope is an assessment method used by many health professionals to evaluate and monitor patients with respiratory diseases [10]. The stethoscope is used by clinicians to aid in the diagnosis of respiratory disorders; however, application of the stethoscope in research studies has been limited due to the inherent inter-observer variability and subjectivity in the interpretation of lung sounds. Computerized analysis of recorded lung sounds may offer a systematic approach to the diagnosis of different respiratory conditions via automated classification of acoustic patterns [1].

Data mining is not all about the used tools or database software. Data mining itself depends on building an appropriate data model and structure which can be used to process, identify, and build the required medical and clinical information. In spite of the source data form and structure, structuring and forming the clinical information in a layout which tolerates the data mining procedures to run in as an actual model as possible.

Therefore, this study is targeted to build a framework to predict the respiratory system diseases by applying Data Mining Techniques (DMT) on clinical information.

The remainder of this paper is structured as follows. First, we pondered the related work in Section II. This is followed by a portrayal of the clinical data and the parts of our framework in Section III. The experimental outcomes are argued in Section IV. We conclude our paper in Section V and give a perspective to the future work.

II. RELATED WORK

In [1], Arati Gurung et al. sought to perform a systematic review and meta-analysis of studies implementing computerized lung sound analysis (CLSA) to aid in the detection of abnormal lung sounds for specific respiratory disorders. They searched for articles on CLSA in MEDLINE, EMBASE, Cochrane Library and ISI Web of Knowledge. They conducted a meta-analysis to estimate the sensitivity and specificity of CLSA for the detection of abnormal lung sounds. They selected eight studies for review. Most studies employed either electret microphones or piezoelectric sensors for auscultation, and Fourier Transform and Neural Network algorithms for analysis and automated classification of lung sounds.

In [2], Sandra Reichert et al. described state of the art, scientific publications and ongoing research related to the methods of analysis of respiratory sounds using PubMed and personal experience. The study included a description of the various techniques that are being used to collect auscultation sounds, a physical description of known pathologic sounds for which automatic detection tools were developed. Modern tools are based on artificial intelligence and its techniques; such as artificial neural networks, fuzzy systems, and genetic algorithms.

In [3], they developed an audio signal processing algorithm that detects rales (gurgling noises that are a distinct symptom of common respiratory diseases in poultry). They derived features from the audio by calculating mel frequency cepstral coefficients (MFCCs), clustering the MFCC vectors, and examining the distribution of cluster indices over a window of time. The features were classified with a C4.5 decision tree. Their training data consisted of eight minutes of manually labeled audio selected from 25 days of continuous recording from a controlled study. The experiment group was challenged with the infectious bronchitis virus and became sick, while the control group remained healthy. They tested the algorithm on the entire dataset and obtained results that match the course of the disease.

In [4], they developed a non-invasive method of classifying respiratory sounds that are recorded by an electronic stethoscope and the audio recording software that uses various machine learning algorithms. In order to store respiratory sounds on a computer, they developed a cost-effective and easy-to-use electronic stethoscope that can be used with any device. Using this device, they recorded 17,930 lung sounds from 1630 subjects. They employed two types of machine learning algorithms; mel frequency cepstral coefficient (MFCC) features in a support vector machine (SVM) and spectrogram images in the convolutional neural network (CNN). Since using MFCC features with a SVM algorithm is a generally accepted classification method for audio, they utilized its results to benchmark the CNN algorithm.

In [5], Jose Semedo et al. aimed to develop and evaluate an open-source computer-assisted learning application for respiratory sound recording, analysis and interpretation, the Computerized Lungs Auscultation – Sound Software (CLASS). Participants highlighted its utility and great potential to be used in academic and clinical environments, and they also identified some drawbacks. Despite of the drawbacks addressed in the usability assessment, the utility has great potential to be used both in academic and clinical environments.

In [6], Cátia Pinho et al. proposed a study aimed to develop a usable interface to collect and organize respiratory-related data in a single multimedia database. The usability of the user-centred interface was assessed in a pilot study and in an evaluation session. However, some drawbacks were identified, such as a certain difficulty to intuitively navigate in the great amount of the available information, which will inform future developments.

In [7], the aim of this study was to implement a pattern recognition system to classify lung sounds. They used the cross-validation procedure to analyze the classifiers performance and the Tukey's Honestly Significant Difference criterion to compare the results. Their results showed that the Genetic Algorithms outperformed the Fisher's Discriminant Ratio for feature selection. Moreover, each lung class had a different signature pattern according to their cumulants showing that HOS is a promising feature extraction tool for lung sounds. Besides, the proposed divide-and-conquer approach can accurately classify different types of lung sounds. The proposed approach achieved good results even using only one feature extraction tool (higher-order statistics). Additionally, the implementation of the proposed classifier in an embedded system is feasible.

In [8], this research highlighted the importance of machine learning in computer-based lung sound analysis. A brief description of the types of lung sounds and their characteristics is provided. They examined specific lung sounds/disorders, the number of subjects, the signal processing and classification methods and the outcome of the analyses of lung sounds using machine learning methods that have been performed by

previous researchers. A brief description on the previous works is thus included.

In [9], they presented a novel low-complexity wheeze detection method based on frequency contour tracking for automatic wheeze detection. Two hardware friendly variants of the algorithm have also been proposed. Applying the proposed feature extraction algorithm we achieved very high classification accuracy ($> 99\%$) at considerably low computational complexity ($3\times-6\times$) compared to earlier methods and the power consumption of the proposed method is shown to be significantly less ($70\times-100\times$) compared to 'record and transmit' strategy in wearable devices. The algorithm produces highly distinguishable features for wheeze signals using nominal computational overhead. High classification accuracies were obtained for both software and hardware simulations. The method (coupled with a suitable classifier) can be used for low power implementation of both on-chip wheeze detection and selective transmission strategies.

In [10] Ana Oliveira and Alda Marques made a systematic review aimed to characterise respiratory sounds of healthy people. The Web of knowledge, MEDLINE, EMBASE and SCOPUS databases were searched and studies using computerised analyses to detect/characterise respiratory sounds in healthy people were included. Data were extracted using a structured table-format. The results showed that 13 investigated normal respiratory sounds (frequency, intensity and amplitude) and 3 adventitious respiratory sounds (crackles and wheezes). The highest sound frequencies were observed at the trachea (inspiration: $447e1323$ Hz; Expiration: $206 e540$ Hz). Women ($444e999$ Hz) and infants ($250e400$ Hz) presented the highest frequencies at maximum power. Inspiratory sounds were more intense at the left posterior lower lobe ($5.7e76.6$ dB) and expiratory sounds at the trachea ($45.4e85.1$ dB). Nevertheless, studies establishing direct comparisons between inspiratory and expiratory sounds showed that inspiratory sounds presented the highest intensities ($p < 0.001$). Amplitude was higher at the left upper anterior chest (1.7 ± 0.8 V) and lower at the right posterior lower lobe (1.2 ± 0.7 V). Crackles were the adventitious respiratory sound most frequently reported. Conclusions: Respiratory sounds show different acoustic properties depending on subjects' characteristics, subjects' position, respiratory flow and place of recording.

In [11] M F Syahptura, et al. designed a filter called a wavelet transform based filter. The filter that is designed in this study is using Daubechies wavelet with four wavelet transform coefficients. Based on the testing of the ten types of breath sounds data, the data is obtained in the largest SNRdB bronchial for 74.3685 decibels. The experimental results showed that the biggest SNRdB value is 74.368 decibel which is obtained from the reduction of bronchial. If the threshold value is too large, the noise cannot be reduced from the breath sounds signal. Conversely, if the threshold value is too small, breath sounds signal will lose a lot of important information.

In [12] Nandini Sengupta, et al. studied short-term spectral characteristics of lung sounds to characterize the lung sounds for the identification of associated diseases. Motivated by the success of cepstral features in speech signal classification, they evaluated five different cepstral features to recognize three types of lung sounds: normal, wheeze and crackle. Subsequently for fast and efficient classification, they proposed a new feature set computed from the statistical properties of cepstral coefficients. Experiments are conducted on a dataset of 30 subjects using the artificial neural network (ANN) as a classifier. Results showed that the statistical features extracted from mel-frequency cepstral coefficients (MFCCs) of lung sounds outperform commonly used wavelet-based features as well as standard cepstral coefficients including MFCCs. Further, they experimentally optimized different control parameters of the proposed feature extraction algorithm. Finally, they evaluated the features for noisy lung sound recognition. They have found that their investigated features are more robust than existing features and showed better recognition accuracy even in low signal-to-noise ratios (SNRs).

In [13] Plamen Bokov, et al. developed a wheezing recognition algorithm from recorded respiratory sounds with a Smartphone placed near the mouth. A total amount of 186 recordings were obtained in a pediatric emergency department, mostly in toddlers (mean age 20 months). After exclusion of recordings with artefacts and those with a single clinical operator auscultation, 95 recordings with the agreement of two operators on auscultation diagnosis (27 with wheezing and 68 without) were subjected to a two phase algorithm (signal analysis and pattern classifier using machine learning algorithms) to classify records. The results showed that the best performance (71.4% sensitivity and 88.9% specificity) was observed with a Support Vector Machine-based algorithm. They further tested the algorithm over a set of 39 recordings having a single operator and found a fair agreement ($\kappa = 0.28$, CI95% [0.12, 0.45]) between the algorithm and the operator. The main advantage of such an algorithm is its use in contact-free sound recording, thus valuable in the pediatric population.

In [14] F. Jin, et al. proposed a new set of features based on temporal characteristics of filtered narrowband signal to classify respiratory sounds (RSs) into normal and continuous adventitious types. RS signals were first decomposed in the time-frequency domain and features were extracted over selected frequency bins containing distinct signal characteristics based on auto-regressive averaging, the recursively measured instantaneous kurtosis, and the sample entropy histograms distortion. The presented features were compared with existing features using a modified clustering index with different distance metrics. Mean classification accuracies of 97.7% and 98.8% for inspiratory and expiratory segments respectively have been achieved using Support Vector Machine on real recordings.

In [15] Yusuf A. Amrulloh, et al. developed a method to

automatically identify cough segments from the pediatric sound recordings. Their method is based on extracting mathematical features such as non-Gaussianity, Shannon entropy, and cepstral coefficients to describe cough characteristics. These features were then used to train an artificial neural network to detect coughs segment in the sound recordings. They worked on a prospective data set of 14 subjects (sound recording length 840 min), proposed method achieved sensitivity, specificity, and Cohen's Kappa of 93%, 98%, and 0.65, respectively. These results indicated that the proposed method has the potential to be developed as an automated pediatric cough counting device as well as the front-end of a cough analysis system.

In [16] Roberto Marani and Anna G. Perri proposed a new pressure sensor-based electronic device for the analysis of lung sounds. The device allows the effective auscultation, the accurate processing and the detailed visualization (temporal and frequency graphs) of any lung sound. It is suitable for the continuous real-time monitoring of breathing functions, resulting very useful to diagnose respiratory pathologies. It provides medical specialists with a totally noninvasive high-engineering device able to detect and analyze the widest number of data for the monitoring of the respiratory system by the simple recording and evaluation of lung sounds being a substantiated correlation between lung sounds and diseases. The innovative solutions allowed a high miniaturization level, automatization and simplicity of use, since they have employed last-generation programmable integrated circuits.

In [17] Roneel V. Sharan and Tom J. Moir carried out a review of some important contributions in ASR techniques, mainly over the last one and a half decades. Similar to speech recognition systems, the robustness of an ASR system largely depends on the choice of feature(s) and classifier(s). They took a wider perspective in providing an overview of the features and classifiers used in ASR systems starting from early works in content-based audio classification to more recent developments in applications such as sound event recognition, audio surveillance, and environmental sound recognition. They also reviewed techniques that have been utilized in noise robust sound recognition systems and feature optimization methods.

III. MATERIAL AND METHODS

A. Clinical Data

Data has been collected from multiple sources. As well as the selection of real data to reflect the extent of the problem within our community. The collected data was pulmonary function tests (PFT) for approximately 250 patients. The most important elements of the reports were described below

TABLE 1: A DISCRIPTION OF CLINICAL DATA

Appreivation	Stands for	Notes
VC	Vital Capacity	Maximum volume of air that can be expired by forced expiration after forced inspiration
FEV1	Forced Expiratory Volume	The volume of air that is forcefully exhaled in 1 second
FVC	Forced Vital Capacity	Volume of air that can be maximally forced exhaled
FEV1/FVC	Forced Expiratory Volume/ Forced Vital Capacity	Ratio of Forced Expiratory Volume in 1 second to Forced Vital Capacity
PEF	Peak Expiratory Flow	>200 l/m --- Moderate asthma <200 l/m --- Severe asthma <200 l/m or <33% or unable to do --- life threating
MEF	Mid Expiratory Flow rate	MEF25, MEF50 and MEF75
MEF25	Mid Expiratory Flow rate 25	Maximum flow achievable when 75% of Forced Vital Capacity has been expired (25% remain in lungs)
MEF50	Mid Expiratory Flow rate 50	Maximum flow achievable when lungs are half empty
MEF75	Mid Expiratory Flow rate 75	Maximum flow achievable when 25% of Forced Vital Capacity has been expired (75% remain in lungs)
TLC	Total Lung Capacity	about 6,000 mL, is the maximum amount of air that can fill the lungs (TLC = TV + IRV + ERV + RV).

In addition, written approvals were received from doctors who were treated. In order to ensure transparency and credibility in dealing with the ethics of medical professions.

B. Prediction Model

In this study, a framework has been built to predict the respiratory system diseases by applying Data Mining Techniques (DMT) on clinical info, and then extract the result of DMT to be a Knowledge Base for our application to achieve the prediction procedure. Fig. 1 shows the proposed framework which consists of (1) Data preprocessing phase to prepare the data before applying the DMT; (2) DM phase to apply DMT, (3) evaluation phase to appraise the performance and accuracy of the built model using a data mining evaluation method.

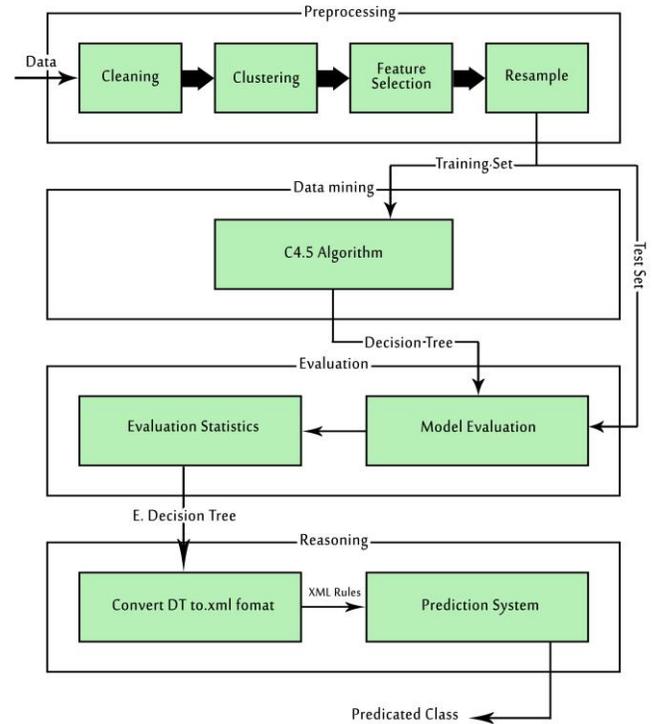


Fig. 1. The Framework of predicting the respiratory system diseases

1) Data Preprocessing Phase

In this phase, a sequence of steps were applied to cleaning, splitting and selecting the most appropriate topographies for the model from the patient data for applying DMT and resample the data sets into training and test sets.

Data cleaning phase is to clean the data and remove the records that contains empty values.

Feature Selection phase is to select a subset of topographies relevant to the target DMT from all the topographies of the data set. In the filtering approach; the feature selection algorithm is independent of the DMT which applied to the selected topographies.

We applied the feature selection on the clinical data and the selected attributes were Height, FVC, FEV3/FVC, FEF25 and FEF50. The class label (Result) is considered as the diagnosis of the disease (Normal Spirometry, Mild Restriction, Very Mild Obstruction, Moderate Restriction or Severe Restriction).

For Resampling; we divided the data into: 80% of the data as a training set and 20% as a testing set using unsupervised resample filter in Weka.

2) Data Mining Phase

Weka implementation of C4.5 (Weka J48) decision-tree learning algorithm was applied the training data sets.

C4.5; is based on the ID3 algorithm and tries to find simple (or small) decision trees (DT's). Some premises on which this algorithm is based will be presented in the following sections.

a) Construction

Some premises guide this algorithm, such as: If all cases are of the same class, the tree is a leaf and the leaf is returned labeled with this class;

For each attribute, calculate the potential information provided by a test on the attribute (based on the probabilities of each case having a particular value for the attribute). Also calculate the gain in information that would result from a test on the attribute (based on the probabilities of each case with a particular value for the attribute being of a particular class);

Depending on the current selection criterion, find the best attribute to branch on.

b) Counting gain

This process uses the “Entropy”, i.e. a measure of the disorder of the data. The Entropy of \vec{y} is calculated by

$$\text{Entropy}(\vec{y}) = -\sum_{j=1}^n \frac{|y_j|}{|\vec{y}|} \log \frac{|y_j|}{|\vec{y}|} \quad (1)$$

Iterating over all possible values of \vec{y} . The conditional Entropy is

$$\text{Entropy}(j|\vec{y}) = \frac{|y_j|}{|\vec{y}|} \log \frac{|y_j|}{|\vec{y}|} \quad (2)$$

and finally, we define Gain by

$$\text{Gain}(\vec{y}, j) = \text{Entropy}(\vec{y}) - \text{Entropy}(j|\vec{y}) \quad (3)$$

The aim is to maximize the Gain, dividing by overall entropy due to split argument \vec{y} by value j [18].

c) Pruning

Pruning is a significant step to the result because of the outliers. All data sets include a little subset of instances which are not well-defined, and vary from the other ones in its neighborhood. After the whole creation processes of the tree, which classify all the training set instances, it is pruned. This is to minimize classification errors which can be occurred because of specialization in the training set; we do this to make the tree more general concise.

d) Results

To show concrete examples of the C4.5 algorithm application, WEKA software tool has been used on training sets. The resulting classes are about the effect of the clinical features on the disease diagnosis, e.g. Normal Spirometry. Fig. 2 shows the resulting DT, using C4.5 implementation from WEKA on the data set.

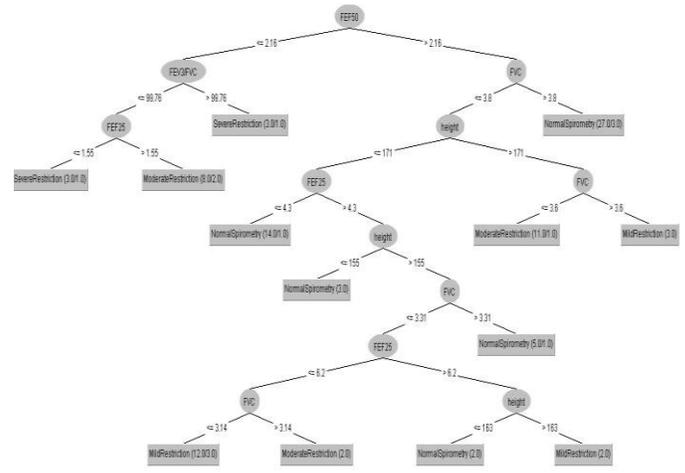


Fig. 2. DT of Respiratory System Diseases, Created by the C4.5 algorithm.

3) Evaluation Phase

The universality of the DT models was validated using the test data sets by the hold-out validation method.

The holdout method is considered as the simplest type of cross validation. The data set is divided into two dissimilar sets, which are known as the training set and the testing set [19]. The purpose approximator uses the training set only to fit a function which is used to predict the output values for the testing set data which has never seen these output values before. Then the faults it makes are gathered as before to give the mean complete test set fault, which will be used to appraise the model. This technique is usually preferable to the other approaches and takes less time to compute.

C. Reasoning

In the reasoning phase an application has been developed with C# programming language to perform the prediction operation. It can be considered as an expert system.

The knowledge base of this application is the model of decision tree algorithm. It is applied on Weka which delivers guidelines which has been rehabilitated into XML rules format

IV. EXPERIMENTAL RESULTS

This section shows an evaluation of the proposed framework performance using the applied DM techniques. At first we have countered serial of problems like there are not enough datasets via internet. Then when we found it as a real data, we still had to wait decision maker's permissions. Also the number of records make it poor data so, we had to gather more records to enhance system accuracy. We discovered that using real data of our local community, raises credibility of data records. We gathered our data from several sources, Mansoura University, Portsaid University and Certified medical clinic. Extensive experimental studies had been conducted in order to evaluate the model performance. Feature selection algorithm has been applied on each group. Height, FVC, FEF3/FVC, FEF25 and FEF50 features were selected. A subset of 20% of the data had been selected to test the model and the 80% used to build the classifier.

After applying the model; a large scale of statistical information were obtained. These performance measures had been used to evaluate the model as shown in Table 2. This table shows the performance for Diseases Diagnosis. The DT had 83.3% sensitivity, 100% specificity and 90.9% accuracy. Figures [3, 4, 5, 6] show the Receiver Operating Characteristic (ROC) curves for the four class labels (Mild, Moderate, Norman and Severe) and their sensitivity and specificity values at the optimal cutoff points.

TABLE 2: PERFORMANCE OF THE DECISION TREES OF THE 4 DAAS COMBINATION AFTER TESTING

Size of the tree	Number of leaves	TP	TN	Positive Predictive value %	Negative Predictive value %	Sensitivity	Specificity	AUC%	Accuracy%
25	13	17	2	100	33.3	89.4	100	93.9	90

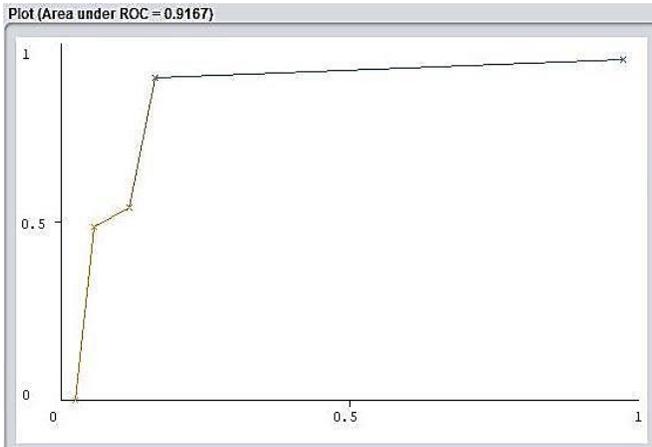


Fig. 3. Receiver Operating Characteristic (ROC) curves for Moderate Restriction.

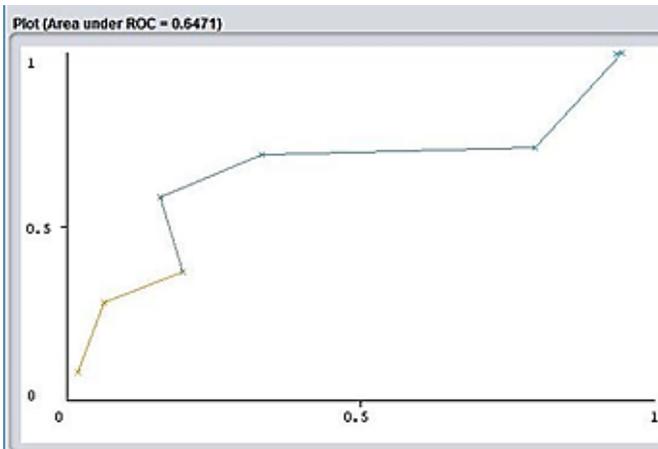


Fig. 4. Receiver Operating Characteristic (ROC) curves for Mild Restriction.

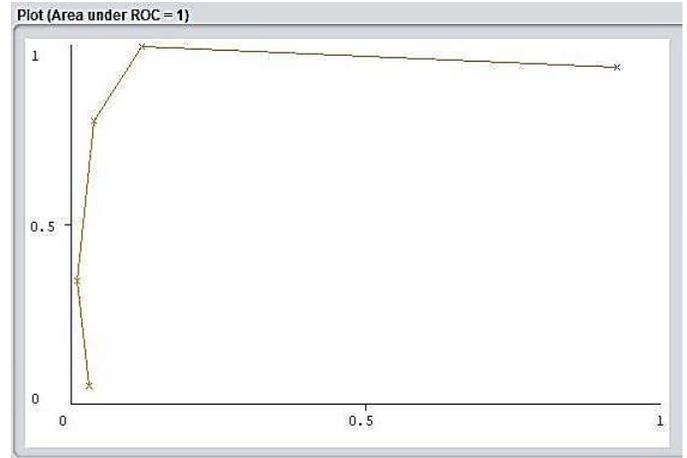


Fig. 5. Receiver Operating Characteristic (ROC) curves for Normal Spirometry.

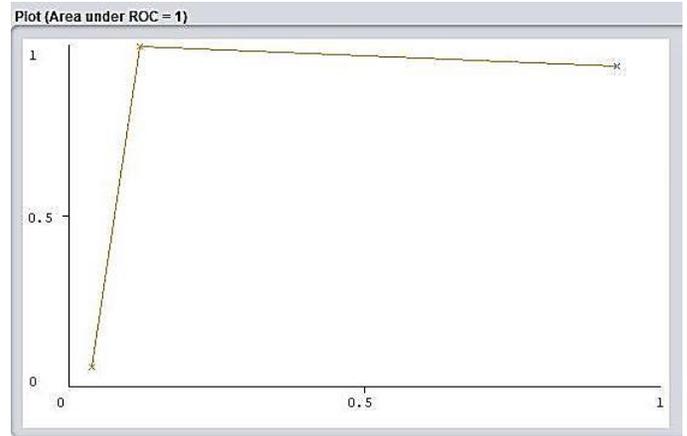


Fig. 6. Receiver Operating Characteristic (ROC) curves for Severe Restriction.

V. CONCLUSION AND FUTURE WORK

In this research, a framework has been built to predict the respiratory system diseases by applying Data Mining Techniques (DMT) on clinical information. Data has been collected from multiple sources. As well as the selection of real data to reflect the extent of the problem within our community. The collected data were pulmonary function tests (PFT) for approximately 250 patients. It has been divided into 80% as a training set and 20% as a testing set. Feature selection algorithm has been applied on each group. Decision Tree has been applied for the prediction, after that extraction of the result of DTs was performed. This constructed a Knowledge Base for our application to perform the prediction operation. The experimental results showed that the four groups give acceptable results.

In the future, more data sets will be used to train other classifiers and to try more experiments. Also other techniques will be applied and more than one technique will be combined to reach as high accuracy as possible.

REFERENCES

- [1] Guring A, Scrafford CG, Tielsch JM, Levine OS, Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med* 2011;105:1396–1403.
- [2] S. Reichert, R. Gass, C. Brandt, E. Andres. Analysis of respiratory sounds: state of the art. *Clinical medicine: circulatory Respir Pulm Med*, 2 (2008), pp. 45-58.
- [3] T Carroll Brandon, V Anderson David, Daley Wayne, Harbert Simeon, F Britton Douglas, W-Jackwood Mark, "Detecting symptoms of diseases in poultry through audio signal processing", *Signal and Information Processing (GlobalSIP) 2014 IEEE Global Conference on*, pp. 1132-1135, 2014.
- [4] Aykanat, M., Kılıç, Ö., Kurt, B and Saryal, S. (2017). Classification of lung sounds using convolutional neural networks, *EURASIP Journal on Image and Video Processing*, 2017:65, pp. 1-9.
- [5] J. Semedo, A. Oliveira, A. Machado, J. Moreira, J. Rodrigues, J. Aparício et al., "Computerised Lung Auscultation-Sound Software (CLASS)", *Procedia Computer Science*, vol. 64, pp. 697-704, 2015.
- [6] Pinho C, Oliveira D, Oliveira A, Dinis J, Marques A. LungSounds@UA Interface and Multimedia Database. *Procedia Technology*. 2012;5(0):803-11.
- [7] Naves R, Barbosa BH, Ferreira DD. Classification of lung sounds using higher-order statistics: A divide-and-conquer approach. *Comput Methods Programs Biomed*. 2016;129:12–20.
- [8] Palaniappan R, Sundaraj K, Ahamed NU: Machine learning in lung sound analysis: a systematic review. *Biocybern Biomed Eng*. 2013, 33: 129-135.
- [9] J. Acharya, A. Basu, & W. Ser, "Feature extraction techniques for low-power ambulatory wheeze detection wearables," 2017, 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2017), Korea, July, 2017
- [10] Oliveira A, Marques A: Respiratory sounds in healthy people: a systematic review. *Respir Med*. 2014, 108: 550-570.
- [11] M F Syahputra, S I G Situmeang, R F Rahmat, R Budiarto, "Noise Reduction in Breath Sound Files Using Wavelet Transform Based Filter", *IOP Conf*. 2017.
- [12] Sengupta N, Sahidullah M, Saha G. Lung sound classification using cepstral-based statistical features. *Comput Biol Med*. 2016;75:118–129.
- [13] Bokov P, Mahut B, Flaud P, Delclaux C. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Comput Biol Med*. 2016;70:40–50.
- [14] Jin F, Sattar F, Goh DYT. New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. *Neurocomputing*. 2014;123:362–371.
- [15] Amrulloh YA, Abeyratne UR, Swarnkar V, Triasih R, Setyati A. Automatic cough segmentation from non-contact sound recordings in pediatric wards. *Biomed Signal Process Control*. 2015;21:126–136.
- [16] Marani R.; Perri A.G. A New Pressure Sensor-Based Electronic Medical Device For The Analysis of Lung Sounds. *Proceedings of MELECON 2010, Valletta, Malta*, 25 – 28.
- [17] A. Sivasankari, S. Sudarvizhi and S. Radhika Amirtha Bai, "Comparative study of different clustering and decision tree for data mining algorithm," *International Journal of Computer Science and Information Technology Research*, Vol. 2, Issue 3, pp. 221-232, 2014.
- [18] Vijay Kumar Mago and Nitin Bhatia, *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies*, 2011.