A Convolution Neural Network Model for Detecting Respiratory Diseases from Lung Auscultation Sounds using MFCC

Aditya Dhavala, Asif Ahmed, B Yasaswini, Pradnya Meshram,
Arpit Mohakul, Srikar Peddapally, Rishabh Verma

ABSTRACT

Lung sound auscultation is an essential procedure to diagnose abnormalities in the respiratory system. Skilled physicians usually have a keen eye for detail while trying to diagnose pulmonary diseases. However, automated analysis of lung sound auscultation may be useful in scenarios where there is a lack of information or resources, as well as when there is uncertainty. This paper proposes using a convolutional neural network architecture to classify diseases using Mel frequency cepstral coefficients. The mel frequency cepstral coefficients are derived by applying cosine transforms on Mel spectrograms. The proposed model has been tested using a publicly available, patient-independent train-validation dataset from ICBHI 2017. Once the proposed model was implemented over the data set, a training accuracy of 97.09% and a testing accuracy of 89.14% were yielded. The model also ensured to avoid overfitting.

Subjects: Artificial intelligence, deep learning, neural networks, Convolution neural networks

Keywords: *Deep learning, CNN, MFCC, spectrogram, mel spectrogram, respiratory disease classification, machine learning*

I. INTRODUCTION

Chronic respiratory diseases are one of the most common noncommunicable diseases in the world., mainly due to the ubiquity of noxious environmental, occupational, and behavioral inhalation exposures [a]. Pulmonary disorders are considered the third largest fatality causing disease mortality globally [1]. Per year, more than 3 million people die from the chronic obstructive pulmonary disease (COPD), tuberculosis, asthma, acute lower respiratory tract infection (LRTI), and lung cancer, according to the World Health Organization (WHO) [2], [3]. These respiratory disorders have a significant impact on the healthcare system as a whole and people's health. Prevention, early diagnosis, and treatment of these deadly diseases are considered essential factors in minimizing the negative impacts. Auscultation of the respiratory sounds with the aid of a stethoscope is the traditional procedure for diagnostics that general practitioners and specialists use in the preliminary investigation of the pulmonary system. Though this method is simple and easy for preliminary diagnosis, it is inefficient for many reasons. It needs an experienced practitioner to identify the disease from the lung sounds right. Human auditory capacity is limited and is often corrupted by many noises such as ambiance, heart sounds, etc. Although the availability of various strategies for a pulmonary sound investigation like arterial blood gas analysis, plethysmography, spirometry, auscultation remains the most preferred strategy opted by physicians owing to its simplicity and inexpensive nature [4]. Wheeze, rhonchus, crackle, stridor, squawk, and pleural rub are the commonly known abnormal (adventitious) lung sounds. Features of these anomalies like frequency, energy, timbre, intensity, the pitch can

come in handy in separating them from a typical lung sound [6], [7]. Therefore, these sounds tend to have great importance in diagnosing specific pulmonary ailments. They are also essential for assessing chronic and non-chronic characteristics of lung disorders. However, the minute variation amongst a few of these adventitious respiratory sound classes becomes a challenging task, even for a physician with a great experience. The ambiguity in some cases might even lead to misinterpretation or subjectivity in diagnosis [8]. Under these circumstances, interpreting the pulmonary data automatically by algorithms developed from artificial intelligence techniques would be highly beneficial. These techniques would help diagnose the disease better, especially in regions where an experienced pulmonary practitioner is available or act as an aid to a physician in arriving at the correct diagnosis whenever there involves any ambiguity or subjectivity. With the advancement of technologies and neural networks and improved knowledge resources, extensive research is currently happening in this region over the past decade. In recent years, a diverse and massive set of algorithms in Machine Learning became available and accessible [8]–[22]. Different types of techniques for extracting features such as statistical features [9], entropy-based features [10], frequency domain features, Mel Frequency Cepstral Coefficients (MFCC), time-domain features [8], scalograms [13], etc. are being adopted and integrated with ML algorithms.

New advances and development were made possible in recent times owing to the introduction of Deep Learning (DL) techniques which opened gateways to many new technologies that gave extremely promising results in various fields, which can be extended to medical diagnostics and biomedical engineering too [23]–[28]. Limitations to the general ML algorithms are also mitigated in Deep Learning (DL) approaches as they offer more generic automated feature learning techniques. In a similar vein, DL-based paradigms are currently in use in recent years for the detection of respiratory irregularities, and maladies from lung auscultation data have demonstrated highly impressive results. [3], [29]–[40]. Deep neural networks, on the other hand, go through a rigorous training process with a computationally large training sample, which takes a long time and necessitates the use of powerful computing resources to achieve proper functionality. Because of a dearth of computational power, wearable devices, and mobile platforms that are currently available, it becomes challenging to incorporate deep learning models. While designing AI-assisted automated medical diagnosis frameworks, patient specificity in the train and validation datasets should be considered a crucial element in generating accurate findings for unseen data of patients, specifically for chronic diseases.[41], [42]. Because of the shortage and inadequacy of medical data, this fact is usually ignored in current literature. Because of the corruption of the dataset by intrapatient dependencies as a result of the random adoption of 80 % -20 % or some other percentage of the train validation so, the obtained results rarely stand out as reliable and generalized in the case of a new patient.[34]. Real world scenarios can be best achieved with a patient independent division even though it takes additional time and effort.

Using the ICBHI 2017 science challenge respiratory sound database, a CNN architecture has been introduced to achieve respiratory disease classification (ternary chronic classification) [43] while preserving a patient-independent dataset splitting strategy for train validation. An approach to obtain mel frequency cepstral coefficients is presented, and later a model is proposed whose merits are evaluated in a detailed manner.

The remaining part of the paper is listed out as follows. To start with, previous research on lung disease classification using machine learning methods is included. Following that are the

techniques to preprocess the data, followed by the feature extraction techniques. The proposed CNN architecture follows that, and lastly, the merits and demerits of the proposed model and technique are evaluated.

II. RELATED WORK

Respiratory sound classification has been the primary goal for much ongoing research works in the field of machine learning and deep learning. A majority of the research lies in detecting and classifying respiratory anomalies [8]-[22],[29]-[34] into various sounds like a wheeze and a crackle rather than predicting pulmonary diseases directly from the sound. Only a few recent works concentrate on the direct classification and diagnosis of the diseases from the lung sound. These works required an elaborate processing and a detailed design of the CNN or RNN model because of the signal's inherent complexity [35]-[40]. Though the complexity involved is higher, the outputs have been divided into three resolutions, the ternary chronic classification (healthy, chronic disease, non-chronic disease) [36], [40], the binary classification (healthy, pathological) [35], [36], and multi-class distinct disease classification [37], [40]. In [35], a novel CNN-based ternary classification method was adopted, and it worked excellently, with an accuracy of 82 percent and an ICBHI score of 88 percent. Later, the same researchers suggested a Mel-Frequency Cepstral Coefficient (MFCC) and Long Short-term Memory (LSTM) based framework capable of both binary and ternary classification of respiratory ailments [36], which demonstrated excellent performance with 99% and 98 accuracies, respectively. Another research using sophisticated RNN architecture achieved a predictive performance of 95.67 % for six pathology-driven diseases [37], except it requires a significant amount of preprocessing. Recent research [38] showed that the use of a CRNN network with a CNN-Mixture-of-Experts (MoE) baseline to learn both spatial and time-sequential features from spectrograms generated a sensitivity of 96% and a specificity of 83% in ternary respiratory disorder identification. The same study found that binary classification has a precision and resilience of 83 percent and 99 percent, respectively. An additional study utilizing the robust Teacher-Student learning schemes with knowledge distillation was done as a follow-up to[38], and it resulted in a drastically reduced specificity while retaining sensitivity [39]. Because the current grossly imbalanced datasets of lung auscultations make it much more difficult to categorize respiratory diseases, a recent study [40] explored various data augmentation approaches, notably SMOTE, Adaptive Synthetic Sampling Method (ADASYN), and Variational autoencoder (VAE). The VAE-based Mel-spectrogram augmentation strategy, in combination with a CNN model, produced the greatest results in chronic ternary categorization, with 98.5 % sensitivity and 99.0% specificity. Despite the fact that the scope of DL-based frameworks with a spectrogram-based feature extraction method has been explored in different texts for direct classification of respiratory disorders from lung auscultations [38]–[40] to the best of the authors' knowledge, this paper reuses their approaches of feature extractions employing MFCCs. This paper comes with an approach to augment the data before preprocessing. Drawing inspiration from all the above research works, this model aims to classify the pulmonary sound data into three classes, namely chronic, non-chronic and healthy, and at the same time maintaining patient independence.

III. MATERIALS AND METHODS

ICBHI 2017 Dataset

The ICBHI 2017 database (International Conference on Biomedical Health Informatics) is a publicly accessible pulmonary auscultations benchmark dataset [43]. Two independent research teams from Portugal and Greece

gathered the data. The study consisted of 5.5 hours of audio recordings sampled at various frequencies (4 kHz, 10 kHz, and 44.1 kHz) in 920 audio samples, varying from 10s to 90s, collected from 126 subjects in different body positions using a variety of equipment.[44]. Two systems are used to annotate the samples professionally:

1. According to the corresponding patients' pathological condition, i.e., healthy and seven distinct disease classes, namely Pneumonia, Bronchiectasis, COPD, URTI, LRTI, Asthma
2. Each respiratory cycle is based on the existence of respiratory anomalies, such as crackles and wheezes. The dataset and data collection methods are discussed in more detail in[44].

There are 37 files pertaining to Pneumonia, 16 files pertaining to Bronchiectasis, 793 files pertaining to COPD, 35 files pertaining to healthy subjects without any ailments, 23 files pertaining to URTI, 13 files pertaining to Bronchiolitis, two files pertaining to LRTI, and lastly one file corresponding to a subject suffering from asthma.

Noise Filtering

Filters enable desired frequency signals to pass while rejecting unwanted frequency signals. Since the recommended frequency range for lung auscultation signals is 50 Hz to 2500 Hz [4], the reported signals are filtered with a 6th order Butterworth bandpass filter, and the 50Hz to 2500Hz frequency components are preserved. Then, for ensuring consistency, all the samples are resampled to 22050 Hz, and for attaining the device homogeneity, we normalized in the range of [-1,1].

Data Augmentation

Data augmentation is a data processing technique that involves inserting slightly changed copies of existing data or developing new synthetic data from existing data to increase the amount of data available.

Benefits of Data Augmentation are:

1)Adds more training data in the models
2)Removes the class imbalance issue in classification problems
3)Reduces data overfitting and create variability in data
4)Prevents data scarcity for better models
5)Reduces the cost of collecting and labeling data

Data Augmentation for audio files

To produce multiple audio data, we can use time shifts, pitch changes, and speed changes. With only one line of code, Librosa (library for speech and audio file recognition and organization) can adjust the pitch and speed of audio [48]. NumPy, on the other hand, makes it easy to deal with time shifts.

Shifting Time: Shift the time by a random second to the left or right. If we need to change audio to the right with y seconds last, we will set y seconds to 0, and if we need to shift audio to the left with y seconds first, we will set y seconds to 0.

Changing Pitch: This augmentation is a wrapper for the librosa feature, and it changes pitch at random.

Changing Speed: Changing speed is equivalent to changing pitch, except the augmentation is handled by librosa, which effectively stretches the duration by a set amount of time.

The ICBHI 2017 data set is highly imbalanced. To improve the efficiency of the model and to prevent model imbalance because of the data set, we are using the above-mentioned augmentation techniques. Every file which is not labeled as "COPD" is undergoing the above three transforms for data augmentation purposes. So for every file not labeled as "COPD," we are getting three more augmented audio files. Firstly, for shifting time transform, the audio signal is shifted right by two seconds. The second file is obtained by applying a pitch factor of 2 to the audio files under consideration. Lastly, we are changing the speed of the audio file by a factor of 1.2.

After data augmentation, we are getting 148 files pertaining to Pneumonia, 64 files pertaining to

Bronchiectasis, 793 files pertaining to COPD, 140 files pertaining to healthy subjects without any ailments, 92 files pertaining to URTI, 52 files pertaining to Bronchiolitis, eight files pertaining to LRTI and lastly four files corresponding to a subject suffering from asthma.

Feature Extraction

Feature extraction is part of the dimensionality reduction method, which breaks and compresses a vast collection of raw data into compact classes [49]. In a nutshell, feature extraction assists in the extraction of the best feature from large data sets by selecting and integrating variables into features, hence minimizing the quantity of data that the model must process. These features are simple to implement, but they are capable of accurately and uniquely describing the real data set. Though we have numerous methods for feature extraction from a dataset, we have used mel frequency cepstral coefficient (mfcc) for feature extraction, and mfcc is the best known and most efficient way of extracting features from a dataset. The steps to extract MFCC are shown in Figure 1.

MFCC (mel frequency cepstral coefficient)

With the assumption that the human ear is a competent speaker recognizer, MFCC computation is essentially a reproduction of the human hearing system [49]. To maintain the phonetically significant components of the speech signal, frequency filters spaced linearly at low frequencies and logarithmically at high frequencies were used, with MFCC features based on observational difference in the human ear's vital bandwidths. The initial step in calculating the MFCC (Mel frequency cepstral coefficient) is windowing the speech sample, which involves separating the data into frames. Because the high frequency formants process has a lower amplitude than the low frequency formants, high frequencies are prioritized to ensure that all formants have the same amplitude.

As soon as the windowing is finished, the Fast Fourier Transform (FFT) has been used to calculate the power spectrum of each frame. The filter bank processing is then done using mel-scale on the power spectrum. To compute MFCC coefficients, the DCT is applied to the speech signal after translating the power spectrum to the log domain [3].

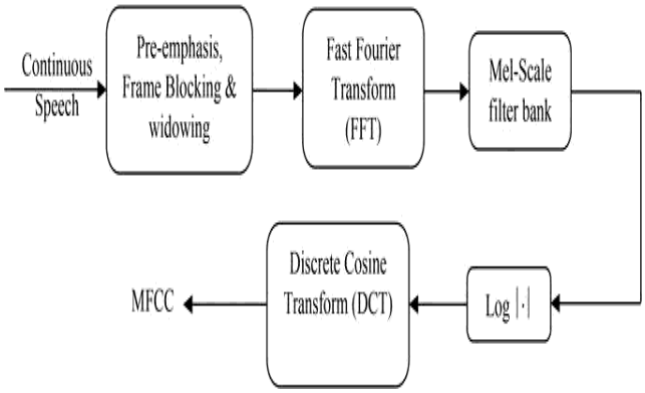

Figure 1: Steps to extract MFCC

Mel Filter Bank

The main reasons for using mel filter banks for the mfcc process:

1)It employs Mel-frequency scaling, a perceptual scale that aids in replicating the human auditory system's operation. It relates to higher resolution at low frequencies and less resolution at high frequencies.

2) The triangular filterbank captures energy at each critical band and gives a rough

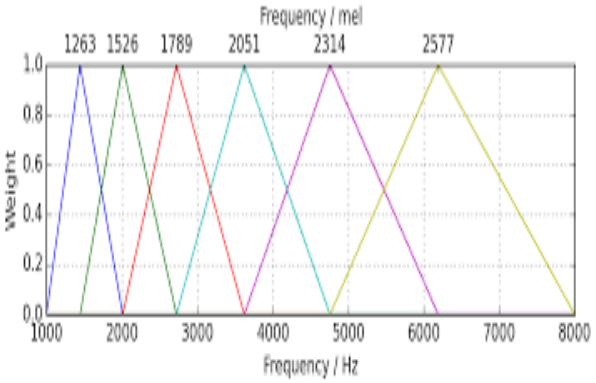

Figure 2: Mel Bands

approximation of the spectrum form while smoothing the harmonic structure. While raw DFT bins can be manipulated, this does not reduce the dimensionality of the features, which is the whole point of filterbank analysis to capture the spectral envelope. A mel filter band is shown in Figure 2.

Librosa is the python library that is being used in the feature extraction process [45]. The audio files in consideration are 20 seconds in duration. The preprocessing on the audio file included the application of a 6th order Butterworth filter followed by the data augmentation techniques on the files, which are not labeled as "COPD." After the feature extraction, the dimensions of the MFCC obtained are (40x862x1), as shown in Figure 3.

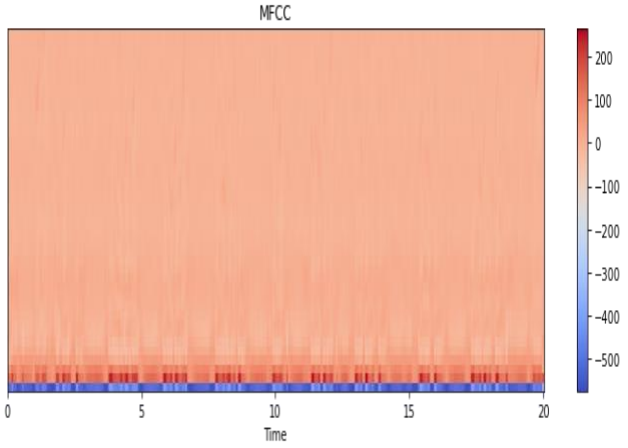

Figure3: MFCC

IV. CNN ARCHITECTURE

Before CNN became well-known, several other image classification algorithms were used. Photographs were used to generate characteristics, which were then submitted to a classification system like SVM.[50]. The pixel level values of images were also used as a function vector in some algorithms. We could, for example, train an SVM with 784 features, each of which is a pixel value for a 28x28 image. CNNs can be known as automated image function extractors, and if we use a pixel vector algorithm, we lose a lot of spatial connectivity between pixels; a CNN (Convolution Neural Network) efficiently uses adjacent pixel information to downsample an image by convolution and then uses a prediction layer at the end. [50].

Padding: When we do the convolution of the image with the filters/kernels, the output dimension gets reduced, and because of this, we lose some data [51]. So, to overcome this problem, we introduce some supplemental zero boundaries to the convolution and then recalculate it to cover all of the input values.

Pooling: A small portion is referred to as pooling in general. Pooling can be classified into two categories. The first is average pooling, which utilizes the average value, and the second is max pooling, where we take the maximum value from the group of values [52].

Function of Activation: It's a node that appears at the end of or in the midst of a neural network. They assist in determining whether or not a neuron will fire [53]. We have a variety of activation functions, but for the objectives of our project, we have primarily focused on Rectified Linear Unit (ReLU).

CNN Model

We have used 4 convolution layers with activation function as “ReLU” with 64, 64, 96 and 96 filters respectively and there are 6 hidden layers with 256, 128, 64, 32, 16, and 8 neurons respectively with dropouts of 60%, 30%, 15%, 7.5%, and 3.25% respectively. Finally, there is one output layer with three neurons as we are classifying three different types of diseases, which are chronic, non-chronic, and healthy. The dropout layers used in between the hidden layers were to ensure that there is a balance between generalization and memorization. Dropout layers in the sequential model of Keras ensure that overfitting of the model is prevented. The proposed model is shown in Figure 4. The input to the CNN file is the MFCC coefficients of dimension (40x 862x1).

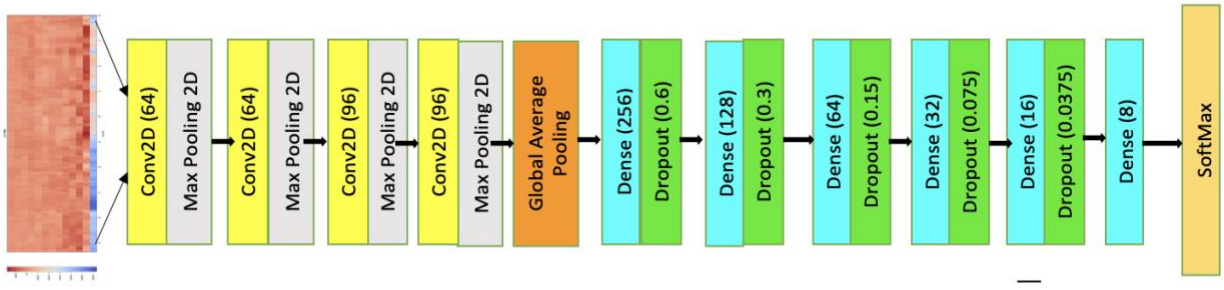

Figure 4: Proposed CNN Architecture

V. EXPERIMENTAL RESULTS

Evaluation Criteria

The test-train split among the augmented dataset is 80%and 20%. A critical aspect that was maintained while dividing the data set randomly was maintaining the patient uniqueness. This makes the model more significant for real life applications.

Well known evaluation metrics such as F1 score, precision, accuracy, recall are used to determine the model's efficiency.

Experimental setup

The Convolution neural network was built using Keras and TensorFlow backend. It was trained using the intel i5 10th generation processor over a Jupyter notebook. A batch size of 128 was employed to run the model over the testing and training data set during each epoch. Because the total length of the input data was 1289, the epoch was split into nine batches according to the batch size. The optimizer used for training was Adam. The Adam" optimizer" learning rate was left at default at 0.001. The values of beta1 and beta two, which are the exponential decay rates of first and second and the second moments, respectively, were also left to default at 0.9 and 0.999, respectively [46].

As stated before in the paper, the model was trained only for ternary classification, i.e., chronic, non-chronic, and healthy. The classification work of the proposed CNN is evaluated, and the performance metrics during training are plotted.

Classification performance of the proposed network

The model initially showed a pre-training accuracy of 61.6279%. The time taken to train the model over 120 epochs was thirty-six minutes and fourteen seconds. Post training, the model showed a training accuracy of 97.09% and a testing accuracy of 89.15%.

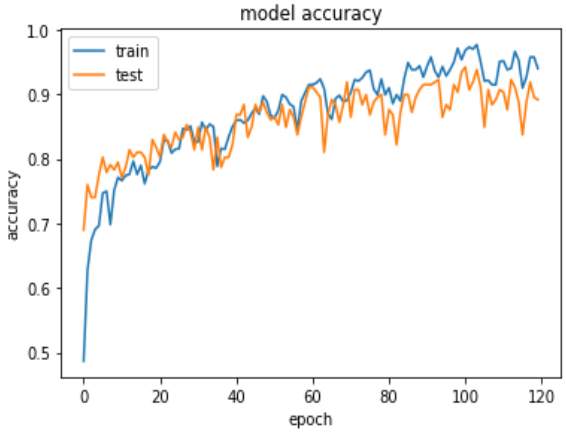

Figure 5: Model Accuracy vs. Epoch

The model showed the least training loss of 0.0622 at the 104th epoch. The results for ternary classification were better than most models proposed in the previous related works.

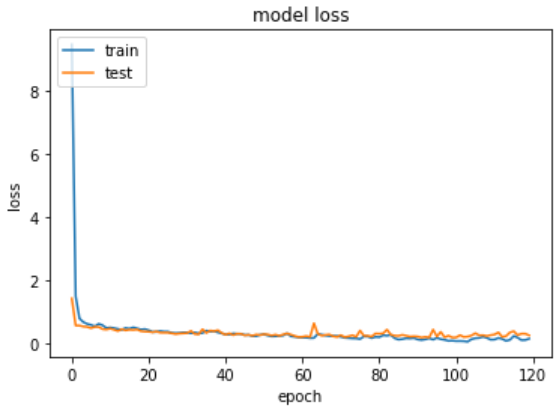

Figure 6: Model Loss vs. Epoch

The ROC curves were plotted for the three output classes, and the areas for chronic, healthy, and non-chronic diseases were 0.99, 0.98, and 0.97, respectively.

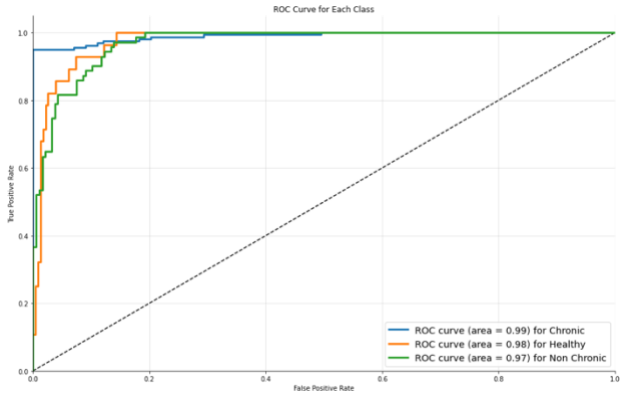

Figure 7: ROC Curve

The Confusion Matrix is given in Table 1.

Table 1: Confusion Matrix

	Chronic	**Healthy**	**Non-Chronic**
Chronic	155	0	4
Healthy	4	21	3
Non-Chronic	12	5	54

The performance metrics such as F1 score, precision, accuracy for the model is mentioned in Table 2.

Table 2: Summary of Classification Performance

	Precision	**Recall**	**F1**	**Support**
Chronic	0.91	0.97	0.94	159
Healthy	0.81	0.75	0.78	28
Non-Chronic	0.89	0.76	0.82	71
Accuracy			0.89	258
Macro Avg	0.87	0.83	0.85	258
Weighted Avg	0.89	0.89	0.89	258

Comparison with Other Models

A detailed comparison has been made between deep learning models that have used ICBHI 2017 database. Three models referenced in [56], [57] and [58] respectively along with the two models proposed in [55] have been taken up for comparison. The model proposed in [57] uses a hidden Markov model and has achieved an accuracy of 39.56%. The proposed model in [56] is using a low-level feature decision tree and achieved an accuracy of 49.62%. The model proposed in [58] uses an SVM with wavelet extraction from STFT achieving an accuracy of 57.88%. Lastly, [55] has proposed two different models which use Deep Feature with CNN along with SVM classifier and transfer learning with CNN with SoftMax classifier which achieved an accuracy of 65.5% and 63.09% respectively. In Table 3, the methodologies and accuracies of various classifiers discussed above along with proposed model is given for comparison.

Table 3: Comparison of Classifiers

Author	**Methodology**	**Accuracy**
Jakovljević et al. [57]	MFCC, Hidden Markov model	39.56%
Chambres et al. [56]	Low level feature, decision tree	49.62%
Serbes et al. [58]	STFT + Wavelet, SVM classifier	57.88%
Fatih Demir [55]	Deep Feature with CNN model, and SVM classifier	65.50%
Fatih Demir [55]	Transfer learning with CNN Model, and SoftMax classifier	63.09%
Proposed Model	MFCC features based CNN classifier	89.15%

VI. CONCLUSION

In this research work, we have proposed a convolution neural network model that uses mel frequency cepstral coefficients from lung sounds to classify pulmonary diseases. An approach using the short time Fourier transform and the cosine transform to obtain MFCCs is extensively discussed. The publicly available ICBHI 2017 data set has been used to classify the diseases according to their chronicity. The proposed model has given a reliable accuracy of 89.15% for ternary chronic classification. It has also outperformed many of the classifiers that were proposed in previous studies. A primary feature that people look for in this pandemic struck world is the quickness in results with considerable accuracy for medical tests. As Covid-19 continues to take lives every day, quicker and reliable systems to detect the disease in its early stages are crucial and necessary. The proposed model can also have a more extensive and significant use considering the reliability and accuracy, given that the model is trained in that specific direction. We believe that implementing the model is the right step in the direction of portable and wearable technology, which can help save many lives.

REFERENCES

1) "The Global Impact of Respiratory Disease Second Edition — CHEST Physician," 2017, [Online]. Available:https://www.mdedge.com/chestphysician/article/140055/society-news/global-impact-respiratory-disease-second-edition.
2) C. D. Mathers and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," PLoS Medicine, vol. 3, no. 11, p. e442, 2006.
3) "WHO — Global tuberculosis report 2019," 2019, [Online]. Available: https://www.who.int/tb/publications/global report/en/.
4) S. Jayalakshmy and G. F. Sudha, "Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks," Artificial Intelligence in Medicine, vol. 103, p. 101809, 2020.
5) S. Reichert, R. Gass, C. Brandt, and E. Andres, "Analysis of respiratory sounds: state of the art," Clinical Medicine. Circulatory, Respiratory and Pulmonary Medicine, vol. 2, pp. CCRPM–S530, 2008.
6) "Auscultation of the respiratory system," Annals of Thoracic Medicine, vol. 10, no. 3, p. 158, 2015, M. Sarkar, I. Madabhavi, N. Niranjan, and M. Dogra
7) "Fundamentals of lung auscultation," New England Journal of Medicine, vol. 370, no. 8, pp. 744–751, 2014, A. Bohadana, G. Izbicki, and S. S. Kraman
8) "New parameters for respiratory sound classification," in Canadian Conference on Electrical and Computer Engineering, vol. 3. IEEE, 2003, pp. 1457–1460, M. Bahoura and C. Pelletier
9) "Machine learning in lung sound analysis: a systematic review," Biocybernetics and Biomedical Engineering, vol. 33, no. 3, pp. 129–135, 2013, R. Palaniappan, K. Sundaraj, and N. U. Ahamed
10) "A novel wheeze detection method for wearable monitoring systems," in 2009 International Symposium on Intelligent Ubiquitous Computing and Education. IEEE, 2009, pp. 331–33, J. Zhang, W. Ser, J. Yu, and T. Zhang
11) "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," Computers in Biology and Medicine, vol. 39, no. 9, pp. 824–843, 2009, M. Bahoura
12) J. Acharya, A. Basu, and W. Ser, "Feature extraction techniques for low-power ambulatory wheeze detection wearables," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 4574–4577.
13) N. Gautam and S. B. Poole, "Wavelet scalogram analysis of phonopulmonographic signals," International Journal of Medical Engineering and Informatics, vol. 5, no. 3, pp. 245–252, 2013.
14) G. Serbes, C. O. Sakar, Y. P. Kahya, and N. Aydin, "Pulmonary crackle detection using time-frequency and time–scale analysis," Digital Signal Processing, vol. 23, no. 3, pp. 1012–1021, 2013.
15) S. ˙Ic¸er and S¸. Gengec¸, "Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds," Digital Signal Processing, vol. 28, pp. 18–27, 2014.
16) F. Jin, F. Sattar, and D. Y. Goh, "New approaches for spectro-temporal feature extraction with applications to respiratory sound classification," Neurocomputing, vol. 123, pp. 362–371, 2014.
17) P. Bokov, B. Mahut, P. Flaud, and C. Delclaux, "Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population," Computers in Biology and Medicine, vol. 70, pp. 40–50, 2016.
18) P. Mayorga, C. Druzgalski, R. Morelos, O. Gonzalez, and J. Vidales, "Acoustics based assessment of respiratory diseases using gmm classification," in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, 2010, pp. 6312–6316.
19) T. R. Fenton, H. Pasterkamp, A. Tal, and V. Chernick, "Automated spectral characterization of wheezing in asthmatic children," IEEE Transactions on Biomedical Engineering, vol. 32, no. 1, pp. 50–55, 1985.
20) H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds: advances beyond the stethoscope," American Journal of Respiratory and Critical Care Medicine, vol. 156, no. 3, pp. 974–987, 1997.

21) Z. Dokur, “Respiratory sound classification by using an incremental supervised neural network,” Pattern Analysis and Applications, vol. 12, no. 4, p. 309, 2009.
22) S. Rietveld, M. Oud, and E. H. Dooijes, “Classification of asthmatic breath sounds: preliminary results of the classifying capacity of human examiners versus artificial neural networks,” Computers and Biomedical Research, vol. 32, no. 5, pp. 440–448, 1999.
23) B. Bozkurt, I. Germanakis, and Y. Stylianou, “A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection," Computers in Biology and Medicine, vol. 100, pp. 132–143, 2018.
24) A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, “Towards domain invariant heart sound abnormality detection using learnable filterbanks,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2189–2198, 2020.
25) U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, “Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," Computers in Biology and Medicine, vol. 100, pp. 270–278, 2018.
26) S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, et al., “Cnn architectures for large-scale audio classification," in 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE, 2017, pp. 131–135.
27) S. Debbal and F. Bereksi-Reguig, “Analysis of the second heart sound using continuous wavelet transform,” Journal of Medical Engineering & Technology, vol. 28, no. 4, pp. 151–156, 2004.
28) A. Meintjes, A. Lowe, and M. Legget, “Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks,” in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018, pp. 409–412.
29) K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, and S. Kido, "Automatic classification of a large-scale respiratory sound dataset based on convolutional neural network," in 2019 19th International Conference on Control, Automation and Systems (ICCAS). IEEE, 2019, pp. 804–807.
30) M. Aykanat, O. Kılıc¸, B. Kurt, and S. Saryal, “Classification of lung ¨ sounds using convolutional neural networks,” EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, p. 65, 2017.
31) F. Demir, A. Sengur, and V. Bajaj, “Convolutional neural networks based efficient approach for classification of lung diseases,” Health Information Science and Systems, vol. 8, no. 1, p. 4, 2020.
32) R. Liu, S. Cai, K. Zhang, and N. Hu, “Detection of adventitious respiratory sounds based on convolutional neural network,” in 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). IEEE, 2019, pp. 298–303.
33) D. Bardou, K. Zhang, and S. M. Ahmad, “Lung sounds classification using convolutional neural networks,” Artificial Intelligence in Medicine, vol. 88, pp. 58–69, 2018.
34) J. Acharya and A. Basu, “Deep neural network for respiratory sound classification in wearable devices enabled by patient-specific model tuning," IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 3, pp. 535–544, 2020.
35) D. Perna, “Convolutional neural networks learning from respiratory data,” in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018, pp. 2109–2113.
36) D. Perna and A. Tagarelli, “Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks,” in 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2019, pp. 50–55.
37) V. Basu and S. Rana, “Respiratory diseases recognition through respiratory sound with the help of deep neural network,” in 2020 4th International Conference on Computational Intelligence and Networks (CINE). IEEE, 2020, pp. 1–6.
38) L. Pham, I. McLoughlin, H. Phan, M. Tran, T. Nguyen, and R. Palaniappan, “Robust deep learning framework for predicting respiratory anomalies and diseases,” arXiv preprint arXiv:2002.03894, 2020.
39) L. Pham, “Predicting respiratory anomalies and diseases using deep learning models,” arXiv preprint arXiv:2004.04072, 2020.
40) M. T. Garc´ıa-Ordas, J. A. Ben' 'ıtez-Andrades, I. Garc´ıa-Rodr´ıguez, C. Benavides, and H. Alaiz-Moreton, "Detecting respiratory pathologies' using convolutional neural networks and variational autoencoders for unbalancing data," Sensors, vol. 20, no. 4, p. 1214, 2020.
41) S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, “Convolutional neural networks for patient-specific ECG classification," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015, pp. 2608–2611.
42) N. U. Maheswari, A. Kabilan, and R. Venkatesh, “Speaker independent speech recognition system based on phoneme identification," in 2008 International Conference on Computing, Communication, and Networking. IEEE, 2008, pp. 1–6.
43) “ICBHI 2017 Challenge,” 2017, [Online]. Available: https: //bhichallenge.med.auth.gr/.
44) B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jacome, A. Marques,' et al., "A respiratory sound database for the development of automated classification," in International Conference on Biomedical and Health Informatics. Springer, 2017, pp. 33–37.
45) **“**librosa doc” https://librosa.org/doc/latest/index.html
46) “adam optimiser”https://keras.io/api/optimizers/adam/
47) James SL Abate D Abate KH et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.*Lancet.* 2018; 392: 1789-1858

48) Data Augmentation for audio by Edward ma- https://medium.com/@makcedward/data-augmentation-for-audio-76912b01fdf6
49) Some commonly used speech feature extraction algorithm by Sabir Ajibola Alim and Nazrul Rashid- https://www.intechopen.com/books/from-natural-to-artificial-intelligence-algorithms-and-applications/some-commonly-used-speech-feature-extraction-algorithms
50) Understanding the Architecture of CNN by Kousai Smeda - https://towardsdatascience.com/understand-the-architecture-of-cnn-90a25e244c7
51) Padding in Convolution Neural Network by Ting Hao Chen - https://medium.com/machine-learning-algorithms/what-is-padding-in-convolutional-neural-network-c120077469cc
52) Understanding Convolution and Pooling in Neural Networks by Miguel Fernandez Zafra- https://towardsdatascience.com/understanding-convolutions-and-pooling-in-neural-networks-a-simple-explanation-885a2d78f211
53) Activation Function in Neural Networks by Sagar Sharma - https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6
54) **"**A Lightweight CNN Model for Detecting Respiratory Diseases from Lung Auscultation Sounds using EMD-CWT-based Hybrid Scalogram" Samiul Based Shuvo, Shams Nafisa Ali, Soham Irtiza Swapil, Taufiq Hasan (Member IEEE) and Mohammed Imamul Hassan Bhuiyan (Member IEEE)
55) "Convolutional neural networks based efficient approach for classification of lung diseases" Fatih Demir1, Abdulkadir Sengur and Varun Bajaj
56) Chambres G, Hanna P, Catherine MD. Automatic detection of patient with respiratory diseases using lung sound analysis. In: 2018 International Conference on Content-Based Multimedia Indexing (CBMI), 4–6 Sept. 2018, La Rochelle, France.
57) Jakovljevic N, Loncar-Turukalo T. Hidden Markov model based respiratory sound classification. Precis Med. 2017;17:39–43.
58) Serbes G, Ulukaya S, Kahya YP. An automated lung sound pre-processing and classification system based on spectral analysis methods. In: International Conference on Biomedical and Health Informatics; 2018, pp. 45–49.