International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 02 | Feb 2023 www.irjet.net p-ISSN: 2395-0072

# Three Feature Based Ensemble Deep Learning Model for Pulmonary Disease Classification

**Aditya Dawadikar[1], Anshu Srivastava[2], Neha Shelar[3], Gaurav Gaikwad[4], Prof. Atul Pawar[5]**

[1]*Student, Dept of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*
[2]*Student, Dept of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*
[3]*Student, Dept of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*
[4]*Student, Dept of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*
[5]*Professor, Dept of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*

---

**Abstract –** *In recent years there has been a rise in the number of patients suffering from acute and chronic pulmonary diseases because of varying reasons like pollution, lung damage, or infections. The following research is regarding a Neural Network based solution for the recognition of the abnormality and possible disease based on lung auscultation. The following paper depicts that RNN-LSTM and CNN were the best-performing techniques. Although a higher percentage of noise while capturing the auscultation audio and limited data leads to a saturation point for the models to improve. The dataset had over 5000 breathing cycles for COPD, whereas only about 100 breathing cycles for LRTI and URTI. This unbalanced data made it difficult for the models to perform well on test audio clips because of the bias introduced by the large count of COPD samples. We adopted a filter-based audio augmentation to rebalance the dataset. To get the most out of the data we had, we utilized multiple features like MFCC, Chromagram, and Spectrogram extracted from the same audio clip. Since these extracted features are not fathomable to humans, we used convolutional neural networks to perform primary feature extraction. Likewise, dedicated CNN models acted as feature extractors whereas the dense neural network served as the actual classifier. We developed multiple versions of the models with fine-tuned parameters. The ML models based on a single feature were considered the benchmark for evaluating more complex, multi-feature DL models.*

***Key Words***: **Pulmonary Diseases, CNN, MFCC, Chromagram, Spectrogram.**

## 1. INTRODUCTION

A huge number of deaths are caused every single year due to various pulmonary diseases such as asthma, bronchitis, pneumonia, etc. As per research, COPD is one of the leading ones causing the death of approximately 3.23 million. Affected people suffering from such pulmonary disorders have different breathing sounds as compared to healthy people. This includes rhonchus, crackles, wheezes, stridor, and plural friction rubs which are present in breathing sounds. In order to distinguish between healthy and infected breathing sound various parameter checks are used. For example, frequency, pitch, energy, etc. For in-depth diagnosis, doctors take the help of several medical tests like lung (pulmonary) function tests, lung volume tests, Pulse Oximetry, Spirometry, chest x-ray, CT scan, arterial blood gas analysis, etc which are time-consuming and costly. In order to reduce this process a lot of researchers have proposed various methods using computer science technologies like ML and DL classification algorithms which help us make an early diagnosis.

A lot of the research is based on the findings in speech analysis. For instance, the feature MFCC is heavily used for speech and sound recognition. The research in [34] and [26] is using MFCC for speaker recognition. While [33],[31], and [27] use similar techniques for bird sound and environment sound classification. MFCC, Spectrogram, and Chromagram are 2D features or audio representations. There is research based on the usage of traditional ML models [2]. Here the 2D image is to be converted to a flat 1D feature vector as input to the models like KNN, SVM, Decision Tree, etc. Other researchers used DL algorithms like CNN and RNN [9],[10], and [11]. While few have proposed that a combination of two or more such algorithms has shown better results. For example, RNN with LSTM, CNN BiLSTM,[17] and [18]. A survey for comparison of various these techniques can be found in [1]. In the following paper, we propose a method where we developed a Tri-feature based 2 step classifiers based on Convolutional Neural Network and Dense Neural Network

## 2. DATASET

A considerable amount of the existing research is based on the ICBHI 2017 [14] dataset. The major focus of researchers has been on the adventitious sound classification. Less attention is given to disease classification. While there are no controversies regarding the authenticity and correctness of the ICBHI dataset, one fundamental note that needed to be considered was the disease classes recognized by the medical community.

The ICBHI dataset consists of 8 classes namely, Asthma, Pneumonia, Healthy, COPD, URTI, LRTI, Bronchiectasis,

Bronchiolitis. In reality, COPD, URTI and LRTI are superset for other pulmonary diseases. COPD consists of Asthma, Pneumonia, Bronchitis, while LRTI consists of chronic Bronchitis, Bronchiolitis, Bronchiectasis and acute exacerbations of COPD. Hence a sample can be classified under Pneumonia, COPD, and LRTI at the same time and it should be considered correct classification by logic. But this will not only reduce the accuracy of the model statistically, and it is of less use in real diagnosis.

A solution to this issue is, making separate classifiers for absolute disorder classes and the umbrella (Superset) classes. Use of such a classification is that, when the same sample is tested on a model trained on the Superset classes, it will tell us about the severity of the disorder. Ex: a sample classified as pneumonia in the disorder classifier and LRTI in the umbrella classifier will tell us that the patient is suffering from acute Pneumonia, while a sample classified as COPD in umbrella classifier will tell us that the patient is suffering from chronic Pneumonia. This segregation helps us in giving the severity of the disorder without use of other medical tests. URTIs consist of throat related infections which are not clinically proven to be a leading cause of adventitious sounds in the lung auscultation, and URTI sounds have more correlation with the Healthy sounds. Hence, we can exclude the URTI class from consideration.

This class labeling requires expert knowledge and patience. This approach can be used when a suitable dataset is available, we are focusing on creating the absolute disease classifier. For this purpose, we are using samples belonging to the classes: Asthma, Pneumonia, Bronchiectasis, Bronchiolitis, Healthy and Lung Fibrosis. The samples used for Lung Fibrosis belong to the Mendeley Dataset [13].

## 3. DATASET AUGMENTATION

The skewed nature of the dataset, made it essential for us to rebalance it. Traditionally audio augmentation means artificially creating more audio data samples from the existing ones. This was achieved in various ways as listed in next subsection. Although these methods could work for other audio classification use cases. These techniques would distort our data to such an extent that the trained model will not be generalized over real-life data.

### 3.1 Limitations of Traditional Audio Augmentation

Following is a list of traditional audio augmentation techniques widely used, but they are not suitable on our data set for various reasons.

#### 3.1.1 Time Shift:

The respiratory cycles are extracted from the audio files. A normal respiratory cycle lasts for 2-4 seconds. It was observed that it may span up to 6 seconds in some cases hence a 6 seconds duration was selected for the samples. Smaller samples were center padded by 0 to get the required duration. Because of the center padding, the time shift will only cause the data to move towards left or right and add more zeros to one of the ends. This will not be effective in real life scenarios where the samples fed to the system will always be center padded.

#### 3.1.2 Pitch shift

Crackles, Wheezes, Rhonchi, Stridor, etc adventitious sounds have a certain pitch and the model will learn this information. Employing pitch shift will lead to hampering of the original data and the model will not generalize for real life scenarios.

#### 3.1.3 Changing speed

The respiratory sounds cannot be made faster as a standard respiratory cycle should last 2-4 seconds minimum. Speeding down will cause stretching of the data and any ambient noise may sound similar to a wheeze because of the time stretch.

#### 3.1.4 Noise injection

The audio samples are already noisy hence using various methods to reduce the noise in the sample needs to be used. Adding more noise to the sample makes no sense as after the filtering process the sample will be restored to the original audio and cause redundancy in the dataset.

#### 3.1.5 Spectrogram masking

Spectrograms are basically visualizations of the audio data. Spectrogram masking suggests that we mask a certain portion of the spectrogram along the time axis, frequency axis or both randomly. This will not be very helpful as in real life scenarios, the audio samples will retain all the features. This will lead to overfitting the model with partially redundant information. Which will lead to higher validation loss and poor generalization.

#### 3.1.6 Generation of Permutations

Some studies suggested creating new samples by creating permutations of the sequence of respiratory cycles. This caused the model to overfit by poor generalization.

### 3.2 Filter Based Audio Augmentation

Since traditional audio augmentation methods cannot be employed to our dataset for various reasons, we came up with a solution of applying filters to generate samples with slight variations. This solves 3 major issues.

1. Less Redundancy:

We are using filters which will modify the data uniformly and generate spectrograms with nuanced differences. But

the filter process will not lead to loss of information in random regions of the spectrograms contrasting to the Spectrogram masking technique. Hence reducing the redundancy with minimum loss of data.

2. No additional noise:

Since a filter reduces noise in the signal, it helps us train the models to perform well on both clean signal as well as with ambient noise similar to real life scenarios hence improving generalization.

3. No miss classification because of Pitch shift:

The filters will only remove unwanted frequencies and abnormalities, this will not affect the crackles or healthy sounds to miss classify as high-pitched wheezes and vice-versa.

Filters used for augmentation are

1. Butterworth Bandpass filter

2. Harmonic-Percussive source separator

3. 3-layer filter Pipeline

Following are the generalization results for the preliminary iterations of the proposed models. We used the dataset directly without augmentation on a preliminary version of our ML model. The results showed a high difference in the validation accuracy and the training accuracy, same is the case with the validation loss and the training loss. This can be observed in Fig-1. This huge difference in the validation and training accuracy is reminiscent of poor generalization meaning high variance.

Upon performing the Filter Based Audio Augmentation, the results were clearly stating that the variance has reduced to a great extent in Fig-2. This reduced variance signifies that the model has generalized well.

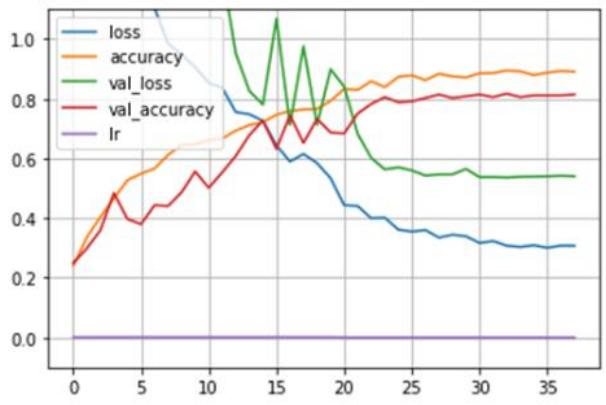


**Fig -1**: Accuracy and Loss on Imbalanced Dataset.

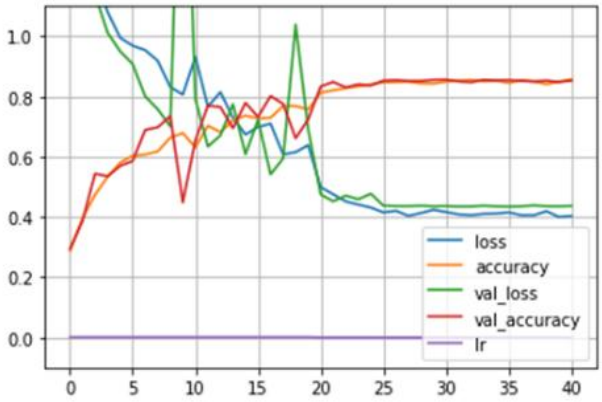


**Fig -2**: Accuracy and Loss on Balanced and Augmented Dataset.

## 4. SOLUTION

The main contribution of our work is implementation of an ensemble CNN model. The raw input sample i.e., the lung auscultation sounds from the dataset is split into training dataset and validation dataset. These samples are standardized by resampling, clipping and padding then passed through a Filter bank and the filtered output is passed on to the level one feature extractor which generates 2D images from the audio. These images are then passed to the level 2 feature extractor which is a fine-tuned CNN model. For feature extraction, the best suited methods for our dataset, has been identified to be Spectrogram, MFCC, and Chroma STFT.

### 4.1 Data Preprocessing

The process of preprocessing and feature extraction is divided into 3 parts i.e Cycle Extraction, Filtering and Feature Extraction. We standardized the audio clips coming from different sources by resampling them to 22050 Hz. the raw sample then undergoes segmentation to capture a single breathing cycle per clip and zero padding to make each clip 6 seconds long. If our sample consists of noise, it will lead us to inaccurate results. The audio is passed through the Harmonic-Percussive Source Separation Filter to reduce the ambient noise, Wavelet denoise filter, Butterworth band pass filter [60Hz to 1000Hz] to remove unwanted frequencies which may be caused by the heart beats. The Butterworth bandpass signal induces very small noise in the form of values close to zero for the zero padded segment of the audio clip. This is the result of the FFT algorithm on Zero values, hence this noise is attenuated. After these steps of filtering the audio samples are ready for classification.

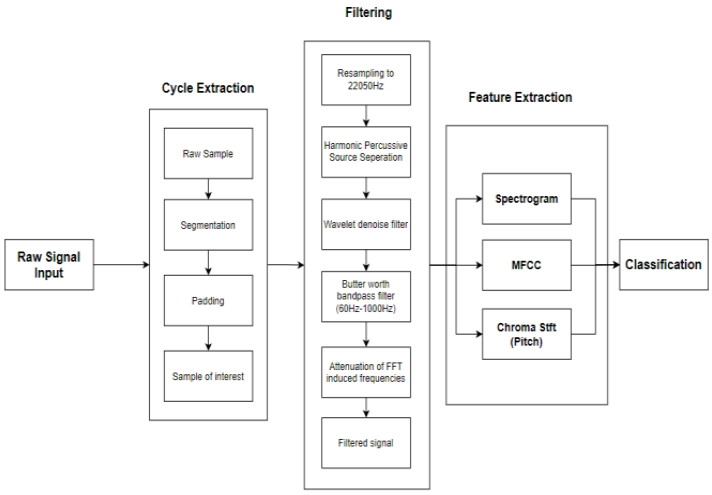


**Fig -3**: Audio Processing Pipeline

### 4.2 Feature Extraction

We selected Spectrogram, MFCC and Chromagram as the features. Each one has a significant use in the classification process. Spectrogram is a visual representation of the loudness of the audio at particular frame [32]. This is the simplest way to represent audio. Chromagram is visualization of the pitch class of the audio frame [22]. Pitch allows us to classify among wheezes and crackles. MFCC represent sound to the machine in same way as the human perceive the audio, which means higher sensitivity to low frequency sound and less sensitivity to higher frequency sound [28]. Hench in a nutshell, MFCC can be interpreted as a kind of spectrogram that represent sounds on nonlinear (logarithmic) scale knows as Mel-Scale.

The dataset consisted of audio clips ranging from 10 seconds to 30 seconds. We extracted Spectrogram, MFCC and Chromagram from once such audio clip which was 14 seconds long which can be seen in Fig-1, Fig-2 and Fig-3 respectively.

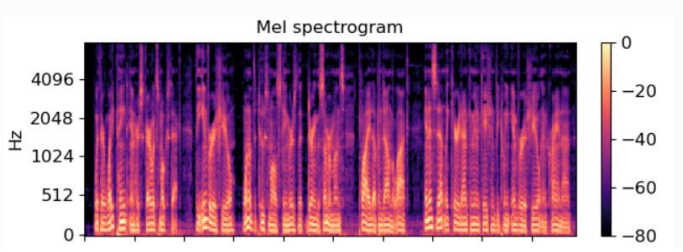


**Fig -4**: Spectrogram

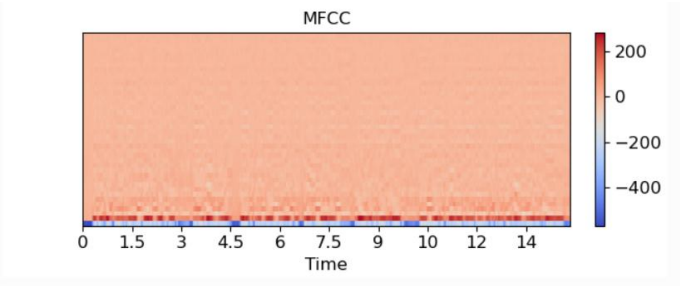


**Fig -5**: MFCC

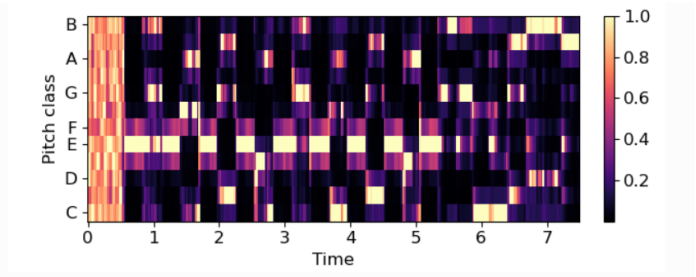


**Fig -6**: Chromagram

### 4.3 Classification

The proposed system uses 3 features to perform the analysis. This architecture allows us to gather information from different audio domains and incorporate them into one aggregated output. The Level 1 feature extraction is used to extract the feature images (visualization) from the raw audio. These images are passed on to the Level 2 feature extractors which are tailored CNN models which learn the essential features from the images. These models are concatenated together and passed on as input to a Multi-Layer Perceptron which gives final output.

Preliminary models suffered from overfitting issue. The train accuracy reached up to 99% while validation accuracy couldn't go past 85%. This issue was resolved using callbacks like Early Stopping with best weight and Learning Rate reduction.

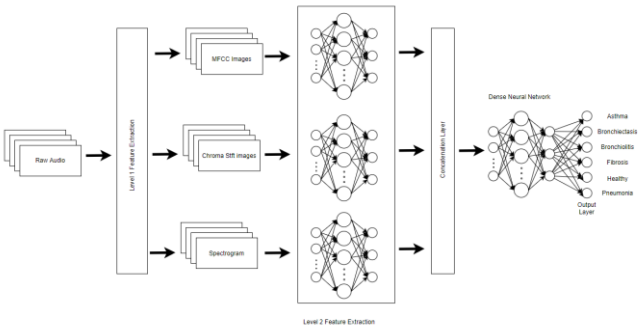


**Fig -6**: Model Architecture

## 5. Model Evaluation

For model evaluation, we trained multiple versions for a single kind of model and with slight variations to tune the model. We gathered compared the data of the best performing models. We segregated the results in three categories, namely MFCC based ML models, MFCC based DL models and lastly Multi Feature based DL models.

MFCC based ML models consist of traditional ML models like Decision Tree, KNN, SVM and Random Forest. The models were trained on flattened MFCC (1-Dimensional) vector. These models are based on a single feature. From Table-1 it was observed that Random Forest algorithm was giving best result among others of the same category. This pointed towards the possibility of achieving a better accuracy with an ensemble model in further research.

**Table -1:** MFCC based ML Model Comparison

| MFCC based ML Model Comparison Table | | |
|---|---|---|
| Model Name | Version | Validation Accuracy |
| Decision Tree | V0 | 0.7069 |
| Fine K-Nearest Neighbors | V1 | 0.8421 |
| Sequence Vector Machine | V0 | 0.7828 |
| Random Forest | V1 | 0.8580 |

MFCC based DL models consist of Single Layer Perceptron, Multi-Layer Perceptron, Convolutional Neural Networks, Recurrent Neural Networks and Long-Short Term Memory models. Some of the Models were used in a combination like RNN-LSTM and CNN+RNN-LSTM. From Table-2 it is observed that RNN-LSTM models are performing well along with the models containing CNNs. This pointed us in the direction that using a combination of multiple models in a kind of sequential pipeline is likely to give better results. Because V4 of CNN+RNN-LSTM model gave better accuracy as compared to purely CNN and RNN-LSTM models.

**Table -2:** MFCC based DL Model Comparison

| MFCC based DL Model Comparison Table | | |
|---|---|---|
| Model Name | Version | Validation Accuracy |
| Single-Layer Perceptron | V1 | 0.8628 |
| Multi-Layer Perceptron | V4 | 0.8549 |
| RNN-LSTM | V3 | 0.8531 |
| CNN + RNN LSTM | V4 | 0.8838 |
| CNN | V4 | 0.8634 |

Multi Feature based DL Models consisted of Deep learning models which are used in a particular combination, such that the input is taken from multiple features. These features were MFCC, Spectrogram and Chromagram as discussed earlier. This time we developed models that used CNNs in a sequence with different combination of Features. From Table-3 we observed that when the three features were used together, their validation accuracy was 92.76% and that of only MFCC and Chromagram was approximately 91.94%.

**Table -3:** Multi Feature based DL Model Comparison

| Multi Feature based DL Model Comparison Table | | |
|---|---|---|
| Model Name | Version | Validation Accuracy |
| CNN with MFCC and Chromagram Feature | V4 | 0.9194 |
| CNN with MFCC, Chromagram and Spectrogram | V3 | 0.9276 |

## 6. CONCLUSION

We used cycle extraction and filtering methods on our dataset to get better audio quality which will give better accuracy for our model. This dataset was then used to trained various traditional ML models like SVM, KNN, etc. which gave accuracy of more than 70% with random forest reaching up to 85%. On contrary DL based models gave even better results with accuracy of more than 85%. Few of these models were Single Layer perceptron, CNN, RNN, etc. Here RNN LSTM with CNN 1-D convolutional model gave us accuracy of 88%. Finally, we developed our model which is combination of MFCC, Chroma STFT and spectrogram as feature extraction technique with CNN as classification model which gave us best results of the highest accuracy reaching up to 92%. This can be inferred as, when multiple features of pulmonary audio with an ensemble CNN model are used, the results are much better as compared to the results of an approach with a single audio feature with a single ML/DL model. Which is in agreement with the works of D.Perna in [16] and [17].

Further, we plan to develop a diagnosis support system that allows the medical practitioner to perform lung auscultation and see the results in real time. The accuracy of the models will be increased as more data is collected over time with the expert feedback. More advancements in the field are possible when the digital stethoscopes become less expensive making it easier for more practitioners to adopt it.

## REFERENCES

[1] A. Dawadikar, A. Srivastava, N. Shelar, G. Gaikwad and A. Pawar, "Survey of Techniques for Pulmonary Disease Classification using Deep Learning," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/I2CT54291.2022.9824879.

[2] Haider, N.S., Singh, B.K., Periyasamy, R. *et al.* Respiratory Sound Based Classification of Chronic Obstructive Pulmonary Disease: a Risk Stratification Approach in Machine Learning Paradigm. *J Med Syst* **43,** 255 (2019).

[3] G. Altan, Y. Kutlu and N. Allahverdi, "Deep Learning on Computerized Analysis of Chronic Obstructive Pulmonary Disease," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 5, pp. 1344-1350, May 2020, doi: 10.1109/JBHI.2019.2931395.

[4] F. Demir, A. M. Ismael and A. Sengur, "Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure," in IEEE Access, vol. 8, pp. 105376-105383, 2020, doi: 10.1109/ACCESS.2020.3000111.

[5] S. Z. H. Naqvi, M. A. Choudhry, A. Z. Khan and M. Shakeel, "Intelligent System for Classification of Pulmonary Diseases from Lung Sound," 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2019, pp. 1-6, doi: 10.1109/MACS48846.2019.9024831.

[6] G. Chambres, P. Hanna and M. Desainte-Catherine, "Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis," 2018 International Conference on Content-Based Multimedia Indexing (CBMI), 2018, pp. 1-6, doi: 10.1109/CBMI.2018.8516489.

[7] Naqvi, S.Z.H.; Choudhry, M.A. An Automated System for Classification of Chronic Obstructive Pulmonary Disease and Pneumonia Patients Using Lung Sound Analysis. Sensors 2020, 20, 6512. https://doi.org/10.3390/s20226512

[8] S. Z. Y. Zaidi, M. U. Akram, A. Jameel and N. S. Alghamdi, "Lung Segmentation-Based Pulmonary Disease Classification Using Deep Neural Networks," in IEEE Access, vol. 9, pp. 125202-125214, 2021, doi: 10.1109/ACCESS.2021.3110904.

[9] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1207-1216, May 2016, doi: 10.1109/TMI.2016.2535865.

[10] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan and M. I. H. Bhuiyan, "A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 7, pp. 2595-2603, July 2021, doi: 10.1109/JBHI.2020.3048006.

[11] S. Kido, Y. Hirano and N. Hashimoto, "Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN)," 2018 International Workshop on Advanced Image Technology (IWAIT), 2018, pp. 1-4, doi: 10.1109/IWAIT.2018.8369798.

[12] L. Pham, H. Phan, R. Palaniappan, A. Mertins and I. McLoughlin, "CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 8, pp. 2938-2947, Aug. 2021, doi: 10.1109/JBHI.2021.3064237.

[13] Z. Tariq, S. K. Shah and Y. Lee, "Lung Disease Classification using Deep Convolutional Neural Network," 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 732-735, doi: 10.1109/BIBM47256.2019.8983071

[14] Fraiwan, Mohammad; Fraiwan, Luay; Khassawneh, Basheer; Ibnian, Ali (2021), "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope", Mendeley Data, V3, doi: 10.17632/jwyy9np4gv.3

[15] Rocha BM et al. (2019) "An open access database for the evaluation of respiratory sound classification algorithms" Physiological Measurement 40 035001

[16] Rocha, Bruno & Filos, D. & Mendes, L. & Vogiatzis, Ioannis & Perantoni, Eleni & Kaimakamis, Evangelos & Natsiavas, Pantelis & Oliveira, Ana & Jácome, Cristina & Marques, Alda & Paiva, Rui Pedro & Chouvarda, Ioanna & Carvalho, P. & Maglaveras, N.. (2017). A Respiratory Sound Database for the Development of Automated Classification. 33-37. 10.1007/978-981-10-7419-6_6.

[17] D. Perna, "Convolutional Neural Networks Learning from Respiratory data," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, pp. 2109-2113, doi: 10.1109/BIBM.2018.8621273.

[18] D. Perna and A. Tagarelli, "Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), 2019, pp. 50-55, doi: 10.1109/CBMS.2019.00020.

[19] L. Cai, T. Long, Y. Dai and Y. Huang, "Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis," in IEEE Access, vol. 8, pp. 44400-44409, 2020, doi: 10.1109/ACCESS.2020.2976432.

[20] Porieva, H. S., Ivanko, K. O., Semkiv , C. I. and Vaityshyn , V. I. (2021) "Investigation of Lung Sounds Features for Detection of Bronchitis and COPD Using Machine Learning Methods", Visnyk NTUU KPI Seriia - Radiotekhnika Radioaparatobuduvannia, (84), pp. 78-87. doi: 10.20535/RADAP.2021.84.78-87

[21] M. A. Islam, I. Bandyopadhyaya, P. Bhattacharyya and G. Saha, "Classification of Normal, Asthma and COPD Subjects Using Multichannel Lung Sound Signals," 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 0290-0294, doi: 10.1109/ICCSP.2018.8524439.

[22] X. Yu, J. Zhang, J. Liu, W. Wan and W. Yang, "An audio retrieval method based on chromagram and distance metrics," 2010 International Conference on Audio, Language and Image Processing, Shanghai, China, 2010, pp. 425-428, doi: 10.1109/ICALIP.2010.5684543.

[23] S. Yuan et al., "Improved Singing Voice Separation with Chromagram-Based Pitch-Aware Remixing," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 111-115, doi: 10.1109/ICASSP43922.2022.9747612.

[24] R. Hidayat, A. Bejo, S. Sumaryono and A. Winursito, "Denoising Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System," 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Bali, Indonesia, 2018, pp. 280-284, doi: 10.1109/ICITEED.2018.8534807.

[25] A. Winursito, R. Hidayat and A. Bejo, "Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2018, pp. 379-383, doi: 10.1109/ICOIACT.2018.8350748.

[26] P. Bansal, S. A. Imam and R. Bharti, "Speaker recognition using MFCC, shifted MFCC with vector quantization and fuzzy," 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI), Faridabad, India, 2015, pp. 41-44, doi: 10.1109/ICSCTI.2015.7489535.

[27] X. Shan-shan, X. Hai-feng, L. Jiang, Z. Yan and L. Dan-jv, "Research on Bird Songs Recognition Based on MFCC-HMM," 2021 International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China, 2021, pp. 262-266, doi: 10.1109/ICCCR49711.2021.9349284.

[28] M. A. Hossan, S. Memon and M. A. Gregory, "A novel approach for MFCC feature extraction," 2010 4th International Conference on Signal Processing and Communication Systems, Gold Coast, QLD, Australia, 2010, pp. 1-5, doi: 10.1109/ICSPCS.2010.5709752.

[29] A. Winursito, R. Hidayat, A. Bejo and M. N. Y. Utomo, "Feature Data Reduction of MFCC Using PCA and SVD in Speech Recognition System," 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam, Malaysia, 2018, pp. 1-6, doi: 10.1109/ICSCEE.2018.8538414.

[30] P. Mahesha and D. S. Vinod, "LP-Hillbert transform based MFCC for effective discrimination of stuttering dysfluencies," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 2561-2565, doi: 10.1109/WiSPNET.2017.8300225.

[31] Z. Chi, Y. Li and C. Chen, "Deep Convolutional Neural Network Combined with Concatenated Spectrogram for Environmental Sound Classification," 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2019, pp. 251-254, doi: 10.1109/ICCSNT47585.2019.8962462.

[32] R. Decorsière, P. L. Søndergaard, E. N. MacDonald and T. Dau, "Inversion of Auditory Spectrograms, Traditional Spectrograms, and Other Envelope Representations," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 1, pp. 46-56, Jan. 2015, doi: 10.1109/TASLP.2014.2367821.

[33] M. S. Towhid and M. M. Rahman, "Spectrogram segmentation for bird species classification based on temporal continuity," 2017 20th International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2017, pp. 1-4, doi: 10.1109/ICCITECHN.2017.8281775.

[34] J. -G. Leu, L. -t. Geeng, C. E. Pu and J. -B. Shiau, "Speaker verification based on comparing normalized spectrograms," 2011 Carnahan Conference on Security Technology, Barcelona, Spain, 2011, pp. 1-5, doi: 10.1109/CCST.2011.6095878.