Development of Classification Methods for Wheeze and Crackle Using Mel Frequency Cepstral Coefficient (MFCC): A Deep Learning Approach

Tinir Mohamed Sadi, Raini Hassan
Department of Computer Science, Kulliyyah of ICT, International Islamic University Malaysia, Kuala Lumpur, Malaysia
tinir.msj@gmail.com, hrai@iium.edu.com.my

Abstract— The most common method used by physicians and pulmonologists to evaluate the state of the lung is by listening to the acoustics of the patient's breathing by a stethoscope. Misdiagnosis and eventually, mistreatment are rampant if auscultation is not done properly. There have been efforts to address this problem using a myriad of Machine Learning algorithms, but little has been done using Deep Learning. A Convolutional Neural Network (CNN) model with Mel Frequency Cepstral Coefficient (MFCC) is expected to mitigate these problems. The problem has been in the paucity of large enough datasets. Results show 0.76 and 0.60 for recall for wheeze and crackle respectively and these number are set to increase with optimization and larger, more diverse datasets.

Keywords— deep Learning, convolutional neural network, mel frequency cepstral coefficient, respiratory sounds, adventitious sounds, sustainable development goals.

I. Introduction

For many reasons, there exists great disparity when it comes to accessibility in medicine. One of which is the lack of specialist pulmonologist to accurately identify symptoms of lung diseases. While this research does not aim to solve this inequality, it can act as a steppingstone to minimize this gap by utilising the latest technologies.

The most common method used by physicians and pulmonologists to evaluate the state of the lung is by listening to the acoustics of the patient's breathing using a stethoscope. This 100-year-old technique is referred to as auscultation. The outcomes and interpretations of these examinations are vastly subjective for multiple reasons. Humans are less sensitive to low frequency; environmental noise exists in rooms and some patterns of lung sounds are very similar. For these reasons, misdiagnosis and eventually, mistreatment are rampant if auscultation is not done properly. There have been efforts to address this problem using a myriad of Machine Learning algorithms, but little has been done using Deep Learning. While there are no solid attempts at completely automating lung sound analysis, there has been major progress.

This research's aim is to create a model to categorize lung sounds with a convolutional neural network approach into two categories: wheeze and crackle. This can aid untrained doctors and general practitioners to provide diagnosis in early stages of lung diseases.

II. Significance of Project

According to the World Health Organization, respiratory illness is one of the most common mortality factors

worldwide [1]. To put the problem into context, it is of paramount importance to acknowledge the gap that exists when it comes to accessing medical equipment. This is especially true in remote regions or in Less Economically Developed Countries (LEDC) [1].

In line with the United Nations Sustainable Development Goals, this project supports in realizing goal number 2. By making common lung diseases more easily diagnosable, we can ensure healthy lives and promote well-being for all at all ages.

Auscultation is a simple technique and generally inexpensive that can be performed by almost all doctors. Theoretically, the physician listens to breathing sounds in multiple locations of the chest - in the front and the back - and diagnoses immediately. Less common examinations include percussion, whereby the examiner taps on the patient's chest wall to produce sounds. It is probable that the experience and perceptual skills, or lack thereof, of doctors could lead to errors.

The objectives of this paper can then be split into two:

- To develop a Convolutional Neural Network (CNN) model as a classification tool for respiratory acoustics.
- To investigate the effectiveness of Deep Learning in classifying wheeze and crackle sounds.

As will be illustrated in subsequent sections, misdiagnosis and mistreatment are very common and often lethal. Diagnostic delays and misdiagnosis in for instance, Interstitial Lung Disease (ILD) at primary healthcare levels are relatively high. ILD is an umbrella term used for a large group of diseases that cause scarring (fibrosis) of the lungs. This is primarily due to reasons such as possible overlap with common endemics and lack of proper knowledge and diagnostic facilities.

III. Review of Previous Works

Previous studies on automated and computerised respiratory sound analysis have been conducted using several Machine Learning algorithms. This section provides a discussion of the few prominent works on computerised respiratory sound analysis [19].

Earlier attempts to automatically differentiate lung noises have wanted to simplify the problem by relying on a single type of lung sound or using an average of 5 to 20 patients with a small number of patients. Some of these experiments have used several sound recordings taken from the same individual, which also decreases the amount of variation in the data significantly. Working with a limited group of patients or concentrating on a specific element of lung sound, very high accuracy results can certainly be achieved since the algorithm can be handcrafted and carefully tailored to fit a limited number of patients' data and features collected. However, as the number of patients is extended to several dozen or several hundreds, the features learned from small datasets could not be generalized. [22].

Another important and fundamental obstacle in any work on classification of Machine Learning is the need to get data classified as "ground facts." In the case of pulmonary data, sounds from one single patient can be captured from multiple locations, and for analysis each sound file can be divided into multiple segments. Finding a qualified pulmonologist who can spend hours listening to thousands of files and mark them manually is then a logistical challenge. By using Machine Learning approaches developed for image classification, we can use unlabelled data to improve the accuracy of the identification of lung sound.

In [22] we see a comparative study between two Machine Learning algorithms, namely SVM and KNN. The dataset of choice for this work was the R.A.L.E database. The R.A.L.E database encompasses more than 70 recordings from numerous subjects that were recorded on the surface of the chest wall using a contact accelerometer (EMT25C, Siemens). These recordings were manually categorised into three different groups, namely normal pathology, airway obstruction pathology, and parenchymal pathology. Features were extracted by using MFCC through a one-way ANOVA and these were then separately fed into each algorithm. In the end, the classification accuracies of the SVM and K-nn classifiers were found to be 92.19% and 98.26%, respectively. A confusion matrix was also produced for analysis. These results are satisfactory; however, the dataset used is free from any form of environmental noise which makes it very dissimilar from real-life situations.

The HMM learnt temporal patterns of crackles, wheezes, normal sounds and crackles and wheezes. Then, the sounds were classified into four categories in of probable lung diseases: asthma, COPD, and pneumonia. Unlike other works, the feature set in this research was based on wavelet packet analysis characterizing data coming from the four sound classes. The respiratory audio was obtained from a competition, International Conference on Biomedical Health Informatics (ICBHI 2017) Challenge [26]. On average, the recognition rate was slightly over 50%. This is below the acceptable rate for use on real patients. Three possible enhancements can be proposed from here onwards, first augmenting the scarce ‘wheeze’ and ‘wheeze and crackle’ classes. Second, employing an amalgamation of spectral and wavelet features and third, to include a discriminative classifier, perchance making a synergistic framework.

In a separate study conducted, a robust Deep Learning framework was designed [7]. The experiments evaluated the ability for the model to classify sounds obtained from ICBHI 2017 Challenge. Pham et al. also highlighted the factors affecting the final prediction accuracy such as respiratory cycle length, time resolution, and network architecture. The novel CNN, called CNN-MoE, uses an array of different trained models. For the task of classifying respirator anomaly, the model demonstrated an accuracy of 0.80 and 0.86 for the 4-class and 2-class subtasks, respectively. For the second task of predicting respiratory disease, the system specificity and sensitivity were 0.83 and 0.96, respectively.

A comparison of five Machine Learning algorithms for 2-class classification (healthy/non-healthy) as well as a multi-class classification (healthy, COPD: basal lower lobe pneumofibrosis, COPD: diffuse pneumofibrosis, another pathology) was performed in [8]. In this work, classifiers of different types for the detection of lung diseases have been investigated and analysed. Namely, the classifier based on the K-nn method, based on the decision trees (DT), support vector method (SVM), Naive Bayesian (NB) classifier, and the logistic regression method were investigated. Poreva et al. used dataset containing only 134 patients which was divided into training and test subsets in the ratio of 85% and 15%. Then, a cross-validation method was employed for forming the training and test sets for teaching the analytical model in situations of inadequate preliminary data or irregular representation of classes. Eventually, the SVM classifier and the decision tree classifier are turned out as optimal with an accuracy rate of 88 and 77, respectively. The obvious weakness in this research is the use of an incredibly

small dataset and less obvious is the vague classification of the classes.

A research team in Turkey, used SVM and CNN to classify lung sounds into numerous classes by building their own stethoscope and recorded their own patients [9]. The recording consisted of 17,930 sounds from 1630 subjects. From a total of 8 datasets, 4 for SVM and 4 for CNN; 17,930 audio clips were split into two classes (normal or pathological), 14,453 audio clips into 13 classes (normal, rhonchus, squeak, stridor, wheeze, rales, bronchovesicular, friction rub, bronchial, absent, decreased, aggravation, or Long Expirium Duration (LED)), 15,328 audio clips into 3 classes (rale, rhonchus, or normal) and lastly, 17,930 audio clips were categorised into 78 classes. Feature extraction was done using MFCC enhanced with Short Time Fourier Transform (STFT) to find base value for accuracy. A spectrogram (800x600 RGBA then 28x28) was built using open source software and Pylab. Aykanat et al. concluded that spectrogram image classification with CNN works as well as SVM does. CNN and SVM algorithms were run comparatively to classify respiratory audio: (1) healthy versus pathological classification, (2) rale, rhonchus, and normal sound classification, (3) singular respiratory sound type classification, and (4) audio type classification with all sound types. Accuracy results of the experiments were found as (1) CNN 86%, SVM 86%, (2) CNN 76%, SVM 75%, (3) CNN 80%, SVM 80%, and (4) CNN 62%, SVM 62%, respectively [9]. Technically speaking, a few matters arise from this paper, downsizing of spectrogram might have affected results and duration of sounds were 8s to 16s, which causes too much variance. Another matter is that the researchers here used recording software and hardware that are different from well know online repos such as RALE and ASTRA. The data gathered, however, had little or no noise pollution, but it was gathered from a real world situation.

In the sixth paper, the authors applied CNN in attempting to detect asphyxia in infants [10]. The crying audio was obtained from Instituto Nacional de Astrofísica, Óptica and the asphyxia dataset from University of Milano-Bicocca. Features were extracted by MFCC and were fed into a CNN architecture consisting of a convolution layer, Rectified Linear Units (ReLU), max pooling layer, fully connected layer and softmax layer. I. M. Yassin et al. achieved a 94.3% accuracy in training set and 92.8% accuracy in testing set. Again, we see that the sound dataset is from controlled environment, making them unfit for real life practic

[11] compared Backpropagation (BP) and Learning Vector Quantization (LVQ) for lung sound recognition. The audio was attained from Linmann Repository. However, the dataset only contained 32 lung sounds which were divided into 8 tracheal sounds, 8 vesicular sounds, 8 crackle sounds and 8 wheeze sounds. After segmentation of audio and MFCC being used from feature extraction, BP had a 93.17% accuracy rate and 86.88% for LVQ. All five stages of MFCC were performed namely, frame blocking, windowing, Fast Fourier Transform (FST), Mel frequency wrapping and cepstrum coefficient using Discrete Cosine Transform (DCT).

Here, the researchers use hidden Markov models fed with Mel-frequency cepstral coefficients [12]. The reason out forward by the authors for choosing HMM is that it is used in speech which has variations just like lung sounds. The authors propose a methodology to classify the respiratory sounds into wheezes, crackles, both wheezes and crackles, and normal using the same dataset as some of the other literatures discussed, the ICBHI. The procedure consists of a noise suppression step using spectral subtraction followed by a feature extraction process. The input of the model consists of MFCCs extracted in the range between 50 Hz and 2,000 Hz in combination with their first derivatives. The method achieves performance results up to 39.37%, in compliance with the ICBHI score. Best official score of 39.56 achieved by class ensemble. Second best result with 6 states and full covariance matrix type yielded 0.4232 sensitivity, 0.5669 specificity and 39.37 official score. Sensitivity is decreasing, indicating that the classifier could not resolve adventitious sound types. Advanced noise suppression techniques can improve the overall score [12].

[13] applied a Semi-Supervised Deep Learning model to lung sounds. Along with that, 2 SVM classifiers were deployed, one to identify wheezes and one to identify crackles. Greedy forward feature selection to identify the best subset of autoencoder features. Performance was evaluated by computing ROC curve and associated AUC for 50 randomly generated sets of 5-fold cross-validation splits. As a result, ROC curves with AUCs of 0.86 and 0.74 for wheezes and crackles, respectively. The data was collected from 284 pulmonary patients from clinical sites in Maharashtra, India. They were labelled by specialists.

In [14] we see the use of a classification system based on a boosted decisional tree algorithm. The monoclass approach is a succession of two monoclass model, where each tree has two leaves: crackles and no crackles for the first one, then wheezes and no wheezes for the second one. The second kind of model is a multiclass model, were a unique tree have 4 leaves for the 4 different classes: normal, crackles, wheezes or both. the final decision is taken by keeping the highest prediction of the four, one for each possible class. Parameters of the boosted decisional tree model were empirically set, the maximum depth of trees is 3, the maximum of boosting iterations is 100 and the learning rate is 0.1. The model proved to not be good enough to classify adventitious sounds.

The literature review revealed that both the feature extraction method and the Machine Learning algorithm play major roles in the recognition of respiratory sounds. The common issue that arises is the lack of a reasonably large

dataset size in addition to absence of natural, environmental noises that exist. It also appears that the sound content in the ICBHI dataset seems to be difficult for a classification as pointed out by et al. Furthermore, the unbalanced data could also be a cause of bad predictions, because it affects the model training. Lastly, like some studies attempt, the records could be pre-processed to avoid noise and interference. This could give better result for the respiratory cycles classification, but it is necessary to prove that the symptom information is not modified by the filtering step.

Although supplementary studies are desired to improve the accuracy, sensitivity and specificity of this method before it can be efficiently realized in clinical and healthcare settings, the analysis of breath sound intensity and airflow can reveal the real physiological conditions of the airway system. Additional investigation into the changes in the behaviour and breathing patterns between breath phases would be advantageous for respiratory rate monitoring and could improve the diagnosis of respiratory illnesses in both clinical and research environments.

IV. Theoretical Background

A. Sound Analysis and Classification

Sound data is classified as an unstructured data, and unstructured data is by far the most suitable data for Deep Learning. Things like Self-driving Cars and face recognition are the by-product of Deep Learning systems on image data. Unstructured data means a lot of data points, with no mean significance to evaluate statistically. Well, any sort of standard pixel deviation is unlikely, so that images are unstructured. Audio data is a series of sequenced wave signals one after the other. Possibly the mean effect on the chart will not be evaluated.

B. Deep Learning and CNN

The training procedure for a CNN is comparable to a standard Neural Network using backpropagation. More specifically, Lecun et al. introduced error gradient to train the CNNs. In the first stage, information is propagated in the feed-forward direction through different layers. Salient features are obtained by applying digital filters at each layer. The values of the output are then computed. During the second stage, the error between the expected and actual values of the output is calculated. Backpropagating and minimizing this error, the weight matrix is further adjusted, and network is thus fine-tuned. Unlike other standard algorithms in image classification, the pre-processing is not frequently performed in CNNs. Instead of setting parameters, as is the case with traditional NNs, we just need to train the filters in CNNs. Moreover, in feature extraction, CNNs are independent of prior knowledge and human interference.

Deep learning is modelled after the anatomy of the brain. As shown in the figure above, the human neuron has Dendrites, Axon, cellular structures, and Synaptic holes in our brain. With the aid of synapse, the signal is transferred from the axon to dendrite. When dendrites receive the signals, the cell body will do some processing and then sends the signal back to the axon, the updated signal will be sent back to some other neuron and this cycle continues again and again

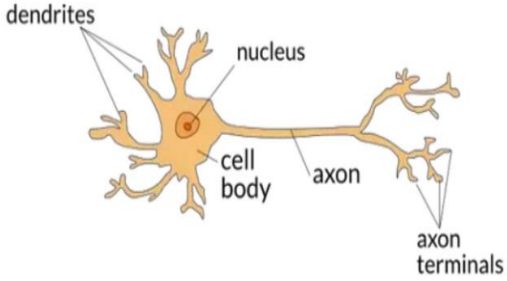

Fig. 1 A Human Neuron [23]

Inputs, weights, outputs, and activation functions are introduced, to replicate the magic in an artificial neural network. A single neuron is useless, but you can recreate the phenomenal operation when we have a bunch of them. The artificial neural network simply imitates the human brain's workings. Weighted inputs are fed to the neural layer, several processing is performed at the very same site that generates the output and is then fed to the next layer and this occurs recursively until the last layer is reached.

C. Adversarial Lung Sounds and Auscultation

The stethoscope was invented in 1816 and is used by physicians and pulmonologists to assess the lung function by listening to breath sounds using a method called auscultation. Using acoustic tests, specialist professionals in patients suffering from diseases such as Pneumonia, Pleural Effusion, pneumothorax, Chronic Obstructive Pulmonary Disease (COPD), and asthma can distinguish between normal and pathologic lungs. Auscultation is a fairly simple technique and generally inexpensive. Theoretically, the physician listens to breathing sounds in multiple locations of the chest - in the front and the back - and diagnoses immediately. Less common examinations include percussion, whereby the examiner taps on the patient's chest wall to produce sounds.

However, the outcomes and interpretations of these examinations are highly subjective, as humans are less sensitive to low frequency, environmental noise and pattern of lung sounds that are very similar. There is a direct correlation to the experience and perceptual skills of doctors and are hence prone to large errors.

1) Crackle sounds: Crackles are less commonly referred to as "crepitations" or "rales". They are short, discontinuous, and nonstationary sounds that can be detected at inspiratory and expiratory cycles. It is usually a sign of too much fluid in the lung. Some of the decisive features of crackles to be extracted include their duration, waveform, and timing. Its corresponding pathologies such as COPD, pneumonia, fibrosis, or bronchiectasis can be identified.

2) Wheeze sound: Wheezes are commonly referred to as adventitious continuous sounds and are heard towards the end of the inspiratory phase or in the early expiratory phase. They are often detected in patients affected with conditions that narrow the small airways in the lungs, such as asthma and COPD. Wheezes are drastically louder than crackles, they even can be heard without a stethoscope.

D. Mistreatment and Misdiagnosis of Lung Diseases

Auscultation is a fairly simple technique and generally inexpensive. Theoretically, the physician listens to breathing sounds in multiple locations of the chest - in the front and the back - and diagnoses immediately. Less common examinations include percussion, whereby the examiner taps on the patient's chest wall to produce sounds. There is a direct correlation [2] to the experience and perceptual skills of doctors and are hence prone to large errors.

Together with improvements in the stethoscope and its ability to record sounds there have been several attempts at slowly introducing them into the healthcare field.

Idiopathic Pulmonary Fibrosis (IPF) is a type of lung disease which causes chronic scarring. Among other symptoms, IPF can present itself as a wheeze and/or a productive cough [3]. A study examining the risk factors involved in delayed diagnosis found that community hospitals attributed to diagnostic delays in IPF and that patients were often misdiagnosed and treated before a final diagnosis of IPF was made. For reasons not mentioned in the study, male sex and older age were risk factors for patient delay and healthcare delay respectively [4].

Setting the problem of accessibility aside, there are financial implications of late-stage diagnosis of lung diseases. This problem manifests itself in two ways: when a patient has a delayed diagnosis and when they are misdiagnosed. Patients are subject to excessive substantial use of healthcare resources and costly and unnecessary diagnostic tests in repeated misdiagnosis situations, which could otherwise be used to treat patients who genuinely needed such a course of action. [17].

There also exist cases of overtreatment and overdiagnosis. A study looking at the economic impact of under and overdiagnosis of Chronic Obstructive Pulmonary Disease (COPD) in primary care, showed that over 50% of the present financial burden of the inhaled drugs is wasted to overtreatment and overdiagnosis, which could cover the cost for all underdiagnosed patients. This finding is astonishing given COPD is generally easily identifiable in computerised sound waves of patients' coarse crackles and expiratory wheezes (Jacome, 2015).

When it comes to the economic impact of over and underdiagnosing COPD patients in primary care, we find that overtreatment increases the financial burden of the disease as well as adverse events due to inhaled drugs overuse [5] . An observational study in Italy showed that 37.9% of COPD patients were receiving appropriate treatment, in 54.9% there was over-prescription. Over and above that, 66.8% were prescribed Inhaled corticosteroids (ICS) along with long-acting bronchodilators and 15.2% used ICS alone [19]. The evidence of benefit of ICS in COPD is limited by methodological problems and have not shown any survival benefit independent of the effect of long-acting bronchodilator. In another example of resource waste, a real-world study in the United Kingdom came to similar conclusions about primary cares in the region. It found that 53.7% of the total COPD population was receiving ICS [20].

E. Mel Frequency Cepstral Coefficient

Machine Learning ML extracts characteristics from the raw data and produces a rich content representation. Without the noise this allows one to know the key knowledge to draw inferences. One common method of extraction of audio features is the Mel-frequency cepstral coefficients (MFCC) that have 39 functions. The count of features is small enough to force us to know the audio content. 12 Parameters are in addition to the frequency amplitude. It provides us with sufficient frequency canals to analyse the audio. The diagram below is the flow of extracting the MFCC features.

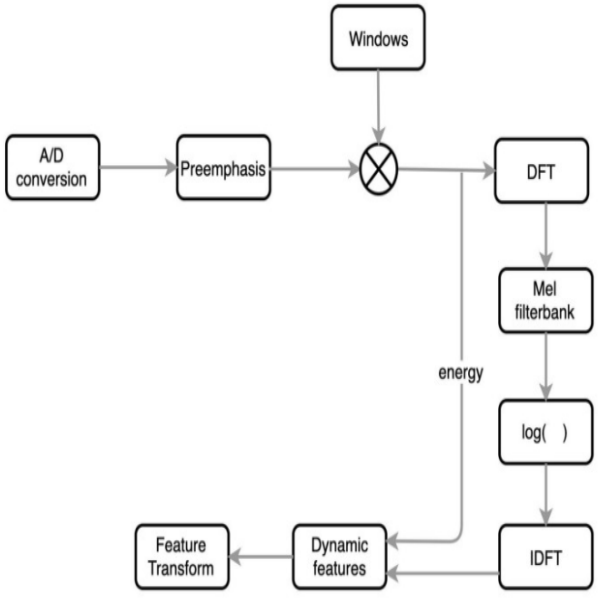

Fig. 2 Flow of extracting the MFCC features [24]

V. METHODOLOGY

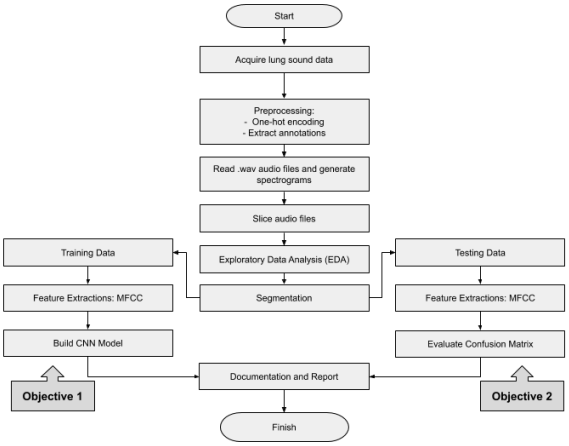

Fig. 1 Summary of methodology

A. Research Design

There are several parts to the initial stages of the data manipulation. First is labelling, a one-hot labelling scheme was used plus annotations were extracted followed by data pre-processing. Here the audio files were each cut into sub slices which is defined by the .txt files that accompanied the dataset. This was followed by Exploratory Data Analysis (EDA) and minimal processing. Feature extraction by utilising the MFCC.

Finally, the model was built and trained. In training, the Mel-Spectrograms were transposed and wrapped around the time-axis to allow the network to learn to identify features occurring at arbitrary times within the recording.

B. Data Description

The chosen dataset titled 'The Respiratory Sound Database' was used for this project. It was built as part of efforts to make large respiratory sound data sets accessible. Two research institutions in Portugal and in Greece developed the database. It consists of 920 recordings collected from 126 patients. The patients span all age groups - children, adults and the elderly. A total of 6898 respiration cycles were recorded making the total recording 5.5 hours. Of them, 1864 contain crackles, 886 contain wheezes and 506 contain both crackles and wheezes as shown in the diagram below in Table 1 [24].

TABLE I
SUMMARY OF DATASET LABELS

No label	3642
Crackles only	1864
Wheezes only	886
Crackles and Wheezes	506

The cycles were annotated by respiratory experts as including crackles, wheezes, a combination of them, or no adventitious respiratory sounds. The recordings were collected using heterogeneous equipment and their duration varied from 10 to 90 s. The chest locations (Fig. 4) from which the recordings were acquired, also provided and were indicated by the numbers in the figure. Noise levels in some respiration cycles were high, which simulated real life conditions [24].

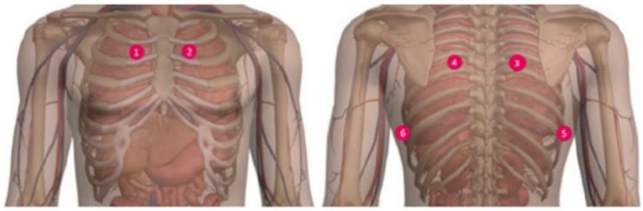

Fig. 4 Chest locations for the recording of respiratory sounds [25]

C. Exploratory Data Analysis

The data was visualised to explore and learn more about the relationships between attributes and observe trends. EDA also provides an overall image as to what the dataset constitutes. In the figure below (Fig. 5), it can be observed that respiratory cycles are generally between 2 to 5.

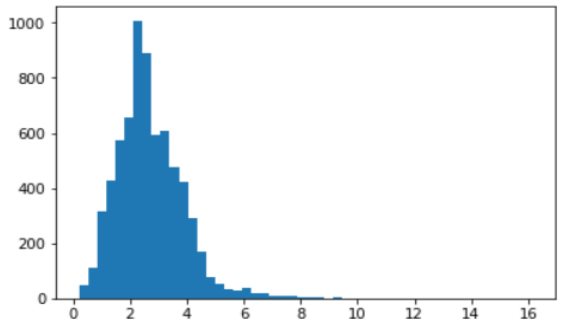

Fig. 5 Distribution of respiratory cycle lengths

It can be seen in the distribution below (Fig. 6) that Chronic Obstructive Pulmonary Disease (COPD) is the leading disease in this dataset while Asthma and Lower Respiratory Tract Infection (LRTI) are the least common.

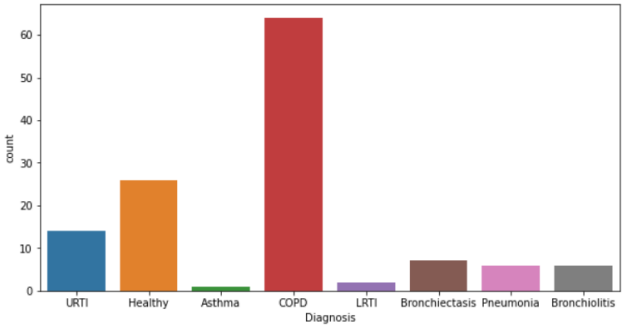

Fig. 6 Distribution of labelled diseases from dataset

D. Data Preparation

The dataset was relatively clean and required minimal edits. The labelling format of choice was one-hot encoding, other methods caused the learning rate to skyrocket and therefore produced undesirable divergent behaviour in the loss function.

The audio files were sliced as illustrated by the diagram below in Figure 7. This was done to ensure that essential features such as change in pitch or volume were not missed. Every deviation from normal sound projection is important in classifying the sounds.

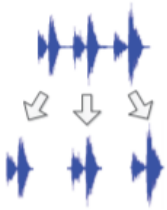

Fig. 7 Illustration of sound signal segmentation [11]

E. Convolutional Neural Network (CNN)

CNNs are one form of discriminative deep architecture and theses models have shown pleasing performance in processing two-dimensional data with grid-like topology, such as images and videos. The inspiration for this architecture comes from the animal visual cortex organization. Sometime in the 1960s, (Hubel and Wisel, 1962) proposed a concept called receptive fields. They found that the complex arrangements of cells were contained in the animal visual cortex responsible for light detection in overlapping and small sub-regions of the visual field.

A CNN is a multi-layer neural network that comprises of two different types of layers, i.e., convolution layers (c-layers) and sub-sampling layers (s-layers). C-layers and s-layers are connected alternately and form the middle part of the network.

VI. Analysis of Results

The CNN model was implemented on Keras on a TensorFlow backend with a batch size of 128 and an epoch of 15.

When analysing statistical model, the goal is to always create the most "accurate" model and increase the "accuracy". However, there are a myriad of metrics to choose from that can state different things about the model. Each metric has its pros and cons depending on the application. For this work, the planned use is for medical application. Therefore, the error, especially False Negatives, should be minimised. In short, if a patient's lung produces an irregular sound, it is vital that the model flags it as such and avoids classifying it wrongly as a healthy lung sound. On the contrary, if a patient's lung has no sign of irregularity, we would also like the model to classify it as healthy. The formula of all the metrics are shown below.

For this experimental work, the efficacy of the model was quantified by precision, recall, f1-score and the number of true values (support). Precision is the ratio of the True Positive to all the Positive results. Recall, occasionally referred to as Sensitivity or True Positive Rate, gives a measure of just how accurately the model can identify the relevant data. F1-scores are just the weighted average between precision and recall. Table 2 shows the confusion matrix of the model run on the *scikit-learn* library. For this study, a higher recall is desired because we would like to detect as many irregular sounds as possible.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

TABLE 2
SUMMARY OF RESULTS

	Precision	Recall	F1-score	Support
None	0.84	0.79	0.81	756
Crackle	0.65	0.76	0.70	379
Wheeze	0.63	0.60	0.62	178
Both	0.61	0.54	0.57	108

The first category identified by the model is the healthy lung sound labelled 'None'. The precision value gives an insight to the model's accuracy in classifying a sample as positive. An 84% precision means that the model was able to accurately classify most of the positive samples, i.e., normal breathing as such, and not misclassify an unhealthy lung sound sample as healthy. The high recall value indicates that the model accurately identifies healthy lung sounds. The harmonic mean of the recall and precision, or the F1-score of this class is relatively high. Support is simply the total number of audio samples that have been classified as true for that class. Thus, we observe that 756 files were classified as having normal breathing sounds, the highest among the classes.

If class has a low precision, but a very high recall, this is a sign that the algorithm is biased towards a positive class. That is the case for the 'Crackle' class. This is not necessarily harmful for our use case; further tests can confirm or deny the preliminary diagnosis. The main reason for this is because the dataset is imbalanced, the number of positive samples are not equal to the negative samples. There are several negative cases like wheeze sounds that might turn into false positives. On the other hand, there are fewer positive cases, which may become false negatives.

For the 'Wheeze' class, it has a similar precision and recall rate. Because F1-score scores are just the weighted average between precision and recall, it is also similar. This indicates

that the model is weary about classifying sounds as wheezes. A possible explanation for this is that there are only 886 labelled audio clips of wheezes compared to the 1864 crackle sounds.

In general, the produced results are still unfavourable for our use case. With further tuning, precision and overall accuracy are set to increase. It is observed, from the results that the 'Both' class has the worst overall performance and is more misclassified than the other classes. This indicates that the patterns revealed by the data coming from the 'Normal' class are comparable to all other classes. A possible solution to this problem is collecting more data form the other three classes, in a manner such that the differences are emphasized.

VII. Conclusions

Wheezes and crackle provide insight into the state of one's lungs in a non-invasive and nonspecialised way. These two acoustics plus a combination of these two can be useful in detecting early-stage lung disease. CNNs have shown great potential in classifying images, by creating visual representations of audio. The application can be expanded to sound classification as well. We have seen from the past works that CNN outperforms all other Machine Learning algorithms by a huge margin.

Acknowledgment

First and foremost, all praise and thanks are due to the Almighty, Allah Subhanahuwata'lla. The authors are deeply grateful to the Department of Computer Science.

References

[1] *The top 10 causes of death*, World Health Organization, 2017.

[2] H. Hafke-Dys, A. Bręborowicz, P. Kleka, J. Kociński, and A. Biniakowski, "The accuracy of lung auscultation in the practice of physicians and medical students," *Plos One*, vol. 14, no. 8, 2019.

[3] J. J. Swigris, A. L. Stewart, M. K. Gould, and S. R. Wilson, "Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives," *Health and Quality of Life Outcomes*, vol. 3, no. 1, p. 61, 2005.

[4] N. Hoyer, T. S. Prior, E. Bendstrup, T. Wilcke, and S. B. Shaker, "Risk factors for diagnostic delay in idiopathic pulmonary fibrosis," *Respiratory Research*, vol. 20, no. 1, 2019.

[5] D. Spyratos, D. Chloros, D. Michalopoulou, and L. Sichletidis, "Estimating the extent and economic impact of under and overdiagnosis of chronic obstructive pulmonary disease in primary care," *Chronic Respiratory Disease*, vol. 13, no. 3, pp. 240–246, 2016.

[6] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Classifying Heart Sound Recordings using Deep Convolutional Neural Networks and Mel:Frequency Cepstral Coefficients," *2016 Computing in Cardiology Conference (CinC)*, 2016.

[7] L. Pham, I. Mcloughlin, H. Phan, M. Tran, T. Nguyen, and R. Palaniappan, "Robust Deep Learning Framework For Predicting Respiratory Anomalies and Diseases," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020.

[8] A. Poreva, Y. Karplyuk, and V. Vaityshyn, "Machine Learning techniques application for lung diseases diagnosis," *2017 5th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 2017.

[9] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, 2017.

[10] A. Zabidi, I. Yassin, H. Hassan, N. Ismail, M. Hamzah, Z. Rizman, and H. Abidin, "Detection of asphyxia in infants using deep learning Convolutional Neural Network (CNN) trained on Mel Frequency Cepstrum Coefficient (MFCC) features extracted from cry sounds," *Journal of Fundamental and Applied Sciences*, vol. 9, no. 3S, p. 768, 2018.

[11] F. Syafria, A. Buono, and B. P. Silalahi, "A comparison of backpropagation and LVQ: A case study of lung sound recognition," *2014 International Conference on Advanced Computer Science and Information System*, 2014.

[12] N. Jakovljević and T. Lončar-Turukalo, "Hidden Markov Model Based Respiratory Sound Classification," *Precision Medicine Powered by pHealth and Connected Health IFMBE Proceedings*, pp. 39–43, 2017.

[13] D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, and R. R. Fletcher, "Application of semi-supervised deep learning to lung sound analysis," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016.

[14] 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.

[15] G. P. Cosgrove, P. Bianchi, S. Danese, and D. J. Lederer, "Barriers to timely diagnosis of interstitial lung disease in the real world: the INTENSITY survey," *BMC Pulmonary Medicine*, vol. 18, no. 1, 2018.

[16] C. Jácome and A. Marques, "Computerized Respiratory Sounds in Patients with COPD: A Systematic Review," *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 12, no. 1, pp. 104–112, 2014.

[17] A. Corrado and A. Rossi, "How far is real life from COPD therapy guidelines? An Italian observational study," *Respiratory Medicine*, vol. 106, no. 7, pp. 989–997, 2012.

[18] D. Price, D. West, G. Brusselle, K. Gruffydd-Jones, R. Jones, M. Miravitlles, A. Rossi, C. Hutton, V. L. Ashton, R. Stewart, and K. Bichel, "Management of COPD in the UK primary-care setting: an analysis of real-life prescribing patterns," *International Journal of Chronic Obstructive Pulmonary Disease*, p. 889, 2014.

[19] P. D. Muthusamy, K. Sundaraj, and N. A. Manap, "Computerized acoustical techniques for respiratory flow-sound analysis: a systematic review," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3501–3574, 2019.

[20] N. Ji, "Incomplete Image Filling by Popular Deep Learning Methods," thesis, UCLA, Los Angeles, 2019.

[21] J. Hui, "Speech Recognition-Feature Extraction MFCC & PLP," *Medium*, 14-Sep-2019. [Online]. Available: https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9. [Accessed: 07-Nov-2020].

[22] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho, and N. Maglaveras, "A Respiratory Sound Database for the Development of Automated Classification," *Precision Medicine Powered by pHealth and Connected Health IFMBE Proceedings*, pp. 33–37, 2017.

[23] S. Ntalampiras and I. Potamitis, "Classification of Sounds Indicative of Respiratory Diseases," *Engineering Applications of Neural Networks Communications in Computer and Information Science*, pp. 93–103, 2019.

[24] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R. P. Paiva, I. Chouvarda, P. Carvalho, and N. Maglaveras, "A Respiratory Sound Database for the Development of Automated Classification," *Precision Medicine Powered by pHealth and Connected Health IFMBE Proceedings*, pp. 33–37, 2017.