Lung Disease Classification using Deep Convolutional Neural Network

Zeenat Tariq, Sayed Khushal Shah, Yugyung Lee
School of Computing and Engineering, University of Missouri-Kansas City, USA
zt2gc@mail.umkc.edu, ssqn7@mail.umkc.edu, leeyu@umkc.edu

***Abstract*—The advanced technologies are essential to achieving the improvement of medicine. More specifically, an extensive investigation in a partnership among researchers, health care providers, and patients is integral to bringing precise and customized treatment strategies in taking care of various diseases. This paper aims to assess the degree of accuracy acceptable in the medical field by utilizing deep learning to publicly available data. First, we extracted spectrogram features and labels of the annotated lung sound samples and used them as an input to our 2D Convolutional Neural Network (CNN) model. Secondly, we normalized the lung sounds to remove the peak values and noise from them. For deep learning classification, publicly available data was not sufficient to conduct the learning process. Finally, we have created a deep learning model called *Lung Disease Classification* (LDC), combined with advanced data normalization and data augmentation techniques, for high-performance classification in lung disease diagnosis. The final accuracy obtained after the normalization and augmentation was approximately 97%. The proposed model paves the way for adequate assessment of the degree of accuracy acceptable in the medical field and guarantees better performance than other previously reported approaches.**

***Index Terms*—Data normalization, Data augmentation, Convolutional neural network, Lungs sound classification, Deep learning.**

I. Introduction

Lung sounds are the acoustic signals generated from breathing. An auscultatory method has been applied widely by physicians to examine lung sounds associated with different respiratory symptoms. The auscultatory method has been the easiest way to diagnose patients with respiratory diseases such as pneumonia, asthma, and bronchiectasis [1], [2]. However, it is a manual process, which takes a lot of time and creates a possibility of more or less accuracy due to the complexity of the sound patterns and characteristics. This may involve a high risk of missed data, leading to underdiagnosed or misdiagnosed results [3], [4]. The accuracy of auscultation is not always correct and reliable since it was found that in one of the studies, the residents were not able to identify 100% of wheezing sounds in a series of pulmonary disease sounds [5].

Machine learning plays an important role in classifying different types of sounds through multiple algorithms [6]. Deep learning is a branch of machine learning, which has attracted a lot of attention due to its high performance in prediction and classification. These learning techniques are among the fastest-growing fields nowadays in the area of audio classification [7]. These classifiers outperform humans due to the ability to ignore noise and memory issues.

In this paper, we have applied deep learning techniques for better classification of our results on the diagnosis of respiratory symptoms. We propose our model that is uniquely designed with a popular deep learning network, Convolutional Neural Network (CNN). Specifically, we introduce various advanced preprocessing techniques such as normalization and augmentation for an effective lung sounds classification. The classification is based on the spectrogram features that are extracted from the audio dataset. The traditional classification results vary due to the existence of noise in the audio samples, which are due to the environmental interference. The existing CNN approaches have adopted a different architecture and therefore obtaining an accuracy between 80 - 95% with very high consumption of memory, which are purely based on audio feature techniques. The dataset used for the experimentation is a public dataset provided for research in [8].

One of the challenges in the research was finding the data that is publicly available and cleaning the data that are not recorded properly and cannot be accepted if it is given as an input to a class. Because of directly recording audio from lungs, the audio samples may have some noise coming from the heart or any other sounds that exist in the body. To improve the accuracy, we have applied the data normalization technique on the original data to rescale the audio samples in a better position and average values for better accuracy.

Deep learning relies on large amounts of data. Due to limited amount of publicly available data, there is limited research progress in this field. To tackle this problem. we have proposed a solution that is known in deep learning field as *Data Augmentation* [9]. Therefore, to improve our results further, we needed larger amounts of data. For that purpose, we have applied our data augmentation techniques, which can help the CNN model report a better accuracy. Finally, our model was observed to outperform all other models that are already researched so far. Large amounts of data could not be experimented by other researchers while data augmentation made us stand out and outperform all other researches.

II. Related Work

Chen et al. [10] proposed a novel solution for lung sounds classification by using a publicly available dataset. The dataset was divided into three categories, i.e., wheezes, crackles and normal. They proposed a detection method using optimized S-transformed (OST) and deep residual networks (ResNets). They performed preprocessing on the audio samples by us-

978-1-7281-1867-3/19/$31.00 ©2019 IEEE

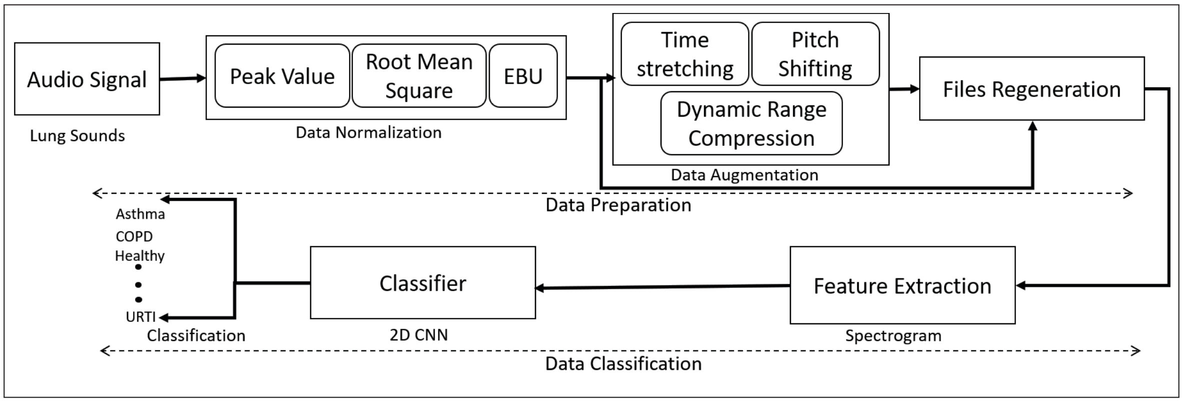

Fig. 1. Block Diagram for LDC System

ing OST, which rescaled the features for ResNets. Dalal et al. [11] has compared four methods of machine learning approaches for the purpose of lung sound classification using lungs dataset. CNN according to their experimentation outperformed all other classifiers. However, this all depend on the batch size and number of epochs. Although they have obtained an accuracy of approximately 97% but their machine utilization was very high by applying almost 1 million or more epochs. Rupesh et al. [12] have reviewed several features extraction and classification techniques for pulmonary obstructive diseases such as COPD and asthma. In their review, the feature extraction used were FFT, STFT, spectrograms and wavelet transform. The best accuracy that was reported for CNN was approximately 95% after all possible efforts. Chen et al. [13] proposed a solution for automatic early detection of a disease using CNN for heart and lungs. They collected data from volunteer patients, which were manually annotated by doctors for the consideration of experiments. The dataset was too limited to have any consequences for results. Salamon and Bello [14] presented the data augmentation technique for environmental sound classification using CNN. The deformation of audio was performed through stretching, pitch shifting, dynamic range compression, and background noise. Piczak [15] proposed a CNN model for classification of environmental sounds. Their 1D CNN architecture consists of two convolutional rectified layers by applying max pooling, two fully connected hidden layers, and a softmax output layer. The data was augmented through random time delays and pitch shifting. Mel spectrograms were extracted from all audio samples, resampled and normalized.

III. Method

A. Data Normalization

In this paper, we have evaluated existing normalization techniques and selected three best ones for the evaluation.

Root Mean Square Normalization In the Root Mean Square (RMS) Normalization, the amplitude level takes the average of a signal amplitude where it does not work as the arithmetic mean of a signal received. The RMS level is useful for finding the signal strength based on the amplitude regardless of positive or negative values of the signal. For a given signal, $x = x_1, x_2, \ldots, x_n$, the RMS value, x_{rms} is:

$$x_{rms} = \sqrt{\frac{x^2}{n}} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \ldots + x_n^2)} \tag{1}$$

The signal amplitude normalization can only be possible if we can figure out the scaling factor that can perform the linear gain change. There is a possibility to scale a signal with an amplitude that is higher than 1 or less than zero 0 decibels (db). For applying the linear gain change we can rearrange the above RMS level formula as shown in Equation 2 where R has a linear scale.

$$\begin{aligned} R &= \sqrt{\frac{1}{n}[(ax_1)^2 + (ax_2)^2 + \ldots + (ax_n)^2]} \\ a &= \sqrt{\frac{nR^2}{(x_1)^2 + (x_2)^2 + \ldots + (x_n)^2}} \end{aligned} \tag{2}$$

Peak Normalization In peak normalization, the peak signal level is analyzed in decibels relative to full scale (dBFS) and for the purpose of normalization, it amplifies the volume of the signal in such a manner that the output gets 0 dB maximum. This process can scale the amplitude of all input audio signals in such a way that the highest amplitude of the signal has a value of 1. The output signal based on above scaling can be mathematically calculated as

$$out = \frac{1}{max(abs(in))}.in \tag{3}$$

EBU Union Standard R128 Normalization European Broadcasting Union (EBU) Standard R128 Normalization focused on measuring the average loudness of a program in the normalization of audio signals.

B. Data Augmentation

We have experimented different types of data augmentation and concluded to experiment our results in three different ways such as time stretching, pitch shifting, and dynamic range compression [16]. Initially, the original data consists of 920 audio samples. After applying the augmentation techniques, the total audio samples obtained including the original audio samples were 11960. The files size that occupied the storage

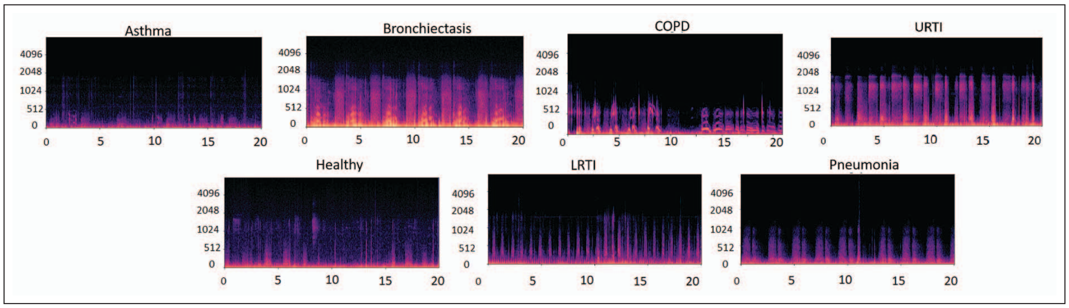

Fig. 2. Spectrogram Feature Extraction

was 26GB. For data augmentation, it is important to select the deformation patterns in such a way that the original labels are maintained and augmented.

Time Stretching For augmentation, the speed of the audio sample is changed and is increased or decreased by some factors [17]. We used four audio speeds, i.e., 0.5, 0.7, 1.2 and 1.5, along with the original audio sample files.

Pitch Shifting For data augmentation, the pitch of the audio samples are either decreased or increased by 4 values (semi-tones) [18]. The duration of the audio samples is kept constant similar to the original audio samples i.e.,4 - 10 seconds. The value changed in semitones ranged between -2, -1, 1, 2.

Dynamic Range Compression This technique compresses the dynamic range of the audio sample by four parameters. Among them, three are taken from Dolby E Standard and 1 is taken from ice cast radio live streaming server.

C. Network Model

Figure 1 shows the block diagram for the LDC system. Data Normalization and data augmentation techniques are applied to Lungs sound data where spectrogram features are extracted from the regenerated audio samples. These extracted features are passed to 2D CNN for classification. There are two main components of a convolutional neural network, i.e.,feature extractor and classifier. The feature extractor extract the spectrogram features from the audio signal and pass them to a classifier to classify the signals into their appropriate categories. The classifier consists of different convolutional and pooling layers, followed by linear activation and fully connected layers, which are used for classification purpose. The mathematical form of the convolutional layers can be found in Equation 4 and 5.

$$[x^l_{i,j,k} = \sum_a \sum_b \sum_c w^{(l-1,f)}_{i,j,k} y^{(l-1)}_{i+a,j+b,k+c} + bias^f] \qquad (4)$$

$$[y^l_{i,j,k} = \sigma(x^{(l)}_{i,j,k})] \qquad (5)$$

The output layer is represented by $y^l_{i,j,k}$ where as the 3-dimensional input tensor is denoted by i, j, k. The weights for filters are denoted by $w^{(l)}_{i,j,k}$ and $\sigma(x^{(l)}_{i,j,k})$ describes the sigmoid function for linear activation. The fully connected is layer is represented by Equation 6 and 7.

$$[x_i^{(l)} \sum_j w_{i,j}^{l-1} y_j^{l-1} + bias_j^{l-1}] \qquad (6)$$

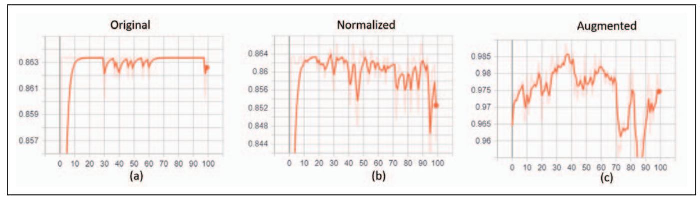

Fig. 3. Classification Accuracy: (a) Original, (b) Normalized, (c) Augmented

$$[y^l_{i,j,k} = \sigma(x^l_{i,j,k})] \qquad (7)$$

The 2D CNN architecture is composed of 5 layers. The first three are the convolutional layers, which are enclosed by max pool layer and finally they are followed two fully connected layers. We extracted librosa features for Mel spectrograms because for noise data spectrograms are considered as the best to differentiate between type of sounds. During the extraction of features, we have used window size and hope the size of 23 ms. As the sound clips vary between 3 to 10 seconds so that we kept the extraction to 3 seconds to make every bit of the sound clip usable. The input from the sound clips is reshaped and $X \in R^{128x128}$ shape is provided to the classifier.

The first layer takes the reshaped features as an input in the form of spectrograms with 24 filters. It takes the shape of [24x1x5x5]. The stride in this layer is [4x2] with ReLU as the activation function. The second layer has 48 filters of the shape [48x24x5x5] with [4x2] stride max-pooling layer and using ReLU as the activation layer. The third layer also takes 48 filters with receptive field [5x5] resulting in shape [48x48x5x5], and the activation is ReLU without pooling. Finally, the fourth layer has 64 hidden units resulting in shape [2000x64] with ReLU activation and [64x10] with softmax activation. In the top layer, we considered [5x5], which is a very small receptive layer due to the localized patterns.

IV. Experimental Design and Results

The dataset is composed of a total 5.5 hours of recording, which are further divided into recording samples of 126 patients. The categories include Asthma, Chronic Obstructive Pulmonary Disease (COPD), Healthy, Upper Respiratory Tract Infection (URTI), Lower Respiratory Tract Infection (LRTI), and Pneumonia. Table 1 shows the categories and the number of data in the dataset. We used librosa [19] for the spectrogram feature extraction. Figure 2 shows the features extracted from the Lung Sounds dataset for spectrograms.

TABLE I
ORIGINAL AND AUGMENTED DATA SIZE

ID	Name of Disease	#Audio File	#Augmented Audio File
1	Asthma	1	13
2	Bronchiectasis	29	377
3	COPD	785	10205
4	Healthy	35	455
5	LRTI	2	26
6	Pneumonia	37	481
7	URTI	31	403

TABLE II
LDC SYSTEM MODEL RESULT COMPARISON

Model	Technique	Accuracy
Model 1	Original Data	83%
Model 2	Peak Value Normalization	86%
Model 3	RMS Normalization	87%
Model 4	EBU Normalization	88%
Model 5	Augmentation applied on Original Data	93%
Model 6	Normalized Peak Value Augmentation	92%
Model 7	Normalized RMS Value Augmentation	94%
Model 8	Normalized EBU Value Augmentation	**97%**

We have designed our experiments to evaluate the proposed lung sound classification based on 2D CNN with the lungs sound dataset. The dataset is split into 70% and 30% for training/testing. The batch size was 32 and the number of epochs was fixed at 100 to avoid any over/under-fitting. The results of each instance for the LDC system is shown in Table II. It was observed during our experimentation stage that the highest accuracy achieved by the existing research is 97%, which is dependent on GPU usage and memory consumption. It can be seen from Table II that the LDC experimentation for Models 1-8. Although the data was not enough for the training, we were able to achieve good results.

Our model is experimented for 2D CNN classification network on the original dataset, which reported an accuracy of approximately 83%. Further, we have applied the three types of normalization i.e., Peak, RMS and EBU, and obtained an accuracy of 86%, 87% and 88%, respectively. The data augmentation is considered as the trend making technique in deep learning for small datasets. Further, the accuracy reported from the 2D CNN for the original data augmentation was 93%. We have also applied three augmentation techniques on normalized data and the highest accuracy achieved was 97%. Even though the data was not enough and had a lot of variation and environmental interference in recording (i.e.,heart beat, running fan), it was observed that our technique has achieved a very good accuracy in comparison with the state of the art research considering feature based approach.

Figure 3(a)-(c) represents the accuracy of 2D CNN network based on lungs dataset for original, normalized and augmented data. It was analyzed that when the data was in original form, CNN ran into overfitting and the highest accuracy reported was between 83%-86%. The accuracy reported for the models has little variations, which is due to the nature of the data. After normalization, we have noticed that the accuracy improved and ranged between 85%-90%. Finally, by applying augmentation we can see a visible increase in accuracy, which was reported approximately between 96%-99%. The result obtained during our experimentation out performs the method proposed in [11].

V. CONCLUSION

In this paper, we developed the Lung Disease Classification (LDC) system combined with advanced data normalization and data augmentation techniques, for high-performance classification in lung disease diagnosis. We have obtained 97% accuracy better than the state-of-the art accuracy. This confirms that the proposed model could be used for the diagnosis of lung diseases with lung sounds in health care.

REFERENCES

[1] E. Pacht, J. Turner, M. Gaillun, L. Violi, D. Ralston, H. Mekhjian, and R. John, "Effectiveness of telemedicine in the outpatient pulmonary clinic," *Telemedicine Journal*, vol. 4, no. 4, pp. 287–292, 1998.

[2] Y. Kahya, EC. Guler, and S. Sahin, "Respiratory disease diagnosis using lung sounds," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering*. IEEE, 1997, vol. 5, pp. 2051–2053.

[3] J. Kaur, K. Chugh, A. Sachdeva, and L. Satyanarayana, "Under diagnosis of asthma in school children and its related factors," *Indian pediatrics*, vol. 44, no. 6, pp. 425, 2007.

[4] A. Mandke and K. Mandke, "Under diagnosis of copd in primary care setting in surat, india," 2015.

[5] S. Mangione and L. Nieman, "Pulmonary auscultatory skills during training in internal medicine and family practice," *Am J respiratory & critical care medicine*, vol. 159, no. 4, pp. 1119–1124, 1999.

[6] J. Geiger and K. Helwani, "Improving event detection for audio surveillance using gabor filterbank features," in *Signal Processing Conference, 23rd European*. IEEE, 2015, pp. 714–718.

[7] L. Deng, D. Yu, et al., "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[8] BM Rocha, D Filos, L Mendes, I Vogiatzis, E Perantoni, et al., "A respiratory sound database for the development of automated classification," in *Precision Medicine Powered by pHealth and Connected Health*, pp. 33–37. Springer, 2018.

[9] I. Rebai, Y. BenAyed, W. Mahdi, and JP. Lorré, "Improving speech recognition using data augmentation and acoustic model fusion," *Procedia Computer Science*, vol. 112, pp. 316–322, 2017.

[10] H. Chen, X. Yuan, Z. Pei, M. Li, and J. Li, "Triple-classification of respiratory sounds using optimized s-transform and deep residual networks," *IEEE Access*, vol. 7, pp. 32845–32852, 2019.

[11] D. Bardou, K. Zhang, and S. Ahmad, "Lung sounds classification using convolutional neural networks," *Artificial intelligence in medicine*, vol. 88, pp. 58–69, 2018.

[12] R. Dubey and R. M. Bodade, "A review of classification techniques based on neural networks for pulmonary obstructive diseases," 2019.

[13] Q. Chen, W. Zhang, X. Tian, X. Zhang, S. Chen, and W. Lei, "Automatic heart and lung sounds classification using convolutional neural networks," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2016, pp. 1–4.

[14] J. Salamon, C. Jacoby, and JP. Bello, "A dataset and taxonomy for urban sound research," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 1041–1044.

[15] K. Piczak, "Environmental sound classification with convolutional neural networks," in *Machine Learning for Signal Processing, IEEE 25th International Workshop on*. IEEE, 2015, pp. 1–6.

[16] LR. Aguiar, Y. Costa, and NC. Silla, "Exploring data augmentation to improve music genre classification with convnets," in *2018 International Joint Conference on Neural Networks*. IEEE, 2018, pp. 1–8.

[17] S. Wei, K. Xu, D. Wang, F. Liao, H. Wang, and Q. Kong, "Sample mixed-based data augmentation for domestic audio tagging," *arXiv preprint arXiv:1808.03883*, 2018.

[18] N. Davis and K. Suresh, "Environmental sound classification using deep convolutional neural networks and data augmentation," in *Recent Advances in Intelligent Computational Systems*. IEEE, 2018, pp. 41–45.

[19] B. McFee, C. Raffel, D. Liang, D. PW Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, 2015, pp. 18–25.