A Systematic Review of Transfer Learning Methods for Identifying Lung Disease Sounds

Vishakha Pagi[1], Jayesh Mevada[2], Mehul S. Patel[3], Ankur J. Goswami[4], Rupal R. Chaudhari[5]

M.Tech Student, Department of Computer Engineering, Sankalchand Patel College of Engineering, Visnagar[1]
Department of Computer Engineering, Sankalchand Patel College of Engineering, Visnagar2,3,4,5

vishakhapagi023@gmail.com[1], jmmevada.ce@spcevng.ac.in[2], mspatel.it@spcevng.ac.in[3], ajgoswami.ce@spcevng.ac.in[4], rrchaudhari.ce@spcevng.ac.in[5]

Abstract: Lung disorders are now the leading cause of death throughout the globe. Despite this, most occurrences of lung illness are only identified at a late stage, when treatment options may be more limited. Technological advances are crucial to today's healthcare delivery system. This state-of-the-art medical research focuses on the value of analyzing lung sounds for the purpose of identifying lung diseases. The capacity to learn new material and use it in novel situations is crucial for patients to make their way through the healthcare system. Several Transfer learning techniques, like ALEXNET, VGGNET, and RESNET, are presented in this paper for classifying lung sounds. In addition to these methods, we will classify lung sound waves using a Transfer learning model that combines a Modified RESNET and a Mel spectrogram. Excellent performance in categorizing lung sounds by these transfer learning models suggests they may one day be employed in the diagnosis of respiratory disorders. In this evaluation, we will look at several Transfer Learning Techniques and talk about their pros and cons. And not even the worst part. To recognize four kinds of breathing noises. In addition, please provide suggestions about how the identification of lung sounds might be improved.

Keywords: Naive Bayes, Decision Tree, Support Vector Machine, Random Forest, Naive Bayes, Artificial Neural Network, AlexNet, VGGNet, RESNET.

I. INTRODUCTION

Lung disease is the third leading killer after heart disease and cancer throughout the world. The World Health Organization (WHO) reports that three million individuals each year lose their lives due to breathing problems. Over 200 million individuals across the world have COPD, and another 235 million have asthma [1,4]. Each year, 8.7 million people are diagnosed with tuberculosis. Pulmonary sound characteristics are reliable predictors of respiratory infections and diseases in this setting. [9.12]. Chronic obstructive pulmonary disease (COPD) and asthma are two of the world's leading killers. While 384 million people have chronic obstructive pulmonary disease (COPD), only 235 million people have asthma. During the duration of an asthma episode, symptoms including wheezing, chest tightness, trouble breathing, and coughing may fluctuate [15]. There are two types of lung sounds, normal and diseased. Two distinct types of breath sounds exist. A person with healthy lungs will make the same noises at each stage of the breathing cycle. It is possible that if you are having breathing problems, you will be able to pick up on more sounds than usual. A secondary respiratory sound is one that is generated in addition to the primary one by the lungs. As an example, apart from the constant sounds of breathing, there is another auditory phenomenon that happens [13].

Each of the topics mentioned below will be explored at further length in the sections that follow. Here, we will examine some of the most recent and noteworthy advances in speech recognition technology. In Section III, we discuss in detail the many techniques that were used to build this structure. In Section IV, we conduct in-depth analyses of many different subjects and draw parallels between them. Finally, the authors draw some firm conclusions from their research and provide suggestions for further research.

II. LITERATURE STUDY

In [1], Fatih Demir and colleagues present a study that delves into the categorization of lung sounds using a Convolutional Neural Network (CNN) model employing a Parallel Polling Structure. Their investigation primarily centers on harnessing deep learning methodologies to precisely classify lung sounds, a development with significant implications for the identification of respiratory conditions. In [2], Valentyn Vaityshyn explores the utility of pre-trained Convolutional Neural Networks (CNNs) in the classification of lung sounds. They are examining the potential of transfer learning to enhance the accuracy and efficiency of lung sound categorization. In [3], Zeenat Tariq presents an approach rooted in deep learning for the classification of lung diseases utilizing a Deep Convolutional Neural Network (DCNN). Their research is primarily directed towards the development of an effective system for automated lung disease diagnosis. In [4], Md. Ariful Islam and their team concentrate on the classification of individuals with normal, asthma, and COPD conditions based on multichannel lung sound signals. Their study underscores the significance of accurately distinguishing between different lung conditions to facilitate precise medical diagnosis. In [5], Joel Than Chia Ming examines the classification of lung diseases using diverse deep learning architectures and Principal Component Analysis (PCA). Their work explores various methodologies aimed at enhancing the accuracy of lung sound classification.

In [6], Anuradha D. Gunasingle places their focus on the early prediction of lung diseases. Their research seeks to develop a system capable of detecting lung conditions at an early stage, potentially enabling timely interventions. In [7], Syed Zohaib Hassan Naqvi proposes an intelligent system for the classification of pulmonary diseases based on lung sounds. Their study explores the application of intelligent systems to enhance the precision of pulmonary disease diagnosis. In [8], D. Jayaraj introduces a classification model for predicting lung cancer using Random Forests applied to computer tomography images. Their research centers on the utilization of machine learning techniques for predicting lung cancer. In [9], Funda Cinyol delves into the classification of lung sounds utilizing Convolutional Neural Networks. Their work underscores the adoption of deep learning models for categorizing lung sounds. In [10], Shreyasi Dutta presents an automated approach for analyzing lung sounds to detect pulmonary abnormalities. Their study investigates the implementation of automated analysis techniques for early detection of pulmonary issues.

In [11], Ramizraja Shethwala explores the classification of lung sounds, specifically wheezes and crackles, through the aid of transfer learning. Their research seeks to leverage transfer learning for improved classification accuracy. In [12], Truc Nguyen and Franz Pernkopf center their research on lung sound classification using a snapshot ensemble of Convolutional Neural Networks. Their study investigates ensemble learning techniques to enhance the classification of lung sounds. In [13], R. X. Adhi Pramono evaluates features relevant to the classification of wheezes and normal respiratory sounds. Their research offers insights into the selection of features that promote precise lung sound classification. In [15], Abdulkadir Sengu and Varun Bajaj present an efficient strategy for classifying lung diseases using Convolutional Neural Networks (CNNs). Their work is dedicated to improving the efficiency of systems for classifying lung diseases. In [16], Adnan Hassal Falah and Jondri propose a method for lung sound classification employing stacked Autoencoders and Support Vector Machines. Their research explores the application of deep learning techniques in conjunction with support vector machines to enhance the precision of lung sound classification. In [17], Gorkem Serbes develops an automated system for preprocessing and classifying lung sounds based on spectral analysis. Their study highlights the importance of effective preprocessing techniques in the analysis of lung sounds. In [19], H. Kamble conducts a frequency response analysis of respiratory sounds and carries out a comparative study on windowing techniques. This research primarily focuses on signal processing and the analysis of respiratory sounds to improve our understanding of the characteristics of lung sounds.

III. THE IDENTIFICATION OF RESPIRATORY SOUNDS

A. Gathering Dataset

The ICBHI 2017 database has 920 annotated audio recordings from 126 people, as stated on Kaggle [1,3,9]. Different stethoscopes were used to capture the sounds in this collection. You may choose the length of the recording from 10 seconds to 90 seconds, and the sample rate from 4000 Hz to 44,100 Hz. Each recording includes a set number of breaths, some introductory and concluding commentary, and the ability to identify crackles and/or wheezes. Here, we use database annotations to separate out individual breaths in audio recordings. The average time for a complete cycle is 2.7 seconds, however, it may range from 0.2 seconds to 16 seconds. The database has a total of 6898 breathing cycles, 3642 of which are regular, 1864 of which include crackles, 886 of which have wheezes, and 506 of which include both.

B. Processing Sound in the Lungs

The purpose of a noise reduction method is to either completely remove or significantly reduce the amount of noise present in an image [2,4,6]. Noise reduction techniques work by smoothing the picture generally while leaving the areas at the contrast limits alone. In contrast, these strategies may obfuscate little elements that have a low contrast. Cut and paste [8,9]: Before a full picture can be shown, it must be resized and maybe translated, and it must also be recognized as to what portion of the image can really be seen. It might be challenging to do this. Parts of the picture are obscured while others are not. Removed are just partly occluded lines and things. Clipping is the process of selecting which parts of a picture will be shown and which will be omitted. Clipping separates an item into visible pieces and those that are not. The scope of what can be seen now has been reduced. When something cannot be seen, it is written off as irrelevant.

C. Tools for Extracting Features from Audio

Waveform analysis in the form of a spectrogram [1,2,11-14] provides a visual representation of the intensity or sound of a signal over time by plotting the signal's amplitude against its frequency. The graph also displays the varying energy levels over time. Our proposed inside-outside model may be informed by spectrograms of our lung sounds, which can be generated via a top-down modeling approach. To create these spectrogram pictures, we utilized the excellent Virdis Color Map, whose colors span the spectrum from blue to green to yellow. The Mel Octave According to the Cestrum Coefficients [3,6,9,12], Mel Frequency is a certain pitch in music. For this assignment, we analyzed audio files by computing Cestrum Coefficients. MFCCs are crucial to the success of speech recognition systems in recognizing human speech. It has also been widely used in previous work on the detection of fake respiration sounds because it provides a measure of the short-term power spectrum of time domain data. Recognizing the many accidental noises that may arise in a single recording at various times and for varying lengths of time requires considering both the frequency and temporal content of the sounds. MFCC is helpful because it captures the evolving frequency content of a signal. The Mel scale is a subjective, nonlinear frequency scale that is used in the field of acoustics. Frequencies are assigned to octaves in the Mel scale according to a formula. MFCC generates a two-dimensional (time and frequency) feature vector that is converted to a one-dimensional array. Convolutional Neural Network [8,9,10,11]: Using this neural network classification approach, images are sorted into classes 24-28. This apparatus is known as a Convolutional Neural Network (CNN). In contrast to conventional neural networks, which assign separate weights to each input feature, CNNs use a shared parameter space for all features. This might help the network get insight into its local environment. Since CNN would pick up on the most vital characteristics without any human intervention, feature extraction is superfluous. CNN-based architectures employ data-driven convolutional kernels to construct a deep layered structure for extracting complicated features. CNN also uses very little processing power. Using convolution and pooling, we may distribute parameters and speed up calculations. When dealing with many frequency bands, as is the situation in most communication networks, it is helpful to convert from the time domain to the frequency domain using the Fourier transform (FFT) [2,18,19]. Furthermore, it has the potential to convert discrete data into continuous data that may be retrieved at varying rates.

D. Machine Learning Methods

Data might be linear or non-linear for use in Support Vector Machine (SVM) classification [2, 6]. Several different kernels are available to the user of an SVM Classifier. To classify new data points according to the values of the closest existing ones, we may use a technique called K-Nearest Neighbor [1-3], which is based on the idea that similar observations in a data collection are the ones that are physically nearest to a given data point. The user may adjust how many nearby observations the algorithm uses by changing a parameter K. Navier the Naive Bayes classifier uses Bayes theorem for statistical inference [2,4,7]. The probability of each category being accurate in the training data is calculated. Inverse probability is used to classify the test data. Consequently, it is possible to utilize the mean and variance to accurately forecast outcomes across a population. The key advantage of this classifier is that it can produce an accurate estimate of the mean and variance using just a small fraction of the training data. Naive Bayes classifiers are a collection of easy-to-understand probabilistic classifiers used in machine learning. They are based on Bayes theorem, which is itself based on the premise of feature independence. This classifier is very scalable since the number of parameters grows linearly with the number of predictors/features in the learning problem. As an ensemble classifier, Random Forests [9,11,19] generates decision trees at random. Bagging and a random sampling of variables are used to build the trees in a random forest. Each tree then votes for a class to which the instances should be allocated after the forest has been built. The winning category is the one that received the most votes. Several characteristics of this classifier make it well-suited for the task of classifying enzyme functions: a) It can be utilized successfully on large datasets without the necessity for pre-existing data normalization. The blanks in (b) pose no problem for it. Sequential Analysis [11,12,15]: To build a model for classification or regression, statisticians employ a tree-like structure called a decision tree. The dataset was further divided into subsets. Similarly, the decision tree associated with this issue is being built incrementally. The result is a structure that looks like a tree with branching-off points and terminal branches.

E. Transfer Learning Methods

The Alex-Net [1,5] findings show that a large, recurrent neural network (RNN may achieve outstanding performance on an extremely challenging data set using just supervised learning techniques. A year after AlexNet's debut, the ImageNet competition began, and all the entrants employed Convolutional Neural Networks. AlexNet was the first CNN, ushering in a new era of research. Despite the proliferation of deep learning frameworks, setting up AlexNet is still a breeze. Res-Net [2,10]: It is a dormant piece of infrastructure with a bypass connection that allows data to flow unhindered through the building. It takes in signal x and generates a signal F using a series of activation curve layers as intermediates (x). A skipped connection is analogous to this modification. In this configuration, the input signal x is compared to the reference signal F, and the differences between the two are described by the residual unit (x). Since the network will have already approximated the output function that creates data at that layer, the optimizer may reduce the weight of the remaining blocks at higher levels virtually to zero, enabling the signal to pass unaltered over the gap. This architecture, named VGG (Visual Geometry Group), is a multi-layer deep convolutional neural network [1,5,10] is a description of the Vgg-Net (DCNN). The difference between

VGG-16 and VGG-19 lies in the overall number of layers, which is 16. A cutting-edge model for object recognition has been built on top of the VGG framework.

IV.DIFFERENTIAL ANALYSIS

TABLE I
DIFFERENTIAL ANALYSIS OF FEATURES

Method	Advantage	Limitation
Mel Frequency Cestrum Coefficients [3,6,9,12]	To mimic the response of the human nervous system more closely, MFCCs frequency bands are arranged logarithmically.	When there is additive noise present, the MFCC results are not especially robust.
Fast Fourier transform [2,18,19]	Compression at low bit rates and compression of continuous tones benefit from this improvement.	Compression times and computation costs may increase.

TABLE II
DIFFERENTIAL ANALYSIS OF MODELS

Method	Advantage	Limitation/Disadvantages
Support Vector Machine [1-4,11,19]	Useful in situations involving several dimensions. is still useful even if there are more dimensions than data points to analyze them in. The decision function is memory-friendly since it uses just a subset of the available training data (the support vectors).	When the number of features exceeds the number of samples, it is more crucial than ever to choose the appropriate Kernel function and regularisation term to avoid over-fitting. SVMs need expensive five-fold cross-validation to assess accuracy instead of supplying probability estimates up front (for more on this, see Scores and probabilities).
K-nearest neighbor [2-8,12,18]	With KNN, a distance formula may be used since the only metric that has to be computed is the distance between two points based on data of various qualities. The model does not need a training period, so new information may be added whenever it is most practical.	Susceptibility to background noise and a lack of required data This approach fails when dealing with a large dataset since calculating distances between each data item is exceedingly time-consuming.
Naive Bayes [1,9-12,21]	Very easy to use and implement. It is easy to calculate the probability of an event occurring under a given set of conditions. A quick calculation of the chances is possible right away. This kind of training is efficient and quick. Possibly favorable if the hypothesis of conditional independence is correct.	Prerequisites for complete autonomy it is risky to draw broad conclusions. There are a lot of interdependent features. If a word is not part of the training data but appears in the test data for a particular class, the model may not assign it any class probability.
Decision Tree [14,17,19]	When compared to other techniques, pre-processing data for use in decision trees is simpler. A decision tree does not need data standardization to be used.	It is possible that decision tree computations will be far more complicated than those of any other approach. Given the time and effort involved, training a decision tree might end up being rather costly. It usually takes more time to train a model that uses a decision tree.

RF (Random Forest) [3-9,13,16,18]	When comparing the random forest technique to the decision tree approach, the latter falls short of the random forests forecast accuracy. With large data sets, it performs well. When it comes to machine learning, the rain forest algorithm is one of the most flexible and straightforward options available.	Compared to a decision tree method, this one takes much longer to complete.
Alex-Net [3,4,6,13]	In contrast to convolutional layers, which only depend on local spatial coherence and a narrow receiving field, fully connected layers may learn features from all conceivable combinations of the attributes of the layer below them.	Layer construction with many interconnections is computationally intensive.
Res-Net [1,2,11]	If you do not want to, you do not have to form any associations at all. It uses batch normalization, which boosts efficiency without compromising accuracy.	Implementation takes a considerable amount of time.
Inception-V3 [11,13,16]	Permits the use of any layer arrangement.	The training budget must be increased. The time spent calculating is seldom worth it.
Vgg-Net [5,10,11,12, 16]	Unfortunately, only 80% of the available parameters are accounted for.	The degree of precision is deteriorating with time.

V. CONCLUSION

The ability to recognize the distinctive characteristics of lung sounds is essential for their accurate diagnosis and categorization. When working with a huge dataset, however, it might be difficult to isolate individual patterns in attributes. The non-linear nature of environmental data makes standard techniques for discovering patterns and creating mathematical models worthless. In this study, we compare and analyze several aspects of noisy conditions. When MFCC fails, time wave late features may be able to restore functioning. However, unlike deep learning approaches, machine learning-based techniques are inefficient when dealing with huge datasets since they are slower and less accurate. Prospects: What We Can Count on Completely connected and soft max layers outperform conventional hard max layers when it comes to the disease-based classification of lung respiratory adventitious sounds using either the RESNET or ALEXNET transfer learning method.

REFERENCES

[1] Fatih Demir, Aras Masood Ismael, Abdulkadir Sengur, "*Classification of Lung Sounds with CNN model using Parallel Polling Structure*", IEEE Access Volume 8, 2020

[2] Valentyn Vaityshyn, Hanna Porieva, Anastasiia Makarenkova, "*Pre-trained Convolutional Neural Networks for the Lung Sounds Classification*", 39th International Conference on Electronics and Nanotechnology (ELNANO), 2019

[3] Zeenat Tariq, Sayed Khushal Shah, Yugyung Lee, "*Lung Disease Classification using Deep Convolutional Neural Network*", International conference on Bioinformatics and Biomedicine, 2019

[4] Md. Ariful Islam, Irin Bandyopadhyaya, Parthasarathi Bhattacharyya and Goutam Saha, "*Classification of Normal, Asthma and COPD Subjects using Multichannel Lung Sound Signals*", International Conference on Communication and Signal Processing, 2018

[5] Joel Than Chia-Ming, Norliza Mohd Noor, Omar Mohd Rijal, Rosminah M. Kassim, Ashari Yunus, "*Lung Disease Classification using Different Deep Learning Architectures and Principal Component Analysis*", 2nd International Conference on BioSignal Analysis, Processing and Systems, 2018

[6] Anuradha D. Gunasingle, Achala C. Aponso, Harsha Thirimanna, "*Early prediction of lung disease*", 5th International Conference for Convergence in Technology, 2019

[7] Syed Zohaib Hassan Naqvi, Dr. Mohammad Ahmed Choudary, Aleena Zaman Khan, Maheen Shakeel, "*Intelligent System*

for Classification of Pulmonary Diseases from Lung Sound" ,13th International conference on mathematics, Actuarial science, Computer science and Statistics, 2019

[8] D. Jayaraj, S. Sathiamoorthy, "Random Forest-based Classification Model for Lung Cancer Prediction on Computer Tomography Images", Second International Conference on Smart and Inventive Technology (ICSSIT 2019), 2019

[9] Funda Cinyol, Huseyin Emre Mutlu, Ugur Baysal, "Classification of Lung Sounds with Convolutional Neural Network", 21st National Biomedical Engineering Meeting, 2017

[10] Shreyasi Dutta, Anirban Dutta Choudary, Parijat Deshpande, Sakyajit Bhattacharya, Arpan Pal, "*Automated Lung Sound Analysis for Detecting Pulmonary Abnormalities*", 39th Annual International Conference of the IEEE Engineering in Medicine & Biology society, 2017.

[11] Ramizraja Shethwala, Shital Pathar, Tejal Patel, Prasann Barot, "*Transfer Learning Aided Classification of Lung Sounds*" Wheezes and Crackles Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021), 2021

[12] Truc Nguyen, Franz Pernkopf, "*Lung Sound Classification Using Snapshot Ensemble of Convolutional Neural Networks*", 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),2020

[13] R. X. Adhi Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "*Evaluation of features for classification of wheezes and normal respiratory sounds*", PLoS One, vol. 14, no. 3, pp. 1–21, 2019, doi: 10.1371/journal.pone.0213659.

[14] D. Jayaraj, S. Sathiamoorthy, "*Random Forest-based Classification Model for Lung Cancer Prediction on Computer Tomography Images*", Second International Conference on Smart and Inventive Technology (ICSSIT 2019), 2019

[15] Fatih Demir, Abdulkadir Sengu, Varun Bajaj, "*Convolutional neural networks based efficient approach for classification of lung diseases Health Information Science and Systems*", 2019

[16] Adnan Hassal Falah, Jondri, "*Lung sounds classification using stacked Autoencoder and Support Vector Machine*", 7th International Conference on Information and Communication Technology (ICoICT), 2019

[17] Gorkem Serbes, Sezer Ulukaya, Yasemin P. Kahya, "*An Automated Lung sound Preprocessing and Classification System Based on Spectral Analysis*", International Conference on Biomedical and Health Informatics, 2017

[18] Shreyasi Dutta, Anirban Dutta Choudary, Parijat Deshpande, Sakyajit Bhattacharta, Arpan Pal, "*Automated Lung Sound Analysis for Detecting Pulmonary Abnormalities*", 39th Annual International Conference of the IEEE Engineering in Medicine & Biology society, 2017.

[19] H. Kamble, "*Frequency Response Analysis of Respiratory Sounds and Comparative Study for Windowing Techniques*" pp. 210–215, 2016.