IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Deep Learning Approaches For Lung Disease Detection Through Voice Analysis

Dr.Y.Sowjanya Kumari
Associate Professor
Department of CSE
QISCET ,Ongole, India

Abstract: Lung disease identification is an important field of research in healthcare, and non-invasive approaches such as human voice analysis have gained traction. This study investigates the use of deep learning approaches, such as 1D Convolutional Neural Networks (1D CNN), Convolutional Neural Networks(CNN), Long Short-Term Memory networks(LSTM), and Gated Recurrent Units (GRU), to automatically diagnose lung disorders using human voice data. The suggested models, which extract relevant features from voice recordings, aim to discover patterns linked with various lung disorders such as COPDchronic obstructive pulmonary disease(COPD) ,pneumonia, and URTI. The 1D CNN and CNN are used to extract features and recognise patterns from audio inputs, while the LSTM and GRU networks are used to capture temporal relationships and sequential patterns. A dataset of speech samples labelled with the respective lung disease categories is utilised to train and evaluate the models. The performance of each deep learning architecture is measured using accuracy. The results show that these models are successful at recognising lung disorders, laying the groundwork for non-invasive, early-stage diagnosis via speech analysis. This study highlights deep learning's potential voice analysis as a viable tool for the rapid and precise diagnosis of lung disorders, perhaps resulting in earlier intervention and better patient outcomes.

***Index Terms* -** 1DCNN,CNN,LSTM,GRU,COPD.

I. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) in healthcare have transformed the way diseases, especially respiratory disorders, are diagnosed and treated. Significant public health issues are presented by Chronic Obstructive Pulmonary Disease (COPD) and other pulmonary illnesses, which are frequently marked by high morbidity rates and delayed diagnosis. Even if they are efficient, traditional diagnostic techniques like spirometry and imaging can be resource-intensive and may not always yield results quickly [1]. Novel directions for non-invasive diagnostics have been made possible by recent developments in the field of voice biomarker research. Research has indicated that vocal characteristics, including speech and cough sounds, can function as trustworthy markers of respiratory health [2, 3]. For example, speech pattern analysis has demonstrated potential in identifying respiratory disorders such as COVID-19 and asthma in addition to healthy persons [4]. This method allows for early detection and monitoring by utilising the distinct audio signatures linked to different lung diseases [5]. AI-based diagnosis models are more reliable when specialised algorithms, like the Bias-Free Network (RBF-Net), are developed to reduce the influence of confounding variables in cough audio data [3]. Additionally, using Convolutional Neural Networks (CNNs), a type of deep learning architecture, has shown promise in the classification of lung disorders using information gleaned from respiratory sounds [6, 7]. Chronic obstructive pulmonary disease (COPD), interstitial lung disease, and asthma are examples of respiratory disorders that pose a substantial global health burden and cause millions of fatalities each year [28, 30]. According to the World Health

Organisation (WHO), these illnesses are the main causes of morbidity and mortality, thus getting a diagnosis as soon as possible is essential to improving patient outcomes [31, 30]. Even if they are useful, traditional diagnostic procedures like spirometry and imaging typically have drawbacks such as high costs, the requirement for specialised equipment, and accessibility issues, especially in settings with low resources [30]. The use of audio-based analysis for respiratory disease diagnosis has gained popularity in recent years. Lung sounds (LSs) can be non-invasively recorded using electronic stethoscopes and offer important insights on lung health [30]. Together with advances in machine learning (ML) techniques, the analysis of these sounds presents a promising path towards the development of automated diagnostic systems that can improve the precision and efficacy of disease classification[32, 30]. Numerous research have shown how ML algorithms can be used to classify respiratory diseases using audio data. For example, [30] created a computer-based method that used lung sound data to categorise different pulmonary disorders with an amazing accuracy of 99.72% [30]. In a similar vein, [33] investigated the potential of voice as a non-invasive biomarker for respiratory illness diagnosis [28] by using voice attributes taken from speech recordings to categorise COPD patients. The field still faces obstacles in spite of these developments, such as the requirement for reliable datasets, efficient feature extraction techniques, and the incorporation of these technologies into clinical practice [32, 34]. With a focus on voice biomarkers and their uses, this paper provides a thorough assessment of the state of AI-driven respiratory disease diagnosis today. It looks at a number of approaches, such as the integration of multimodal frameworks for early diagnosis, the application of machine learning to analyse lung sounds, and the possibility of voice analysis as a clinical decision support system [1,8]. This work intends to showcase the revolutionary potential of AI in improving respiratory healthcare outcomes and addressing the urgent demand for novel diagnostic solutions by synthesising findings from current studies. In conclusion, the field of medical technology that lies at the nexus of AI, vocal biomarker research, and respiratory disease detection is exciting and has the potential to improve patient care by enabling prompt and precise assessments [2,3,5]. The purpose of this research is to add to the increasing corpus of knowledge by examining how well machine learning methods categorise respiratory disorders using audio analysis.

This introduction offers a thorough summary of the study topic's importance, the difficulties with conventional diagnostic techniques, and the potential benefits of machine learning and audio-based analysis for better respiratory disease diagnosis.

Related work was provided in Section 2, proposed work was offered in Section 3, results and discussion were covered in Section 4, and conclusion was covered in Section 5.

II. RELATED WORK

In recent years, there has been a noticeable increase in the use of AI and ML in the diagnosis of respiratory disorders. The results of numerous studies investigating the application of voice biomarkers, cough audio analysis, and deep learning(DL) techniques for the detection and categorisation of respiratory disorders are summarised in this overview of the literature.

The possible of vocal biomarkers as markers for identifying and tracking respiratory disorders has been brought to light by recent study. Demonstrated how speech biomarkers may accurately categorise pathological situations, especially respiratory disorders, by introducing a fresh environment for data gathering and analysis using articulatory speech tasks [2]. The study highlights the value of feature extraction from voice recordings, which can reveal information about a person's underlying medical issues. Parallel to this, Kim et al. (2023) [9] created an AI model to differentiate in the middle of voices in good health and those impacted by laryngeal illnesses, such as vocal cord paralysis and laryngeal malignancy. Their results show that voice analysis can be used as a non-invasive diagnostic tool, with accuracy rates in multiclass classification tasks ranging from 75% to 97% [9]. This emphasises how voice analysis can be used to identify and diagnose respiratory-related disorders early on.

The study of cough sounds has become essential for the diagnosis of respiratory disorders. showed how specific acoustic characteristics in cough signals can be used to track and identify illnesses including COVID-19, asthma, and COPD [3]. Their research focusses on creating AI-enabled models that can distinguish between non-cough and cough sounds, detecting respiratory ailments with high accuracy [3]. Additionally, [13] investigated the spectral fingerprints in cough noises, finding distinct frequency bands linked to a range of respiratory conditions. Ghrabli et al.[13] used machine learning algorithms to categorise cough noises in their study, and the results showed a substantial difference between coughs that were COVID-19 and those that weren't. [13] The viability of employing cough audio as an affordable diagnostic tool is demonstrated by this study. Promising outcomes have been shown when deep learning techniques are applied to the detection of respiratory diseases. used machine learning algorithms to identify aberrant lung function from speech recordings, and they were remarkably accurate in this regard [14]. Their work opens the door for telehealth solutions by demonstrating the possible of speech data as a stand-in measure for lung function. In a different study, [15] achieved good classification accuracy for a variety of lung disorders by using the Tunable Q-factor Wavelet Transform (TQWT) for lung signal disintegration and extraction of statistical characteristics. According to their research, sophisticated signal processing methods can improve machine learning models' capacity to diagnose respiratory disorders.

[1] suggested a multimodal framework for the early diagnosis of COPD that combines lung sounds and CT scan pictures. Their method, which shows increased diagnosis accuracy, combines the extraction of textural features from CT scans with acoustic analysis of lung noises [1]. This emphasises how crucial it is to combine many data modalities in order to strengthen the resilience of diagnostic models. Yan et al. [22] used the Cambridge COVID-19 Sound database to create a deep learning(DL) model for COVID-19 identification from voice recordings. Their method includes training models such as LSTM, CNN, and Hidden-Unit BERT(HuBERT) by extracting voice parameters including Mel-spectrograms and Mel-frequency cepstral coefficients(MFCC). The HuBERT model obtained an accuracy of 86% and an AUC of 0.93, indicating its potential as a non-invasive screening tool for COVID-19 [22].

The experiences of people with Acute exacerbations of mild to severe Chronic Obstructive Pulmonary Disease(AECOPD) were investigated in [23]. They were able to identify themes about the way AECOPD affects daily living and the belief that pulmonary rehabilitation (PR) is necessary through semi-structured interviews. The significance of person-centred interventions in enhancing the management of AECOPD is underscored by their findings. Machado and associates. [23]

[24] presented a deep learning architecture for the classification of several lung disorders, like pneumonia, COVID-19, and COPD, utilising data from digital stethoscopes. Their model, EasyNet, proved highly accurate in identifying respiratory disorders, underscoring the usefulness of audio data as a real-time diagnostic tool [25].

A systematic assessment of 75 papers concentrating on audio analysis for respiratory disease diagnosis was carried out by [26]. The experiments were divided into three categories: voice/speech analysis, lower respiratory symptom identification, and cough detection. The review stressed the promise of machine learning in automating respiratory disease detection and observed a notable surge in research linked to COVID-19 diagnosis, especially during the pandemic. Kapetanidis along with others [26] . [27]presented a method for classifying respiratory audio data using LSTM neural networks(NN). Their finding explains the efficacy of deep learning models in analysing sequential audio data for medical diagnostics, with an accuracy of 98.82% in the diagnosis of lung disorders. Zhang and associates [27].

Chudasama et al.'s research [35] further highlighted how crucial optimised classifiers are for interpreting respiratory sounds. According to their research, using machine learning classifiers in conjunction with audio processing techniques could greatly improve the accuracy of respiratory disease identification. [35] compared deep learning and conventional methods for voice-based pathology identification from respiratory sounds. Using the ICBHI 2017 and Coswara as benchmark datasets, their study revealed that the CNN-based method and Random Forest algorithm performed better, underscoring the promise of speech analysis in respiratory disease detection[35].

[30] created an automated approach based on machine learning to categorise lung disorders based on lung sound data. To improve classifier performance, they used denoising techniques such variational mode decomposition and discrete wavelet transform(DWT). The random forest classifier outperformed the decision tree, k-nearest neighbour(KNN), linear discriminant analysis(LDA), and random forest(RF) classifiers in the study, which compared them all. Its astounding accuracy of 99.72% was attained. [30] A thorough analysis of machine learning techniques for audio-based lung disease identification was presented in [34]. The authors highlighted a range of strategies, algorithms, and datasets utilised in the field, highlighting the necessity for more study to tackle the obstacles and constraints present in the current methodologies [34].

In order to comprehend the experiences of patients dealing with pulmonary hypertension linked to interstitial lung disease (PH-ILD), [31] carried out qualitative research. Significant symptoms like exhaustion and dyspnoea were found in their research, underscoring the need for better patient-centered care and enhanced communication across healthcare systems[31]. [32] investigated the use of DL/ML approaches for the classification and characterisation of respiratory sounds. Their research revealed the potential of these techniques for the prompt identification and categorisation of respiratory disorders, indicating that these models may be used as low-cost screening instruments[32].

[28] looked at the classification of Chronic Obstructive Pulmonary Disease (COPD) using voice characteristics as a digital biomarker. By creating a fresh dataset and using a variety of machine learning

models, the study was able to differentiate between COPD and healthy controls with encouraging results[28].

Further study is required to validate these methodologies, as the current literature indicates a deficit in the systematic use of audio-based analysis across varied demographics and situations. The examined literature emphasises how AI and machine learning have the potential to revolutionise respiratory disease diagnostics. These studies collectively contribute to the development of novel, non-invasive diagnostic methods, ranging from voice biomarkers and cough audio analysis to deep learning approaches and multimodal frameworks. As research advances, the application of these technologies in detached settings may greatly enhance the early diagnosis and treatment of respiratory disorders, improving patient outcomes in the process. This study aims to close the gaps in the literature and improve patient outcomes when managing respiratory illnesses by improving healthcare personnel' diagnostic skills. Every study adds to the increasing amount of data that backs the use of audio-based, non-invasive diagnostic tools in medical settings.

III. PROPOSED WORK

*3.1 **Convolution Neural Network(CNN):***

Inspired by the research conducted by Hubel and Wiesel on the simple and complicated cells that make up the hierarchy of the primary visual cortex, Japanese electrical engineer Kunihiko Fukushima introduced a machine vision architecture in the late 1970s that he named the neocognitron. Based on the LeNet-5 concept, it was a significant development [11]. Like LeNet-5, AlexNet is structured hierarchically, with the first (left) layer representing fundamental visual elements like edges and the deeper levels representing more complex features and abstract concepts [11]. Convolutional neural networks, or CNNs or ConvNets, are a popular type of DL architecture used in machine vision applications nowadays. A CNN is an architecture for a deep learning model that has convolutional hidden layers [11]. CNN's main benefit over its predecessors is its capability to automatically recognise important traits without human monitoring, which makes it the most popular. Consequently, we have thoroughly investigated CNN by summarising its essential components. A CNN is consists of three important layers, or building blocks: convolution, pooling, and fully linked layers. Pooling and convolution, the first two layers, extract features; the fully connected layer, the third layer, converts the extracted features into the desired output, like classification [11]. The convolution layer of the CNN is a crucial part. CNN is made up of numerous mathematical operations, one of which is convolution [10]. A minimal set of framework known as the kernel—a feature extractor with optimisation capabilities—is applied at each picture point in digital photos. This makes it possible to store pixels as a two-dimensional (2D) grid or as an array of numbers. As a result, CNNs process images quite effectively [10, 11]. The complexity of the retrieved attributes may expand gradually and hierarchically as the output from one layer is fed into the next. Training is the process of adjusting parameters, or kernels, by minimising the difference between outputs and ground truth labels through the use of optimisation methods like as gradient descent and back propagation, among others. In essence, CNN is expressed as:

$y_n = \sum_{i=1}^{n} x_i * w_i + b_0$ (3.1)

In 3.1 y is output image, x is an input image ,b is bias, weights.

Patches from the input feature maps are pulled for max pooling, which outputs the maximum value in each patch and discards all other values,is the most often used type of pooling operation. A maxpooling with a stride of two and a size 2X2 filter is commonly employed in real-world applications. Consequently, the in-plane dimensions of feature maps are down sampled by a factor of two [11]. Unlike height and breadth, the depth dimension of feature maps is always the same [10,11].

Pooling layer: A downsampling procedure is used in this layer. After the output has been acquired and normalised, it is down sampled to produce an output with smaller image frame parameters. It is then applied on top of the flattened layer.

Flattened layer: Here, the activation unit of the layer is connected to the down-sampled picture frames.

Layer of activation: Here, the following activation functions are employed: Softmax, ReLu (Rectified Linear Unit).

The formula for ReLu is *f(a)=max(0,a).* (3.2)

The formula for softmax: $S(z_i) = e^{z_i} / \sum_{j=1}^{k} e^{z_j}$ (3.3)

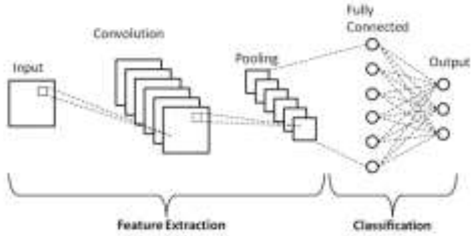

Figure 1: Architecture of CNN [12].

Figure 1 shows the design of Convolution Neural Networks that has convolution layer, pooling, fully connected layer between input and output.

By applying filters over one-dimensional input, 1D CNNs are useful for extracting local patterns from sequential data, such as time-series or audio signals [16]. 1D CNNs are capable of capturing fine-grained information, like fluctuations in pitch, tone, and amplitude, which may be suggestive of respiratory diseases, as well as temporal correlations for the purpose of lung illness prediction from human voice data. They are especially helpful for structured, sequential data, such as sound waves, and are computationallyefficient. Although CNNs are usually used to analyse picture data, they may also be modified to analyse voice data by turning voice signals into spectrograms, which are graphic representations of sound frequencies. CNNs are able to automatically extract spatial characteristics from the spectrograms in this situation [17], recognising patterns that correspond with symptoms of lung illness. They are well-known for their superior performance in tasks involving structured data, such as pictures or spectrograms, and they excel in feature extraction via learning hierarchical representations.

3.2 Long Short-Term Memory Networks (LSTM):

Recurrent Neural Network(RNN)s are a particular class of neural networks that are specifically made to handle sequential data. RNNs are essentially feedback-equipped neural networks. Data sequence modelling is achievable because the network's current state is based on both the input it is getting at any given time and the information it has stored in the past. LSTM networks are sequential neural networks that are capable of deep learning(DL) and information preservation [19]. The ability of this particular type of recurrent neural network [19] to employ RNNs to tackle the vanishing gradient problem makes it unique. Hochreiter and Schmidhuber created LSTM to address the issue that conventional RNNs and machine learning approaches brought about.

Figure 2: Architecture of Basic LSTM

The first section determines whether memory needs to be maintained for the timestamp-based data. From the input of this cell The cell tries to pick up new knowledge in the second segment. In the end, the cell updates the data by moving it from the current timestamp to the timestamp found in the third segment. This is a single LSTM cycle and one-time step [21]. These LSTM Three unit components are referred to as "gates". They control the data that enters and exits the memory cell, sometimes referred to as the LSM cell. The forget gate is the first gate, the input gate is the second, and so on. The final part is the output gate. a feedforward neural network (NN) standard layer, where each neurone has a hidden layer and a current state, can be compared to an LSTM unit. These three gates together with a memory cell—also referred to as an LSTM cell—make up the LSTM unit. RNNs of the long-term dependency LSTM type were created to solve the problem of learning long-term dependencies in data. LSTMs, in contrast to conventional RNNs, have a unique gating mechanism that enables them to maintain information over extended periods of time without experiencing vanishing or exploding gradients. Three main parts, commonly referred to as "gates," each with a distinct purpose, make up an LSTM unit. The network can handle long-term dependence because of these gates, which regulate the information flow into and out of the LSTM memory cell. Additionally, each LSTM cell keeps track of internal states that are updated each time-step. Here, we go over an LSTM cell's three primary parts and the states they affect: The method fresh data is added to a LSTM network is intended to assist the network recall crucial elements over time and filter out irrelevant ones. Without getting too technical, here's a condensed explanation of how LSTMs process new data:

Getting New Input: The LSTM cell gets a fresh set of data at every stage of a sequence. Like a word in a phrase or a new price in a stock market sequence, this is a part of the continuous sequence it is processing.
Input Gate: This new input's significance level is determined by the input gate, which is the first stage. Consider it as a filter that is able to determine which elements of the fresh data should be retained and which ones aren't pertinent to the current work. Both the incoming input and the LSTM's prior output have an impact on this decision-making process.
Candidate Memory Content: The LSTM generates a candidate memory, or a suggestion for changing the cell's memory state, in addition to evaluating the new input. Potentially new information that has been extracted from the input data and is intended to enrich the cell's memory with relevant content is contained in this candidate memory.
Forget Gate: There is a forget gate in place that operates in tandem with the process of taking in new knowledge. The forget gate determines which portions of the current memory should be deleted since they are no longer needed. This system makes sure that information that is out-of-date or unnecessary doesn't accumulate in the memory.
Updating Memory: Next, the old memory (after removing the bits indicated by the forget gate) and the fresh, filtered input (as permitted by the input gate) are combined to update the cell's actual memory. This refreshed memory now represents an ideal combination of new insights and prior knowledge for the current task. Lastly, the output gate chooses which regions of the updated memory to employ in order to produce the output for this timestep. The output is created by filtering the memory state, and it can be utilised as the last result in a series of outputs or for making additional judgements.
LSTMs are able to manage jobs requiring sequences where the timing and relevance of information alter dynamically because of their methodical approach to managing both new and old information. The LSTM's gates allow it to ignore distractions and concentrate on important details, which makes it incredibly efficient for jobs like financial forecasting and natural language processing, among others.

The following is the formula for forget gate

$$fg_t = \sigma(x_t * u_f + H_{t-1} * w_f) \qquad (3.4)$$

x_t: the current timestamp's input.
u_f: the input's allocated weight.
H_{t-1}: The prior timestamp's hidden state.
w_f:T his is the weight matrix corresponds hidden state.
Next, it is subjected to the sigmoid function.As a result, fgt will acquire a value ranging from 0 to 1 [21]. The previous timestamp's cell state is then multiplied by the ground.

Formula for input gate

$$ip_t = \sigma(x_t * u_i + H_{t-1} * w_i) \qquad (3.5)$$

x_t: Information input at timestamp t.
u_i: input weightmatrix
H_{t-1}: A concealed state at the earlier timestamp
w_i: The weight matrix related with the hidden state upon entry

Now that we've used the sigmoid function once more, timestamp t is between 0 and 1. Formula for fresh dataThe activation function in this case is tanh, and as a result, the value of new information can range from -1 to 1 [21]. Data is added to the cell state at the current timestamp if the value of Nt is positive; if it is negative, data is removed from the cell state.

$$N_t = \tanh(x_t * u_c + H_{t-1} * w_c) \qquad (3.6)$$

The N_t will not be appended to the cell state immediately. The most recent formula is this one.

$$c_t = fg_t * c_{t-1} + ip_t * N_t \qquad (3.7)$$

In this instance, Ct-1 represents the cell state as of the timestamp in question, and the other variables are the ones we calculated earlier.
Formula for output gate

$$op_t = \sigma(x_t * u_o + H_{t-1} * w_o) \qquad (3.8)$$

The sigmoid function ensures that the function's result also falls between 0 and 1. We will now use the tanh and ot of the updated cell state to determine the current concealed state. as demonstrated below.

$$H_t = o_t * \tanh(c_t) \qquad (3.9)$$

It finds out that the current output and the long term memory (Ct) control the concealed state.

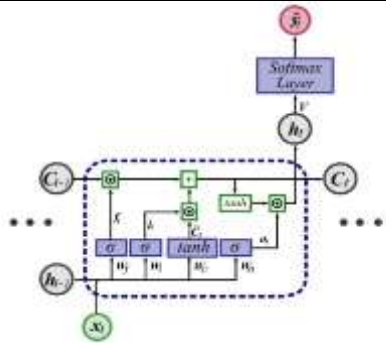

Figure 2: Architecture of LSTM [20].

Figure 2 shows the architecture and operation of LSTM.

LSTM are especially helpful for jobs that require long-term dependencies [18]. LSTMs can represent the temporal dynamics of the voice in voice-based lung disease prediction, including long-term trends in vocal patterns associated with abnormalities in breathing or voice alterations. This aids in the differentiation of patients' breathing patterns between healthy and sick.

3.3 Gated Recurrent Units (GRU):

In particular, this work focusses on time series prediction, for which the RNN approach is a popular deep learning method. Nevertheless, RNNs may run into issues like gradient ballooning and disappearing while learning long-term dependencies in the data[29]. To address these problems, research has proposed LSTM, which enhances gradient flow inside a network via a gating mechanism. An LSTM variation with two gates as opposed to the original three is called the GRU. Consequently, the GRU delivers higher training efficiency, reduces computational costs and model complexity, and shows improved capacity to recognise and comprehend long-term connections in time-series data. Because of its improved ability to handle long-term dependencies in time-series data, the GRU is the suggested choice [29]. Moreover, the GRU is suitable for managing large datasets because it requires less storage. Consequently, the basic GRU model was chosen as the study's principal model.

$$r_t = \sigma(w_r x_t + u_r h_{t-1} + b_r) \qquad (3.10)$$

$$z_t = \sigma(w_x x_t + u_z h_{t-1} + b_z) \qquad (3.11)$$

$$\widetilde{h_t} = \tanh(w_h x_t + u_h(r_t * h_{t-1}) + b_h) \qquad (3.12)$$

$$h_t = z_t * h_{t-1} + (1 - z_t) * \widetilde{h_t} \qquad (3.13)$$

in this case indicates the element-wise product formula; The weightmatrices w_r and w_z, respectively, are used for the r_t and z_t gates; uh; is the weight matrix representation of the output. The input data at time t is denoted by x_t; the candidate state and output state are denoted by h_t and $\widetilde{h_t}$, respectively; b_r, b_z, and b_h are constants; and the activation functions sigmoid and tanh [29], which are used to activate the control gates and candidate states, are represented by σ and tanh, respectively.

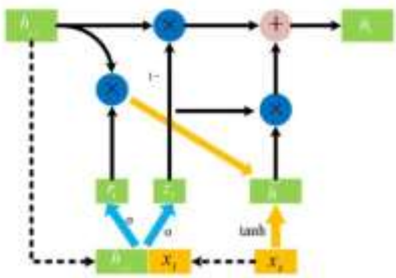

Figure 3 : Architecture of GRU [29].

Figure 3 shows the architecture of GRU.

The flow transmission procedure consists of the following stages when information is entered into the GRU unit:

Concatenation is performed between the output of the hidden layer (h_{t-1}) at time t − 1 and the input data (x_t) at time t [29]. Formula (9) [29] is used to obtain the reset gate r_t's output signal.Equation (3.11) is used to obtain the update gate z_t's output signal [29].Equation (3.12), which simply combines the input data and

hidden layer state at time t-1 after filtering by the resetgate [29], is used to produce the current state hidden unit candidate set (h_t)˜.[29] Equation (3.13) provides the hidden layer's output at time t.
GRUs are less computationally expensive than LSTMs when it comes to modelling temporal dependencies in the context of lung disease prediction from human speech. This is particularly helpful when working with sizable datasets or when there is a training time limit.

In this work, we proposed a comparative study using four different deep learning algorithms like 1DCNN,CNN,LSTM,gru

Algorithm:

1. Load the Respiratory voice data set.
2. Pre process the data set by removing noise, and normalize the data.
3. Apply Featue extraction.
4. Create model like 1DCNN or CNN or LSTM or GRU.
5. Compute the accuracy.

The above algorithm explains how each model works on Respiratory data set.

IV. RESULTS AND DISCUSSION

We observed four deep learning models for classifying the healthy and unhealthy voices. In unhealthy voices the diseases have the categories like Bronchiolitis, COPD, pneumonia, URTI, LRTI, Asthma. We computed accuracy for every model. The below Figure 5 is a confusion matrix when we applied 1D CNN. In respiratory dataset we have a total of 6898 audio files are there.

Table 1: Accuracies of Various Models

Model	Accuracy
1DCNN	92.56
CNN	98.56
GRU	96.71
LSTM	97.12

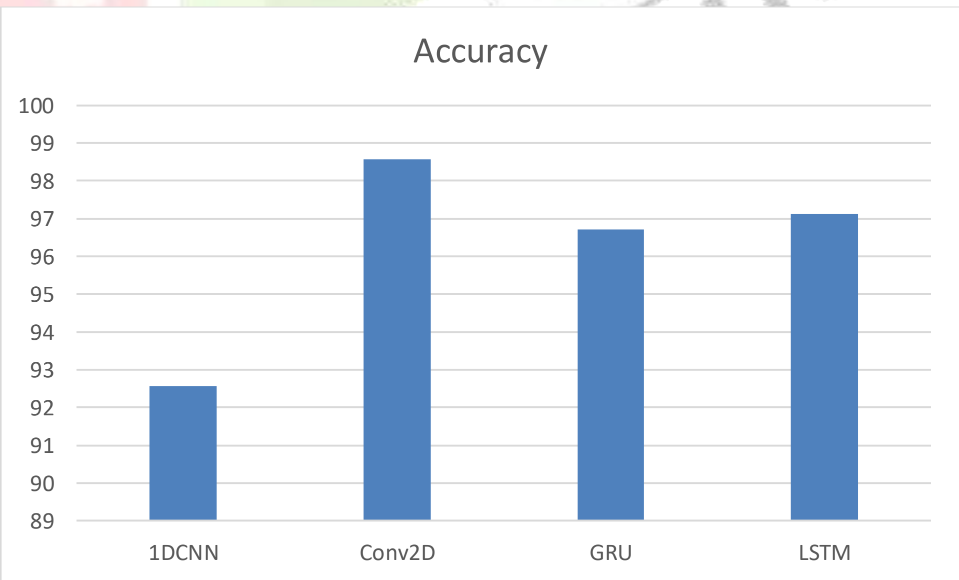

Figure 4: Comparison of various accuracies

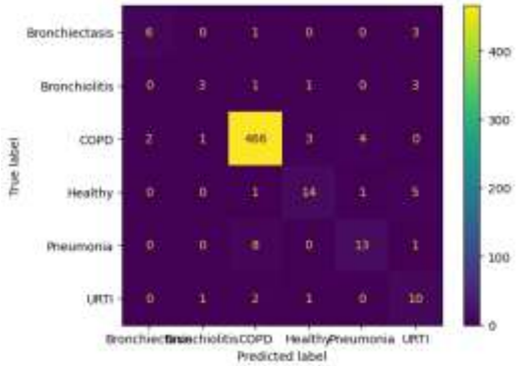

Figure 5: Confusion Matrix for 1D CNN.

Figure 4 clearly shows the comparisons of accuracies of various models.

From the results it is observed that 1DCNN model achieved low accuracy among all models,it achieved 92.56% accuracy . Whereas 2D CNN achieved high accuracy among all,it achieved higest accuracy as 98.56%. The model performed the best out of all the models that were examined because it was able to collect spatial and frequency information by transforming speech data into a 2D representation. GRU (Gated Recurrent Unit) Achieved an accuracy of about 96.71%. Voice signals represent sequential data, which is a good fit for GRUs. Their performance was probably enhanced by their capacity to manage long-term dependencies. They did not perform as well as CNN, though, which implies that convolutional features that are extracted in 2D space might be better appropriate for this task. LSTM (Long Short-Term Memory) - Also performed well, with an accuracy 97.12%. When managing long-range dependencies in sequences, LSTMs and GRUs are comparable. Their somewhat superior performance over GRU suggests they might have done a better job of modelling the audio data's sequential structure.

CNN had the highest accuracy, indicating that the most discriminative features for lung illness prediction came from converting the voice data into a format that could use 2D convolutions. Comparable results from LSTM and GRU highlight how crucial it is to extract temporal dependencies from the voice data. The least successful CNN was 1D, maybe because it was hard to get the required voice features in a one-dimensional environment. It seems that a 2D representation mixed with convolutional layers is quite good for lung disease prediction from speech data.

V. CONCLUSION:

This work shows that human speech analysis combined with deep learning models can effectively diagnose lung problems non-invasively. With an accuracy of 98.56%, the 2D Convolutional Neural Network (CNN) performed best out of all the models that were examined, demonstrating its capacity to extract features based on both spatial and frequency information by transforming speech data into a 2D format. This implies that because 2D CNNs can extract significant patterns from voice data, they are especially well-suited for this task. With accuracy rates of 96.71% and 97.12%, respectively, the GRU and LSTM models also demonstrated strong performance, highlighting the significance of identifying temporal correlations in sequential speech data. The 1D CNN, on the other hand, performed less accurately, suggesting that a one-dimensional representation might not be enough to extract important speech elements. Overall, the results indicate that lung disorders like COPD, pneumonia, and URTI may be accurately identified from voice recordings using deep learning techniques, particularly 2D CNNs. These models offer a viable path forward for the creation of non-invasive diagnostic instruments that can result in early intervention and identification, thus enhancing patient outcomes. These models can be further optimised to save computing costs without sacrificing accuracy. Effective models might make deployment possible in circumstances with limited resources, like low-power mobile devices or remote healthcare facilities.

REFERENCES

[1] Kumar, S., Bhagat, V., Sahu, P., Chaube, M. K., Behera, A. K., Guizani, M., Gravina, R., Di Dio, M., Fortino, G., Curry, E., & Alsamhi, S. H. (2024). A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases. Computer Methods and Programs in Biomedicine, 243, 107911. https://doi.org/10.1016/j.cmpb.2023.107911

[2] Bashford, T. (2024). AI Classification of Respiratory Illness Through Vocal Biomarkers and a Bespoke Articulatory Speech Protocol. International Journal of Simulation: Systems, Science & Technology. https://doi.org/10.5013/ijssst.a.25.01.13

[3] Saeed, T., Ijaz, A., Sadiq, I., Qureshi, H. N., Rizwan, A., & Imran, A. (n.d.). An AI-enabled Bias-Free Respiratory Disease Diagnosis Model using Cough Audio: A Case Study for COVID-19.

[4] Kaur, S., Larsen, E., Harper, J., Purandare, B., Ahmet Uluer, Mohammad Adrian Hasdianda, Ahmet Uluer Mohammad Adrian Hasdianda, & Nikita Umale, James Killeen, Edward Castillo, Sunit Jariwala Nikita Umale James Killeen, Edward Castillo Sunit Jariwala. (n.d.). Implementation and evaluation of a novel respiratoryresponsive vocal biomarker to aid in the detection of COVID-19.

[5] Bringel, K. A., Leone, D. C. M. G., de C. Firmino, J. V. L., Rodrigues, M. C., & de Melo, M. D. T. (2024). Voice Analysis and Neural Networks as a Clinical Decision Support System for Patients With Lung Diseases. Mayo Clinic Proceedings: Digital Health, 2(3), 367–374. https://doi.org/10.1016/j.mcpdig.2024.06.006

[6] Alghamdi, N. S., Zakariah, M., Karamti, H., & * Norah Saleh Alghamdi Norah Saleh Alghamdi. (n.d.). A deep CNN-based acoustic model for the identification of lung diseases utilizing extracted MFCC features from respiratory sounds.

[7] Wanasing, T., Bandara, S., Madusank, S., Meedeniya, D., De, I., Díez, L. T., 1 Department of Computer Science and Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka 47011 Valladolid, Spain, & 2 Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid,. (n.d.). Lung Sound Classification With Multi-Feature Integration Utilizing Lightweight CNN Model.

[8] Sfayyih, A. H., Sabry, A. H., Jameel, S. M., Sulaiman, N., Raafat, S. M., Humaidi, A. J., & Kubaiaisi, Y. M. A. (2023). Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview. Diagnostics, 13(10), 1748. https://doi.org/10.3390/diagnostics13101748

[9] Kim, H.-B., Song, J., Park, S., & Lee, Y. O. (n.d.). Classification of laryngeal diseases including laryngeal cancer, benign mucosal disease, and vocal cord paralysis by artificial intelligence using voice analysis.

[10] RikiyaYamashita,MizuhoNishio , Richard KinhGian Do , Kaori Togashi"Convolutional neural networks: anoverview and application in radiology"*Insights into Imaging* (2018) 9:611–629https://doi.org/10.1007/s13244-018-0639-9

[11] Wurtz RH. Recounting the impact of Hubel and Wiesel. J Physiol. 2009 Jun 15;587(Pt 12):2817-23 doi: 10.1113/jphysiol.2009.170209. PMID: 19525566; PMCID: PMC2718241.

[12] https://www.upgrad.com/blog/basic-cnn-architecture/

[13] Ghrabli, S., Elgendi, M., & Menon, C. (n.d.). Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments.

[14] Alam, Md. Z., Simonetti, A., Brillantino, R., Tayler, N., Grainge, C., Siribaddana, P., Nouraei, S. A. R., Batchelor, J., Rahman, M. S., Mancuzo, E. V., Holloway, J. W., Holloway, J. A., & Rezwan, F. I. (2022). Predicting Pulmonary Function From the Analysis of Voice: A Machine Learning Approach. Frontiers in Digital Health, 4. https://doi.org/10.3389/fdgth.2022.750226

[15] Cansiz, B., Kilinc, C. U., & Serbes, G. (2024). Tunable Q-factor wavelet transform based lung signal decomposition and statistical feature extraction for effective lung disease classification. Computers in Biology and Medicine, 178, 108698. https://doi.org/10.1016/j.compbiomed.2024.108698

[16] Kiranyaz, S., Ince, T., & Gabbouj, M. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151, 107398. https://doi.org/10.1016/j.ymssp.2020.107398

[17] O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458.*

[18] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[19] Meghana Reddy Vangala, "Long Short-Term Memory with Spin-Based Binary and Non-Binary Neurons", (2021), Electronic Theses and Dissertations, 2020-. 779, https://stars.library.ucf.edu/etd2020/779

[20] D. Bermúdez,P. Achanccaray, I. D. Sanches, L. Cue,P.Happ,R. Q. Feitosa," EVALUATION OF RECURRENT NEURAL NETWORKS FOR CROP RECOGNITION FROM MULTITEMPORAL REMOTE SENSING IMAGES", Anais do XXVII Congresso Brasileiro de Cartografia e XXVI Exposicarta 6 a 9 de novembro de 2017, SBC, Rio de Janeiro - RJ, p. 800-804.

[21] https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/

[22] Yan, Y., Aljbawi, W., Simons, S. O., & Urovi, V. (2024). Developing a Multi-variate Prediction Model For COVID-19 From Crowd-sourced Respiratory Voice Data (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2402.07619

[23] Machado, A., Almeida, S., Burtin, C., & Marques, A. (2022). Giving Voice to People – Experiences During Mild to Moderate Acute Exacerbations of COPD. Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation, 336–348. https://doi.org/10.15326/jcopdf.2022.0283

[24] Lella, K. K., Jagadeesh, M. S., & Alphonse, P. J. A. (2024). Artificial intelligence-based framework to identify the abnormalities in the COVID-19 disease and other common respiratory diseases from digital stethoscope data using deep CNN. Health Information Science and Systems, 12(1). https://doi.org/10.1007/s13755-024-00283-w

[25] Hassan, U., Singhal, A., & Chaudhary, P. (2024). Lung disease detection using EasyNet. Biomedical Signal Processing and Control, 91, 105944. https://doi.org/10.1016/j.bspc.2024.105944

[26] Kapetanidis, P., Kalioras, F., Tsakonas, C., Tzamalis, P., Kontogiannis, G., Karamanidou, T., Stavropoulos, T. G., & Nikoletseas, S. (2024). Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review. Sensors, 24(4), 1173. https://doi.org/10.3390/s24041173

[27] Zhang, P., Swaminathan, A., & Uddin, A. A. (2023). Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks. Frontiers in Medicine, 10. https://doi.org/10.3389/fmed.2023.1269784

[28] Idrisoglu, A. (2024). Voice for Decision Support in Healthcare Applied to Chronic Obstructive Pulmonary Disease Classification. https://books.google.com/books/about/Voice_for_Decision_Support_in_Healthcare.html?hl=&id=_4ed0AEACAAJ

[29] Chi Chen , Lei Xue and Wanqi Xing, "Research on Improved GRU-Based Stock Price Prediction Method", Appl. Sci. 2023, 13, 8813. https://doi.org/10.3390/app13158813.

[30] Sangeetha Balasubramanian, & Periyasamy Rajadurai. (2023). Machine Learning-Based Classification of Pulmonary Diseases through Real-Time Lung Sounds. International Journal of Engineering and Technology Innovation, 14(1), 85–102. https://doi.org/10.46604/ijeti.2023.12294

[31] Piccari, L., Kovacs, G., Jones, S., Skaara, H., Herms, C. R., Jeanneret, G. S. B., Vinardell, M. C., Guix, N. G., Delgado, M. F., García, H. G., & Schwicker, D. (2024). The European Voice of the Patient living with pulmonary hypertension associated with interstitial lung disease: Diagnosis, symptoms, impacts, and treatments. Pulmonary Circulation, 14(2). https://doi.org/10.1002/pul2.12405

[32] Karaarslan, O., Belcastro, K. D., & Ergen, O. (2024). Respiratory sound-base disease classification and characterization with deep/machine learning techniques. Biomedical Signal Processing and Control, 87, 105570. https://doi.org/10.1016/j.bspc.2023.105570

[33] Idrisoglu, A., Dallora, A. L., Cheddad, A., Anderberg, P., Jakobsson, A., & Sanmartin Berglund, J. (2024). COPDVD: Automated classification of chronic obstructive pulmonary disease on a new collected and evaluated voice dataset. Artificial Intelligence in Medicine, 156, 102953. https://doi.org/10.1016/j.artmed.2024.102953

[34] Sabry, A. H., I. Dallal Bashi, O., Nik Ali, N. H., & Mahmood Al Kubaisi, Y. (2024). Lung disease recognition methods using audio-based analysis with machine learning. Heliyon, 10(4), e26218. https://doi.org/10.1016/j.heliyon.2024.e26218

[35] Chudasama, V., Bhikadiya, K., Mankad, S. H., Patel, A., & Mistry, M. P. (n.d.). Voice Based Pathology Detection from Respiratory Sounds using Optimized Classifiers.

Dr.Y.SowjanyaKumari working as Associate Professor in QISCET (Affiliated to JNT University, Kakinada), Ongole, India,and she has 20 years of teaching experience. she received her Ph.D.degree from JNT University, Kakinada in 2019. She has receivedM.Tech. in Computer Science & Engineering from JNTU,Kakinada A.P and B.Tech in Computer Science & Engineeringfrom N.B.K.R.I.S.T, Vidyanagar, Nellore(dt) A.P. Her areas of interestingincludes Digital image processing and Artificial Intelligence, Machine Learning.