Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 4, August, 2025

Unleashing Effective Identification of ALS Based on Vowel Phonation: A Deep Learning Approach

Hussein Al-Dossary [1*], Mohamudha Parveen Rahamathulla [2], Mohemmed Sha [3]

[1] *University Hospital, Prince Sattam Bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia.*

[2] *School of Life and Health Sciences, University of Roehampton, London, United Kingdom.*

[3] *Department of Software Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al Kharj 11942, Saudi Arabia.*

Abstract

ALS (Amyotrophic Lateral Sclerosis) is one of the fatal diseases across the world. Therefore, early detection can save patients suffering from ALS from life-threatening consequences. Typically, ALS can be identified based on different factors, and one such factor is voice analysis. Detection of ALS using sound signals is convenient and simpler than other methods, as it is a non-invasive approach, which makes the process faster and more efficient for detection. However, detection of ALS using traditional approaches is challenging, as it is a time-consuming process and heavy reliance on medical experts is needed. Therefore, AI-based models can be used for effective classification of ALS and non-ALS patients, as AI-based models possess the immense ability to examine vast amounts of data, including audio files, effectively. Owing to these factors, the proposed model focuses on employing an AI-based model for ALS classification based on vowel phonation /a/ and /i/. The process is carried out using the Minsk2020 dataset, where important features needed for the proposed model are extracted using MFCC (Mel-frequency cepstral coefficients) by removing the shakiness and jitteriness of the voice. The MFCC feature extraction technique extracts features based on the mel scale, as this reflects human auditory perception, thereby extracting features that are useful for classification. These extracted features are fed to CNN-LSTM (Convolutional Neural Network – Long Short Term Memory) with rapid dilatenet for classifying ALS and non-ALS patients accurately by identifying even the subtle changes in audio signals using maximizing the expansion/dilation rate and aid the context information for interpreting and analyzing the sound of vowels accurately and correctly without any loss of information. Finally, the efficacy of the proposed model is assessed using evaluation metrics. The proposed research work can assist medical professionals in detecting patients with ALS based on vowel phonation.

Keywords:

Amyotrophic Lateral Sclerosis;
Mel–Frequency Cepstral Coefficients;
CNN; LSTM;
Dilation Rate Classification;
Minsk2020;
Vowel Phonation.

Article History:

Received:	26	April	2025
Revised:	01	July	2025
Accepted:	08	July	2025
Published:	01	August	2025

1- Introduction

ALS, or amyotrophic lateral sclerosis [1-3], is a progressive neurodegenerative disease that affects the nerve cells in the brain and spinal cord [4]. "Amyotrophic" derives from the Greek language, in which "A" signifies "no" [5], "myo" refers to muscle, and "tropic" refers to nourishment. Thus, the term "amyotrophic" represents "no muscle nourishment" [5]. Likewise, "lateral" refers to the area in a person's spinal cord [6] where a portion of the nerve cells that signal and control the muscles are located. As this area degenerates, it leads to "sclerosis," which is scarring or hardening in the

* **CONTACT**: hm.aldossary@psau.edu.sa

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-04-012

region [7]. Figure 1 shows the normal nerve cells and muscle cells vs. affected nerve and muscle cells. Thus, ALS is considered one of the fatal diseases [8, 9] that needs to be identified and treated as early as possible in order to avoid further complications. Statistics indicate that 1 in 500 adults dies [10] from ALS in the USA, and over 16 million people who are alive can suffer from this disease due to an imbalance in diet or lifestyle changes. Therefore, it is important to detect ALS as quickly as possible in order to avoid life-threatening situations.

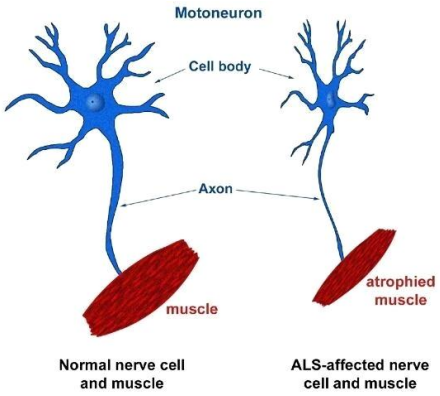

Figure 1. Normal nerve vs ALS affected nerve [11]

In order to detect ALS, different approaches are employed. Some of the conventional approaches implemented for detecting ALS using sound data include manually assessing the speech patterns by listening to the recordings [12, 13] of patients to detect the changes in tone, rhythm, and intonation. Similarly, manual inspection using cough and swallowing sounds also detects the changes that happen in patients with ALS [14]. Besides, clinicians also utilize different equipment for listening to the sounds produced by the muscles of the patients. Furthermore, the employment of laryngoscopy by clinicians to visually inspect the vocal cords of the patients and speech production mechanism for indicating the early signs of ALS [15]. Despite the performance of the conventional tactics, there are certain limitations that need to be overcome to obtain better outcomes, such as the time-consuming process of traditional approaches, proneness to errors, heavy dependency on medical professionals and experts, and laborious and tedious processes. Therefore, AI-based models can be used to attain promising results, as AI is an advanced technology that can work and analyze massive amounts of data efficiently [16]. AI-based models can assist with identifying and monitoring ALS by assessing the sound signals, including vowels [17, 18]. Therefore, by analyzing the characteristics of vowel sounds, AI models can identify the changes in speech patterns effectively. Owing to the advantages of AI, different existing works have used AI techniques [19].

Different ML and DL algorithms [20], like NB (naïve Bayes), KNN (k Nearest Neighbor), DT (Decision Tree), SVM (Support Vector Machine), ANN (Artificial Neural Network), and LDA (Linear Discriminant Analysis), were used in the study for classification of ALS and non-ALS patients using speech phrases. Relatively, the performance of the model has been further enhanced by using PCA compression. The experimental outcome obtained from the model showcased that the SVM model has delivered a better accuracy rate for the detection of ALS and non-ALS patients. Similarly, SVM and DT [21] are used for the identification of ALS patients based on the articulatory phenotypes, which include coordination of the speech, speed of the speech, precision, and consistency of the speech.

Though the existing models have focused on different sound-based phonatory tasks, identification of ALS based on vowel phonation is limited. Besides, there are varied limitations faced by the state-of-the-art approaches, such as low accuracy, overfitting of the model, lack of studies using the Minsk2020 dataset, inability to work with a wide range of data, and scalability issues. Further, many existing models may not fully capture the temporal aspects of speech, which are crucial for understanding the progression of ALS due to the implementation of ineffective algorithms. Moreover, the existing models are known to be challenging, as the acoustic features can consist of jitter, shimmer, and harmonics-to-noise ratio, which can vary significantly among individuals. This variability can lead to difficulties in establishing a reliable baseline for distinguishing between healthy individuals and those with ALS, particularly in early stages when symptoms may be subtle. Thus, the proposed model focuses on designing a model that detects the patients with ALS depending on the sustained vowel phonation, as identification of ALS using vowel sounds requires less articulation and provides a more reliable outcome for classification, as vowel sounds are more acoustically stable and can be easily detected using software tools. Thus, the vowel phonation approach is focused on the proposed model for the classification of ALS and non-ALS patients as/a/ and /i/. Unlike other existing works, which extracted vowels from running speech

tests, sustained phonation of vowels is focused on effective classification outcomes. In order to accomplish this, the proposed work focuses on implementing a hybrid CNN-LSTM with a rapid dilatant model for efficiently classifying patients suffering from ALS based on the phonations of vowels.

1-1-Motivation and Research Contribution

ALS is considered one of the fatal motor neuron diseases [22, 23], which is characterized by progressive degeneration of nerve cells in the spinal cord and brain. Therefore, it is important to detect ALS symptoms as quickly as possible to avoid life-threatening situations. There are different ways to identify the symptoms of ALS in a patient. However, detection of ALS using sound signals like vowel phonation helps with better outcomes, as it is a non-invasive procedure that uses samples collected from EMG and lumbar puncture methods and also acts potentially faster than other approaches. However, manual approaches for ALS detection can be challenging, as pitch, duration, coordination of speech, and frequency of patients' vocal cords suffering from ALS may vary from non-ALS patients, and it can be extremely time-consuming and tedious to detect ALS effectively using conventional approaches. Thus, AI-based models can be used for detecting ALS effectively, as they can handle huge amounts of data and reduce the risk of human error in diagnosis. However, existing researchers could not work with the substantial size of datasets, which can result in imprecise, ineffectual outcomes and inaccurate accuracy of the models for ALS classification based on vowel phonation due to the incorporation of ineffective algorithms.

Motivated by these factors, the proposed work utilizes MFCC for feature extraction, which captures the spectral characteristics of sound, as it is considered to be effective for vowel phonation. Moreover, when a patient produces a vowel sound, the vocal cords vibrate at specific frequencies, creating formants in the sound spectrum, and these frequencies aid in identifying the vowels. In healthy individuals, formants transition smoothly within a vowel sound, whereas, due to muscle weakness, the vocal cords in ALS patients can become slower, shakier, and unstable; thus, the proposed work utilizes the MFCC feature extraction technique for extracting features based on the mel scale, as this reflects human auditory perception, thereby extracting features that are useful for the classification of ALS. Then, the classification of ALS is employed using the proposed CNN-LSTM with a rapid dilatant model for effectively classifying the patients as suffering from ALS and non-ALS patients. Features extracted from MFCC are sent to the proposed CNN-LSTM model, in which CNN aids in learning the features automatically for detecting the significant patterns in the signals, whereas the LSTM utilizes the spectral features from CNN to comprehend the flow and context within the vowel phonation efficiently.

Further, long-term dependencies of the LSTM model refer to the changes between sounds that are all separated by relatively long durations within vowel sounds. However, audio signals can be intricate and complex to dissect the changes effectively; hence, the proposed CNN-LSTM uses a rapid dilatenet function, which helps in identifying even the subtle and delicate changes in audio signals and effectively classifying ALS and non-ALS patients by maximizing the expansion/dilation rate and aids the context information for interpreting and analyzing the sound of vowels accurately and correctly without any loss of information. Therefore, the contribution of the research involves:

- To employ MFCC for effectively extracting the features for the model using vowel phonations of the patients using the Minsk2020 dataset
- To implement the CNN-LSTM model for the classification of ALS based on vowel phonation.
- To perform binary classification of ALS and non-ALS patients using the proposed CNN-LSTM model with rapid DilateNet function by addressing the contextual information is essential for accurately interpreting and analyzing vowel sounds without any loss of detail.

To evaluate the performance of the proposed model, a range of metrics are employed, including accuracy, sensitivity, and specificity.

2- Literature Review

Different existing works done for the classification of ALS and non-ALS patients are reviewed in the subsequent section.

ALS is defined as one of the neurodegenerative diseases that specifically affect speech impairments, the spinal cord, and swallowing difficulties [22-24]. Moreover, the rise of ALS has increased gradually in old-age people, and it cannot be diagnosed easily. Therefore, different classification algorithms are used for detecting ALS effectively. SVM [25] has been used in the study for the classification of ALS patients and HC based on sustained vowel phonation, as early diagnosis of ALS can inevitably enhance the quality of life to a certain extent. Similarly, the biomechanical process of voice production has been used for distinguishing ALS patients from non-ALS patients. In order to accomplish the process, RF (Random Forest) [26] has been utilized for classification, and the result is projected to demonstrate the

potential of using vocal fold dynamics for ALS identification. The drawback of the model is the generalizability of the findings. Likewise, a study focused [27] on developing a robust and efficient system for detecting voice pathologies using the LSTM method. This research integrated a novel combination of feature sets, including MFCCs, Zero Crossing Rate and Mel Spectrograms. The implementation of the LSTM approach significantly enhanced the accuracy of voice pathology identification when applied to samples from the SVD (Saarbruecken Voice Database), and the experimental results were evaluated utilizing accuracy, precision, specificity, sensitivity, and F1 measures. Similarly, the OSELM (Online Sequential Extreme Learning Machine) model [28] has been used to classify voice signals as either healthy or pathological. Voice features were extracted using MFCC, with samples of the vowel /a/ collected equally from the SVD. The OSELM approach was assessed using three widely recognized metrics: accuracy, sensitivity, and specificity. The results demonstrated that the maximum values achieved were 85% for accuracy, 87% for sensitivity, and 87% for specificity.

Approximately 80-96% of people suffering from ALS automatically lose their ability to speak during the disease progression. Thus, Milella et al. [29] have aimed to assess the detection of ALS clinical phenotypes using acoustic voice parameters. Acoustic voice analysis used in the study is considered to be useful for differentiating flaccid dysarthria and spastic dysarthria and also for assessing the degree of bulbar involvement in ALS. However, the small sample size used in the existing work has restricted the generalizability of the findings and also the robustness of the model. Likewise, the classification of ALS patients has been focused on the existing work [30], in which, based on the vowel patterns, ALS patients have been distinguished without any bulbar involvement. The vowel pattern used in the study has been produced from quasi-periodic components for detecting the deficiency in females and males via utterance of the vowels, and the classification process has been carried out using the RF technique. Correspondingly, DT (Decision Tree) has been used for separating the patients with ALS and non-ALS using acoustic analysis in assorted voice signals of different degrees of impairment.

Bulbar dysfunction is one of the terms utilized in ALS. It is a motor neuron disability that could lead to dysfunction of swallowing and also speech issues [31, 32]. Hence, voice deterioration is considered one of the early symptoms of bulbar dysfunction [32]. Therefore, the research work focused on identifying and diagnosing the problem automatically at the early stages of the disease. The process can be accomplished by using RF and SVM [34] models. However, from the experimental outcome, it was demonstrated that RF delivered a better outcome than SVM. Precise detection of features [34, 35], which have been extracted from the acoustic analysis of the vowels produced by patients suffering from ALS, aids in performing the classification of ALS and non-ALS. PCA [35] was utilized for obtaining features, and SVM with a 50% classification threshold was implemented for employing the classification process. The inherent variability is considered one of the limitations of the study in addition to the small sample size data drawback. As speakers differ in production, even identical speakers in an identical context do not possess the ability to produce 2 completely indistinguishable utterances; therefore, manual methods used in the study for processing the speech are not precise and require proper manual correction for obtaining effective outcomes.

Detection of ALS using speech and voice symptoms can be challenging for both automatic systems and human specialists [36]. Hence, Vashkevich & Rushkevich [37] have focused on using MFCC for feature extraction. Further, a set of acoustic features was used for classifying the patients based on vowel phonations. In order to perform classification, LDA (Linear Discriminant Analysis) has been carried out, and the LASSO technique has been used for feature selection. Employment of these techniques aids in better classification of HC from ALS patients. Likewise, the LDA classifier [38] has been used for determining the group of ALS patients and HC. The model used in the study was developed using a mobile application and was processed using a Minsk dataset with 64 voices. Usage of the LDA classifier has delivered better outcomes at a low cost. Supervised learning approaches [17] such as SVM, LR, NB, DR, and RF have opted for the classification of ALS and non-ALS. Classification of the model is based on vowels, sentences, and coughs of the patients using the HomeSenseALS dataset and Minsk dataset, and different results have been obtained for different approaches.

Likewise, the Bayesian LR classifier [39] has been implemented to distinguish between ALS and HC. The dataset used in the study consisted of 119 ALS and 22 controls, which aided in the classification of ALS. As swallowing and speech difficulties are considered as one of the early signs of ALS, it is important to employ techniques that aid in the detection of ALS. Thus, the correlation-based feature selection technique [40], PCA for reducing the dimensionality of the dataset, and SVM for classification of the model have been implemented in the study for ALS classification, and the analytical outcome displayed by the model, like accuracy and sensitivity, was 84.2% and 77.8%. Running a speech test [41] was preferred in the study for the detection of ALS, and selected vowels were extracted from the input audio signals. The process is accomplished by using an LDA classifier, and the detection accuracy attained by the model for the classification of ALS and non-ALS patients is 84.8%.

Table 1. Summary of the existing works

S.No	Reference	Method	Methodology	Outcome
1	Simmatis et al.(2024) [39]	LDA classifier	The model employed in the study was developed by Bayesian logistic regression classifier and was processed using a Minsk dataset with 64 voices by LDA classifier through automated assessment app.	Usage of the LDA classifier has delivered better outcomes at a low cost
2	Lv et al. (2024) [42]	A novel audio-visual fusion model	The model has used audio-visual samples from 130 PD patients and 90 healthy participants. The classification process has attained by based on Transformer cross attention module	The model has been attained average performance of 92.68%.
3	Alqahtani et al. (2024) [43]	RSFFNN-CNN (Resemble Single Feed Forward Neural Network-Convolutional Neural Network)	The model has been classified the ALS clinical associations to analyze the ALS. The model has customized each hidden layer by "k" parameter.	The model has been attained average performance
4	Al-Dhief et al. (2024) [28]	OSELM (Online Sequential Extreme Learning Machine)	Voice features were extracted using MFCC, with samples of the vowel /a/ collected equally from the SVD. The OSELM approach was assessed using three widely recognized metrics: accuracy, sensitivity, and specificity.	The results demonstrated that the maximum values achieved were 85% for accuracy, 87% for sensitivity, and 87% for specificity.
5	Mahum et al. (2024) [44]	Tran-DSR (Transformer Dysarthric Speech Recognition)	The model has encompassed strength of Transformer encoder and ensemble deep networks. The existing research has included two ensemble scenarios and self-attention approach has been used to construct the Transformer encoder	The model has been attained better performance
6	Rong et al. (2024) [45]	The model combined acoustic instrumental and facial sEMG techniques to compare the multimodal performance	The model has been used SVM with RBF kernel based on RF to classify the ALS	The model has been attained average performance of 88%.
7	Mehra et al. (2024) [46]	Deep BiLSTM-GRU model	The model has been used SepFormer and Swim transformer to extract the audio signals from the dataset. The classification process has achieved by deep BiLSTM-GRU.	The model has been attained average performance

2-1-Research Gaps

The gaps identified by the existing works have been explicated as follows:

- The small sample size used in the existing work has restricted the generalizability of the findings [28] and also the sturdiness of the model [28]. Thus, larger and more diverse samples can improve the robustness and strengthen the results of the paper, which is focused on the proposed work.
- Manual detection [35] used in the suggested model for processing the speech is not precise and can potentially lead to inaccurate classification of ALS and non-ALS patients; hence, this needs to be overcome by employing effective AI methods.
- Though the accuracy [37, 41] obtained by the models for classifying ALS and non-ALS is considerable, better accuracy can be attained by using capable algorithms.

The LDA classifier [39] does not consider issues such as class imbalance, as well as the influence of feature selection methods, which are known to affect model performance for voice-based disease classification.

3- Proposed Methodology

ALS is defined as one of the incurable neurological diseases with a rapidly progressive course that can threaten the life of a human being. Therefore, it is important to detect this disease early in order to prevent serious consequences. There are different ways of detecting ALS in the human body, which include muscle weakness, twitching, muscle cramps, and difficulty in walking or speaking; however, detection of ALS through sound can be effective since it is non-invasive, easy to administer, and does not require any specialized equipment for detection. Besides, the sound of vowels for ALS detection is emphasized by the experts, as ALS first affects the muscles involved in speech production. Hence, any changes in speech patterns, like the forming of words or slurred speech, are considered to be early signs of ALS. However, detection of ALS using vowels can be challenging since voice symptoms differ from person to person, which makes it tedious to identify ALS effectively.

Moreover, these two vowels are phonetically maximally distinct, because the vowel /a/ is articulated with the tongue in a low and back position in the mouth, and the lips are open when articulating. Acoustically, the vowel is defined by both a low first formant (F1) and a low second formant (F2). The vowel /i/ is, however, articulated with the tongue in a high and front position of the mouth, and the lips are unrounded. Acoustically, /i/ has a high F1 and a high F2, making it unambiguously distinguishable from /a/ in both the position of the tongue and formant profile. These differences in articulation and acoustics are basic to the differentiation of vowel sounds in speech analysis.

The vowels /a/ and /i/ stand at the opposite ends of the vowel space, with /a/ being low-back and /i/ high-front, thus making these two vowels very sensitive to changes in motor control and vocal tract stability. Under ALS, due to muscle weakness or instability, a reduction in the vowel space and lessened articulatory precision tend to occur, with very evident

changes in formant transition, jitter, and shimmer seen especially in these vowel types. Studies have found /a/ and /i/ to be the most sensitive to bulbar involvement and dysarthria in ALS, demonstrating early and measurable changes in formant structure, duration, and spectral tilt as the disease progresses.

However, even though more approaches are used for detecting ALS in individuals effectively, there are certain drawbacks that need to be taken into consideration, such as a time-consuming process, proneness to error, heavy dependency on experts and medical professionals, subjectivity, and lack of standardization. Hence, in order to overcome these drawbacks, AI-based models can be implemented, as they possess various advantages like the ability to analyze huge amounts of data effectively and examine the patterns and relationships between the data efficiently. However, the classification of ALS using vowel phonation in existing works lacked the delivery of better accuracy and capable outcomes that can lead to the proficient classification of ALS and non-OIP patients. Hence, the proposed model works with algorithms that are capable of producing desirable outcomes for the classification of ALS and non-ALS based on vowel phonation /a/ and /i/. Thus, Figure 2 shows the process involved in the proposed mechanism for ALS classification.

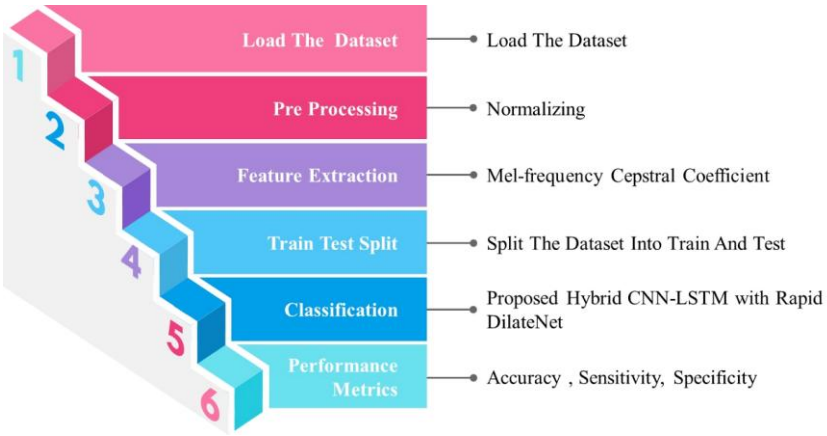

Figure 2. Overall Flow of Proposed Model

The proposed process is depicted in Figure 2. Where the process is carried out by loading the dataset. Once the dataset is loaded, data is pre-processed using the normalizing technique. The normalization method takes place by adjusting the amplitude of an audio signal to a desired level. After pre-processing, feature extraction is done by using the Mel frequency method. As the pronunciation of vowels and the lagging time of patients suffering from ALS may vary from person to person, it is important to detect even the slightest changes in the motion of the articulators appropriately.

Voice analysis takes advantage of acoustic feature extraction namely, fundamental frequency (F0), jitter, shimmer, harmonics-to-noise ratio (HNR), and formant frequencies (F1, F2, F3)—to identify speech anomalies and articulatory impairments typical of neurodegenerative conditions like ALS. The use of methods like Mel-Frequency Cepstral Coefficients (MFCCs) is especially beneficial in detecting symptoms like dysarthria, slurred speech, and vocal tremors, making it possible to systematically detect disease progression. The cost-effective, scalable, and non-invasive characteristics of voice analysis enable early detection, remote monitoring, and ongoing assessment, making it an important technology for the creation of AI systems that can diagnose neurological disorders prior to overt symptoms and ultimately enhance patient outcomes.

Therefore, MFCC employment is used in the proposed model, as it inherently possesses the capability to model an irregular movement of the vocal folds and aids in extracting suitable features needed for the model to perform the classification of ALS using vowel phonations. After feature extraction, the dataset is split as a train-test split (80%-20%). After the train-test split, the classification process takes place by employing the CNN-LSTM model with the rapid dilatant technique. Implementation of CNN helps with automatically learning the features to detect significant patterns in MFCCs, and these patterns could denote formants that are extremely crucial for vowel phonation. Moreover, the employment of the LSTM model for handling sequential data like speech.

In addition, the internal memory of the LSTM model permits the model to remember information from previous sounds and utilize the context to comprehend the current sound. This function is important for capturing the nuances of vowel phonation. However, it is extremely crucial to identify the subtle changes in the sound signals of vowel phonation, as it helps with the identification of ALS patients easily. In order to achieve this, the proposed work utilizes a rapid dilatant model. The proposed rapid dilatenet controls the spacing between the filter elements within the kernel. By inserting zeros between elements, the filter can examine a wide range of the input data and aids in identifying even the subtle changes in vowel phonation. This could be due to the dilation rate used in the proposed model, which comprehends the temporal relationships and variations in the vocal characterization of the data that lead to the effective classification of ALS. Eventually, evaluation metrics are used for gauging the efficacy of the proposed model. The overall process that takes place in the proposed mechanism is illustrated in Figure 3.

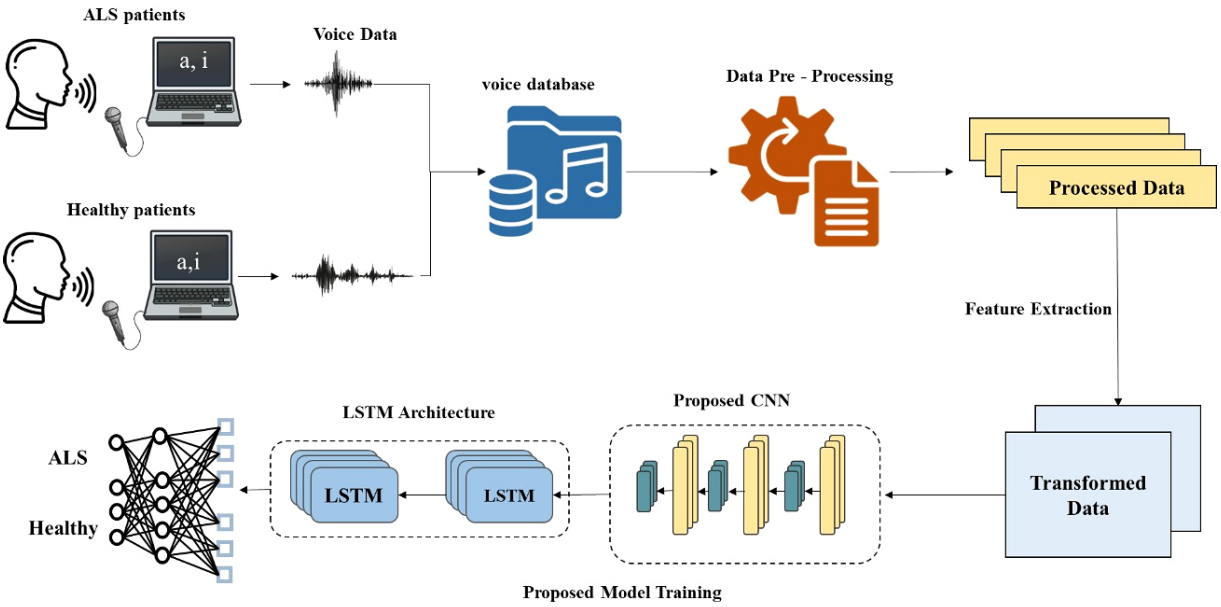

Figure 3. Architecture of Proposed Mechanism

3-1- Dataset Description

The Republican Research and the Clinical Center of Neurology and Neurosurgery (Belarus, Minsk) has the driven voice database. With corresponding sustained vowel phonations of 128 (64 of vowel/a/ and 64 of vowel/i/) among 64 speakers, 31 have been spotted with ALS. For each single one, the speaker has been demanded to render the upheld phonation of vowels /a/and /i/ at a convenient standard of promotion for a while. The voice database is comparatively well-balanced and comprises pathological voices of 48% and healthy voices of 52%. Table 2 shows the number of speakers and percentage of ALS and non-ALS patients.

Table 2. ALS and non-ALS speakers

Speaker Group	Number of Speakers	Percentage
ALS	31	48
Healthy	33	52

In addition, Table 2 represents the gender of ALS, HC male, and ALS, HC female, with their age range and mean age (SD) in Table 3.

Table 3. Age range and Mean

Gender	Age Range	Mean Age (SD)
ALS Male	40-69	61.1 (7.7)
ALS Female	39-70	57.3 (7.8)
HC Male	34-80	50.2 (13.8)
HC Female	37-68	56.1 (9.7)

Table 2 shows that the study included 17 male patients aged 40–69 years (mean 61.1 ± 7.7) and 14 female patients aged 39–70 years (mean 57.3 ± 7.8). Among the Healthy Controls (HC), there were 13 men aged 34–80 years (mean 50.2 ± 13.8) and women aged 37–68 years (mean 56.1 ± 9.7). Speech samples were recorded using various smartphones with standard headsets and stored as uncompressed 16-bit PCM files. The mean phonation duration for the HC group was 3.7 ± 1.5 seconds, while for the ALS group it was 4.1 ± 2.0 seconds, as presented in Table 4.

Table 4. Mean and SD

Recording Duration (s)	Mean	Standard Deviation
HC	3.7	1.5
ALS	4.1	2

3-2- Feature Extraction

Feature extraction is carried out in the proposed model, as it aids in extracting significant information from the audio signals, which can be utilized for identifying the patterns and characteristics indicative of the disease. As the feature extraction process primarily focuses on extracting relevant features needed for the model, effective feature extraction techniques should be taken under consideration for obtaining outcomes like the pitch of the sound, intensity of the sound, and formants from the speech signals. Moreover, the implementation of the feature extraction process also minimizes the dimensionality of the data, thereby making the process of classification effective and efficient. Thus, in order to carry out the process of feature extraction, the proposed model emphasizes using the Mel frequency technique, primarily due to the model's ability to meticulously resemble the human auditory system response to sound. Besides, the Mel frequency scale is considered to be a logarithmic scale, which represents how humans perceive sound frequencies as opposed to a linear scale. Thus, features extracted using Mel frequency are likely to capture significant and detailed characteristics of the sound signals, which are relevant for human perception, and it is achieved due to the ability of MFCC to capture spectral characteristics of sound, which are important for vowel phonation.

The process of the Mel Frequency model typically involves applying techniques such as:

- Windowing – Windowing aims to minimize the discontinuous effect on the signal after the framing process. Thus, ideal windowing should be used so that the features of each sound are not wasted.
- Fourier Transform—This is used to obtain the frequency spectrum of a signal by converting a time-domain signal into a frequency-domain signal.
- Mel Filter bank – Employment of Mel FilterBank helps with capturing the perceived pitch difference more precisely and efficiently.
- Logarithmic scaling – Logarithmic scaling is applied to the output of the Mel FilterBank with the aim of representation similar to that of human perception of sound.

Thus, Algorithm I shows the process involved in Mel frequency for the feature extraction process.

Algorithm I. Mel Frequency

```
Input: Signal (Vowel phonations signal)
Output: MFCC (MFCC of Vowel phonations signal)

Function MFCC (parameters):
   Initialize Parameters:
Number of frames, Frame Size, Number of filters in the mel filterbank,
Number of MFCCs to extract, Other relevant parameters.
   Frame the Signal:
Split the Vowel phonations signal into overlapping frames of a specified size.
   Apply Windowing:
Apply a Hamming window to each frame to reduce spectral leakage.
   Compute the Spectrogram:
For each frame:
Apply the Fast Fourier Transform (FFT) to obtain the magnitude spectrum.
   Apply Mel Filterbank:
Construct a mel filterbank with triangular filters spaced according to the mel scale.
Apply the filterbank to the magnitude spectrum of each frame, resulting in a mel-scale energy
representation.
   Take Logarithm:
Take the logarithm of the mel-scale energies for each frame.
   Compute MFCCs:
Apply the Discrete Cosine Transform (DCT)to the log mel scale energies for each frame.
Keep the first N coefficients as the MFCC features.
End Function
```

Initially, the process starts by feeding input to the proposed model for feature extraction. Then, parameters are initialized accordingly, such as the number of frames, frame size, number of filters in the mel FilterBank, number of MFCCs for extraction, and other parameters used. Once the parameters are initialized, signals are framed, in which the vowel phonation signals are split into overlapping frames of a specified size. Once the signals are framed, a windowing technique is implemented, which aids in reducing the spectral leakage that appears when examining finite duration

signals. More specifically, the Hamming window is used in the proposed work as it tapers the edge of the signals to minimize the distortion in the frequency domain, thereby resulting in a precise analysis of the frequency components of the signals. After this process, a spectrogram is computed for visualizing the frequency content of a signal over time, and for each frame, FFT is applied for obtaining the magnitude spectrum.

Further, the Mel FilterBank is utilized with triangular filters spaced according to the mel scale. The purpose of applying the Mel FilterBank is to extract relevant features from the audio signals that are characteristic of the diseases. Application of this filter aids in capturing the frequency distribution of the vowel sound, thereby making the process of classification much easier depending on the patterns. Eventually, the Mel spectrogram is compressed by utilizing the logarithmic function to reduce the dynamic range of the signal. Later, MFCC is computed by applying DCT to the log mel. DCT can also be applied for further compressing the MFCCs into a reduced set of coefficients while preserving the significant information needed for the model. Further, Standard Scalar is used for removing the mean and scaling each feature to unit variance, and PCA decomposition is used for reducing the dimensionality of the data, thereby reducing the noise of the data to enhance the performance of the model by focusing on the significant features. Thus, Figure 4 shows the features extracted using MFCC.

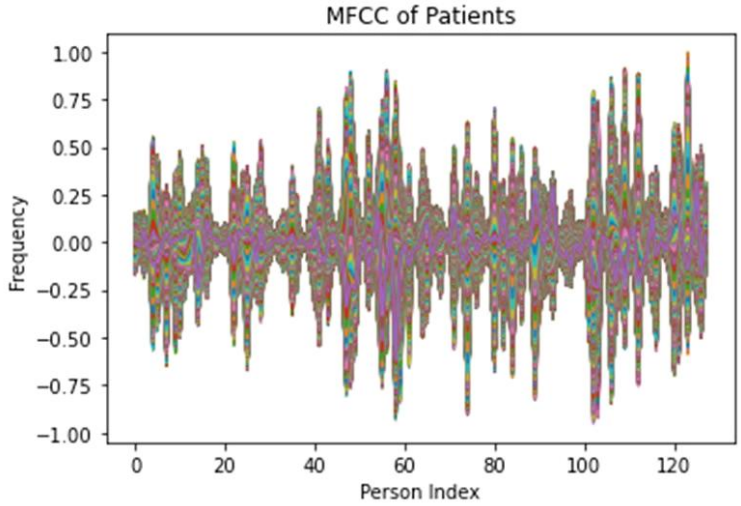

Figure 4. MFCC of patients

From Figure 4, the x-axis identifies the index of each individual coefficient feature extracted from the audio signal, whereas the y-axis denotes the frequency at which each coefficient occurs in the speech signal. It can be depicted that when the index lies at 60, it reaches the frequency level of 1 hertz. By examining the frequency distribution of this coefficient, it can identify the patterns and differences between different phonation types that can be helpful in discriminating between ALS and non-ALS individuals. After MFCC, a standard scalar is used for standardizing the data by centering it around the mean and scaling it to have a unit variance. Therefore, Figure 5 shows the output obtained after the standard scalar.

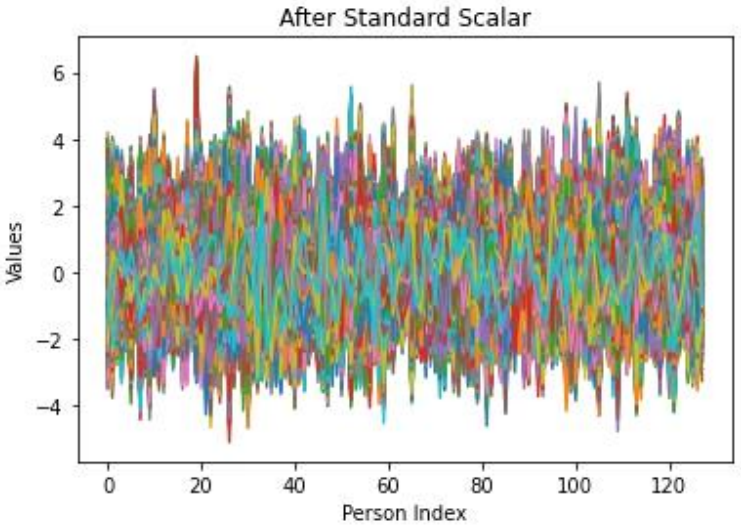

Figure 5. Standard Scalar

Figure 5 depicts the outcome obtained after the standard scalar approach. Employing a Standard scalar ensures that all input features are on the same scale. It also rescales the input features to efficiently learn the relationships between target variables and the features, making the model more robust to outliers. In the figure, the x-axis is denoted as the person index, and the y-axis is denoted as the values. Similar to a standard scalar, PCA decomposition is depicted in Figure 6.

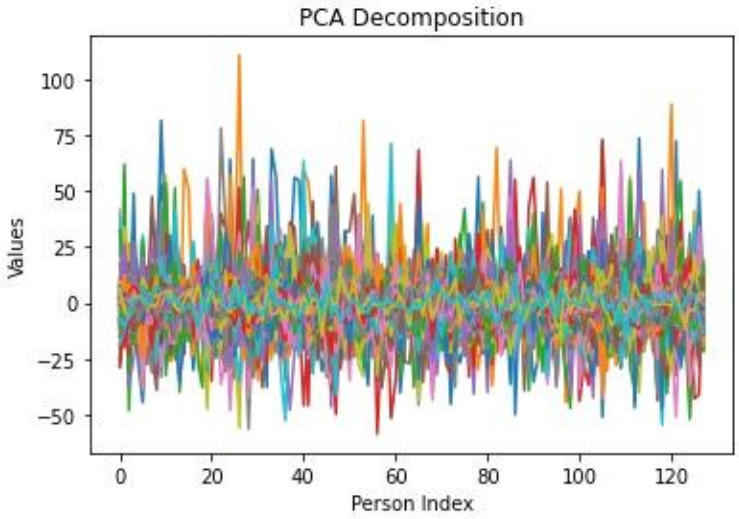

Figure 6. PCA Decomposition

Figure 6 shows the audio signals obtained using PCA decomposition. This technique aided in extracting the relevant features from the sound signals and reduced the dimensionality of the data. Therefore, by using PCA, key acoustic characteristics can be differentiated between ALS and HC. Moreover, PCA aids in removing shaky or lagging features from the audio data, as these can hinder the performance of classification. Once the features are extracted, classification is processed using a Hybrid CNN-LSTM with a rapid dilatenet model.

3-3- Classification Using Proposed CNN-LSTM with Rapid DilateNet Model

CNN is one of the capable models that is preferred for the classification process, as CNN is effective in terms of examining sound data due to its ability to identify patterns and features at different scales. Moreover, the CNN model is considered to be prevailing in capturing the spatial and temporal dependencies in data, which is known to be significant for assessing time series data such as waves of sound. Thus, CNN is used, and the process involved in CNN for audio data is depicted as follows. Initially, audio signals are passed to the input layer as the input of the model. Then, the audio signals are passed to CL (convolutional layer), where CL is responsible for learning features from the audio data.

Further, the activation function is also applied to capturing the complex patterns in the data. Following the activation function, PL (Pooling Layer) is also used. PL is utilized for downsampling the features produced by CL, and it is also used for reducing the spatial dimensions of the features, thereby making the model computationally effective. Then, the features are flattened into 1D vectors and passed via FCL (Fully Convolutional Layer). FCL focuses on classifying the features that are extracted into different classes by optimizing the loss function. Finally, the output layer classifies the patient with ALS and the patient without ALS.

Though the CNN model performs considerably for the classification of ALS patients, it lacks in effectively capturing the temporal dependencies in the data and loss of information during the processing stage due to the fixed input size of the CNN model, which may not be ideal for sound signals with variable length sequences. Moreover, the CNN model lacks in remembering past information, which is assessed to be extremely important for sound signals. Therefore, in order to overcome these shortcomings, CNN can be hybrid with the LSTM model, as LSTM is specially designed to capture the long-range dependencies in the speech signals of ALS patients and comprehend the complex patterns associated with the speech pattern of the patients suffering from ALS. Equation 1 shows the process involved in the CNN-LSTM model for ALS classification. Equation 1 denotes the input values fed to the model.

$$x_i^0 = [x_1, x_2, \ldots, x_n] \quad (1)$$

Once the input is fed, it is passed to the CL layer, where the process involved is depicted in the Equation.

$$c_i^{ld,j} = act(b_j + \sum_{m=1}^{M} w_m^j x_{i+1m-1}^{0j}) \quad (2)$$

where, act is defined as the activation function, l is defined as the layer index, b is represented as bias term, M is represented as size of the kernel and w_m^j is defined as the weight for j^{th} feature map. Once it passes through CL, signals are passed to the max PL. Equation 3 represents the process of max PL as it aids in reducing the feature size.

$$p_i^{ld,j} = c_{i \times T+r}^{ld,j} \quad (3)$$

In Equation 3, T is represented as the pooling stride, and R is represented as the size of the pooling window. Once the spatial dimensions are reduced using max PL, output of max PL is passed to the LSTM network. LSTM network commonly consists of LSTM units such as input gate, output gate, and forget gate. LSTM is an RNN model, and the RNN model can typically minimize the complexity of the network and enable training by utilizing the states of the present neuron and the states of the previous neurons. Typically, the LSTM unit recollects values in any time interval, and the flow of information into and out of the LSTM unit is controlled by these 3 gates. Figure 7 shows the architecture of the LSTM model.

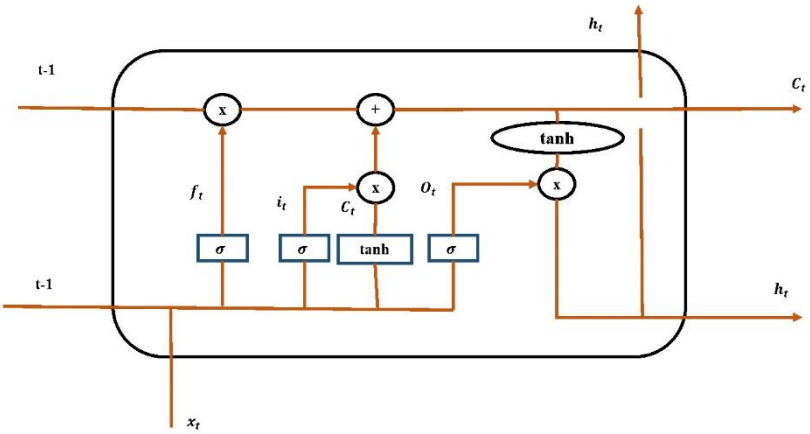

Figure 7. Architecture of LSTM

Here, σ is represented as a sigmoid function, f_t is represented as a forget gate, i_t is denoted as the input gate, and O_t is defined as the output gate, $t-1$ is defined as the cell state, C_t is represented as the candidate gate. Implementation of LSTM for ALS classification aids in examining the nuances and temporal patterns in the vowel sound for making precise and correct predictions with patients suffering with ALS. The forget gate implemented in the model is depicted in Equation 4.

$$f_t = \sigma\left(w\left[x_t, act_{t-1}, C_{t-1}\right] + b_f\right) \quad (4)$$

where x is denoted as the input sequence, act_{t-1} is represented as the output of the preceding block, bias vector is represented as b_f, C_{t-1} is represented as the previous memory block of LSTM, σ is denoted as the sigmoid function, and separate weight vectors for each input is represented using W. Input gate is a section, where a new memory is generated by using a trivial neural network with $tanh$ activation function and this is depicted Equations 5 and 6.

$$i_t = \sigma\left(W\left[x_t, act_{t-1}, C_{t-1}\right] + b_i\right) \quad (5)$$

$$C_t = f_t \cdot C_{t-1} + i_t \; tanh \; tanh\left(\left[x_t, act_{t-1}, C_{t-1}\right]\right) + b_c \quad (6)$$

Output gate is the section, where output generated by the current LSTM block is generated by using output gate and these outputs are estimated using Equations 7 and 8.

$$\sigma_t = \sigma\left(W\left[x_t, act_{t-1}, C_t\right] + b_o\right) \quad (7)$$

$$act_t = o_t . tanh\left(C_t\right) \quad (8)$$

Thus, the connection between the units of LSTM permits the information to cycle between adjacent time steps. This produces an internal feedback state that allows the network to comprehend the concept of time and understand the temporal dynamic in the present data. The combination of CNN and LSTM architectures in the proposed model effectively captures both spatial and temporal features from vowel phonation, by enhancing ALS classification accuracy as CNN model extract spatial features from audio spectrograms by applying convolutional layers that identify local patterns, such as formants and harmonics. This spatial analysis allows the model to recognize distinctive vocal characteristics associated with ALS and LSTMs are designed to handle sequential data this process extracted spatial features over time, by enabling the model to understand how vocal characteristics evolve during phonation. This temporal analysis is essential for identifying subtle changes in speech patterns indicative of ALS. Although the CNN-LSTM model can deliver considerable results, the effectiveness of the model for classifying sound signals can be challenging. Thus, proposed Rapid DilateNet function is applied. Hence, the proposed work focuses on employing Rapid DilateNet for the

effective classification of ALS patients using their speaking manner. Rapid DilateNet is implemented in the proposed model for solving the context information for interpreting and analyzing the sound of vowels accurately and correctly without any loss of information. The absence of context information can lead to errors while analyzing the sound signals. Thus, the proposed rapid DilateNet model helps with context information for providing reliable outcomes. Although the CNN-LSTM model can deliver considerable results, the effectiveness of the model for classifying sound signals can be challenging. Thus, the Rapid DilateNet function is applied.

Hence, the proposed work focuses on employing Rapid DilateNet for the effective classification of ALS patients using their speaking manner. Figure 8 shows the process involved in the proposed CNN-LSTM model with Rapid DilateNet.

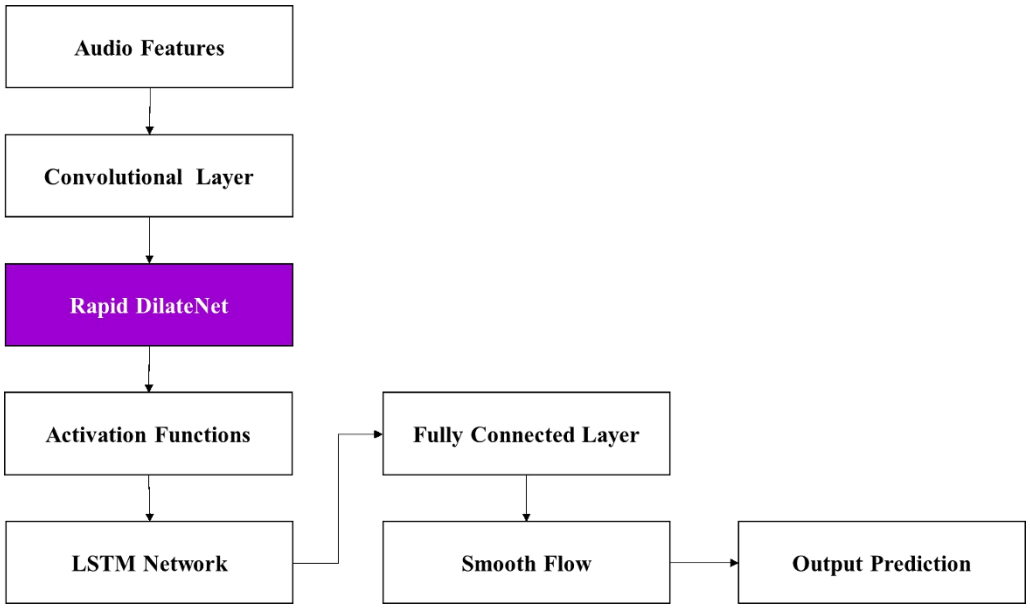

Figure 8. Proposed CNN-LSTM with Rapid DilateNet

Here, the process is initiated by passing the extracted features from the mel frequency feature extraction technique to CL. CL with $64\ filters, 128\ and\ 256\ filters$ are used in the proposed model, and from CL, it is further passed to the proposed rapid DilateNet model. This rapid DilateNet model permits the model to see a larger receptive field by maximizing the dilation rate without employing additional parameters, thereby making the model computationally effective for classification of patients suffering from ALS. Moreover, in conventional neural network models, context information is typically augmented by extending the receptive field, and this extension of the receptive field can be accomplished by either increasing the size of the network or by enlarging the size of the convolution kernel and dilation rates. However, these techniques can tremendously increase the training time. Thus, rapid DilateNet is implemented in the proposed model for solving the context information for interpreting and analyzing the sound of vowels accurately and correctly without any loss of information. The absence of context information can lead to errors while analyzing the sound signals. Rapid dilated convolution enhances the performance of the CNN-LSTM model by allowing it to capture the input audio signals without increasing the number of parameters. Dilation introduces gaps between the kernel elements and enables the model to analyze larger portions of the audio signal at once. By identifying these subtle changes in vowel phonation, it enables the model to recognize patterns with longer temporal ranges. By preserving this context information, the dilation ensures the model retains the features related to the phonetic characteristics of vowels by distinguishing between ALS and non-ALS patients. The combination of CNN and LSTM with rapidly dilated convolutions results in a feature extraction process by enhancing the model's ability to interpret complex audio patterns effectively. Thus, the proposed rapid DilateNet model helps with context information for providing reliable outcomes.

After employing rapid DilateNet, an activation function is used. The proposed model utilizes the ReLU activation function, as the ReLU activation function is more effective for handling the nonlinearity present in the data, and it can also prevent gradient vanishing issues. Further, the LSTM network is used, and a dropout layer of 0.5 is added to the output passed from the LSTM network. This implementation of the dropout layer also prevents the overfitting of the proposed model. In order to overcome overfitting, the proposed model employed dropout at a rate of 0.5, which involves randomly deactivating half of the neurons during training. This keeps the model from becoming overly dependent on particular patterns in the training data. By motivating the network to acquire more resilient, generalized features, this introduces regularization. Nevertheless, considering the small dataset and lack of speaker-independent evaluation, the incredibly high reported metrics (such as 99.99% accuracy) raise questions about potential overfitting. Although dropout aids the model's capacity to generalize to new data. Eventually, the classification of ALS patients based on vowel phonations is predicted. Figure 9 shows the conventional convolutions and proposed rapid DilateNet convolutional.

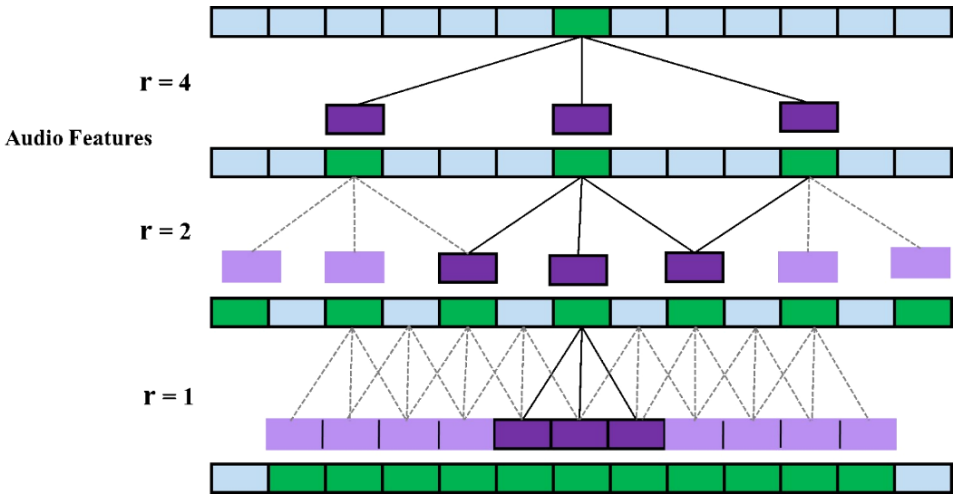

Figure 9. Rapid DilateNet Convolutions

In Figure 10, 1D CNN with 3 conventional convolutions is added, and the expansion rate $'r'$ for each layer is 1. Similarly, 1D CNN with a dilated convolutional layer is present in Figure 9, where the expansion rate of each layer is $r = 1$ for the first layer, $r = 2$ for the second layer, and $r = 4$ for the third layer. The top green unit present in Figure 9 and 11 is denoted as the unit of interest, whereas the other green unit signifies its receptive field in each respective layer. However, when compared to conventional convolutions, dilated convolution expands its receptive field in the convolutional kernel without surging the parameters.

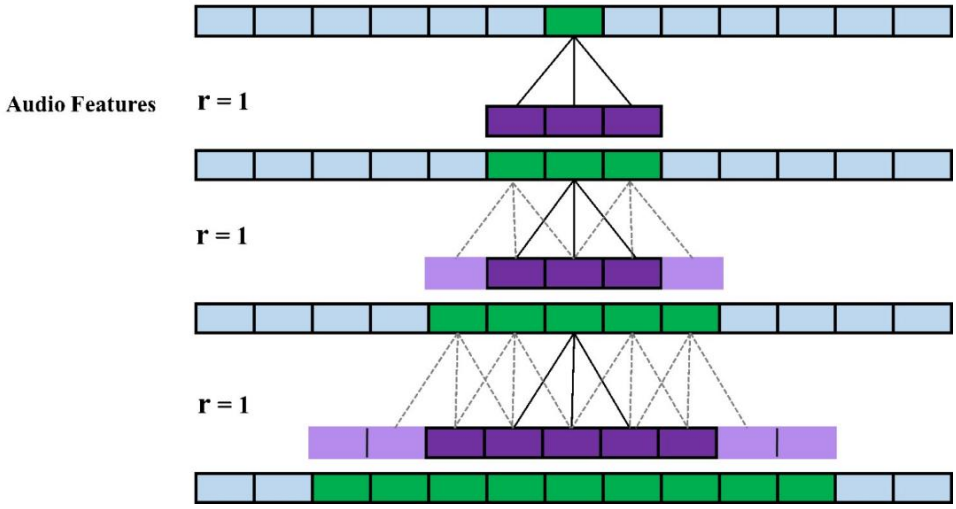

Figure 10. Conventional Convolutions

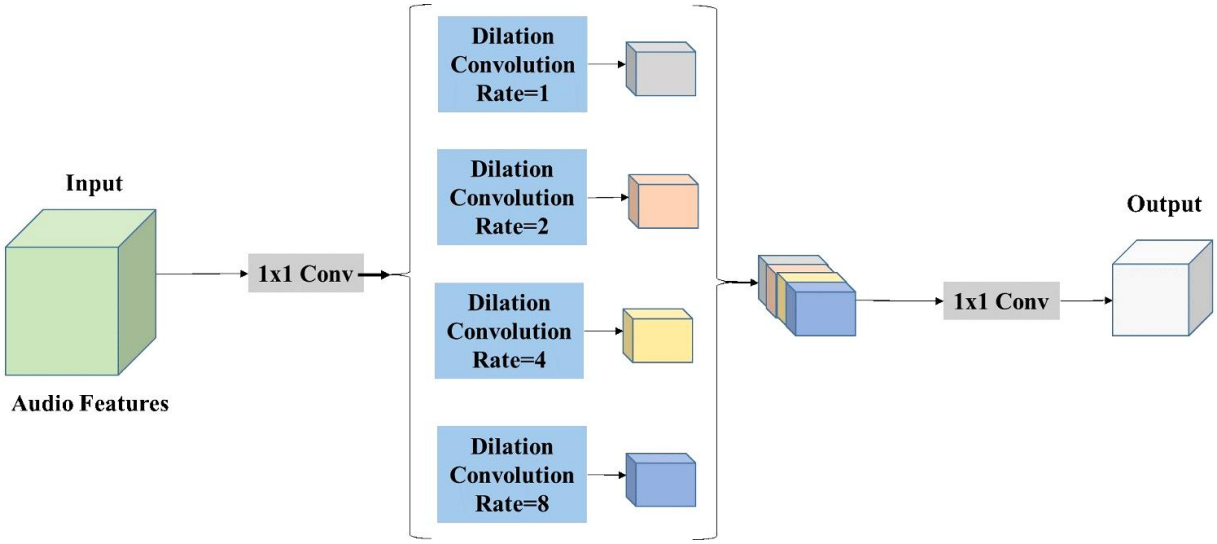

Figure 11. Rapid DilateNet Rate

When $r = 1$, the original convolutional kernel size is 3×3. This can be stated as conventional convolution. Whereas, when $r = 2$, receptive field is 7. Likewise, when $r = 4$, receptive field is 15. Thus, the implementation of dilation convolution can result in a grid effect when assorted dilated convolutions are stacked. Thus, the proposed model uses rapid dilatenet model which maximizes the receptive field, thereby increasing the clarity of the output of the vowels which helps in comprehend the phonations by concatenating the dilation convolutions, which is represented in Figure 11.

Besides, Figure 12 depicts the allocation of different weights to the model.

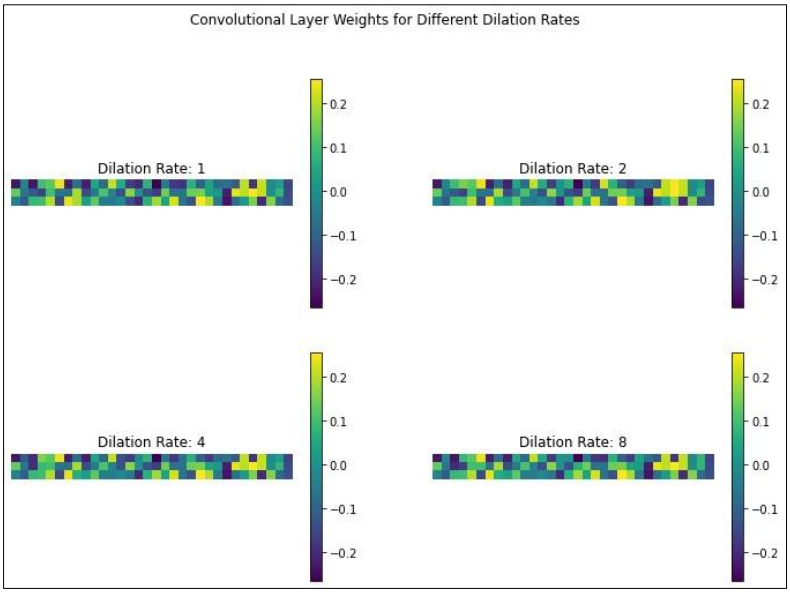

Figure 12. Weights for different dilation rates

Weights for each dilation rate would be learned during the training process, and the weights of the models are adjusted according to account for different spacing between elements. Moreover, adjusting the weights aids in optimizing the performance of the given tasks, as different weights are given for different dilation rates. The pseudocode for the proposed dilate is depicted in Algorithm II.

Algorithm II. Proposed Rapid DilateNet

```
Define a function named Rapid_Dilate_Block (inputs, filters, and dilation_rates):
  Initialize x as inputs
For each rate in dilation_rates:
Apply a 1D convolutional layer with filters, kernel_size, padding, dilation_rate, and ReLU activation to x.
  Return x
Define a list of dilation_rates [1, 2, 4, 8]
For each rate in dilation_rates:
  Create an input layer with shape (100, 1)
     Apply rapid_dilate_block to inputs with filters and the current rate
     Apply a 1D max pooling layer with pool_size to the result
     Apply an LSTM layer to the result
     Apply a dense layer and ReLU activation to the result
     Apply a dropout layer with rate of 0.5 to the result
     Apply a dense layer and sigmoid activation to the result to get outputs
  Create a model with inputs as input and outputs as output
Compile the model
  Evaluate the model on test data and store the test loss and accuracy in variables test_loss and test_acc
Return the test accuracy for the current dilation rate
```

Pseudocode explains the implementation of the proposed rapid dilatenet model. Initially, a 1D convolutional layer is applied with filters. Then, an input layer with shape (100,1) is created. Then, a rapid dilate block is implemented. Further, the LSTM layer is used, and the dense layer and ReLU activation function are employed. After applying those layers, a dropout layer with the rate of 0.5 is employed along with sigmoid activation to fetch results. Finally, the efficacy

of the proposed model will be identified by using metrics. Therefore, the efficiency of the proposed model is evaluated in the subsequent section. The parameters to train the model is depicted in Table 5.

Table 5. Parameters used in the model

CNN Parameters • Filter Size: Convolutional Layers: 3×13×1 (for 1D data) • Number of Filters:32 • Activation Function: Relu • Dilation Rates:[1, 2, 4, 8] (one layer for each rate) • Pooling Layer Type: Max Pooling • Pooling Size: 22
LSTM Parameters • Number of LSTM Units: 64 • Number of LSTM Layers: 1 • Dropout Rate: 0.5
Dense Layer Parameters • Number of Dense Units: 128
Output Layer Parameters • Output Units: 1 • Activation Function: Sigmoid

4- Results and Discussion

This section elucidates results obtained using the proposed model, such as EDA, performance analysis, and comparative analysis.

4-1- Performance Metrics

Performance of the proposed mechanism can be gauged by using metrics such as accuracy, sensitivity and specificity.

A) Accuracy

Accuracy is represented as a metric that describes the performance of the proposed model across all classes. Equation 9 depicts the mathematical formula for accuracy,

$$Accuracy = \frac{TN+TP}{TN+FN+TP+FP} \tag{9}$$

where TN denoted as true negative, TP is denoted as true positive, FN is defined as false negative and FP is defined as false positive.

B) Sensitivity

The sensitivity is defined as the ratio of TP with TP and FN, and it is given in Equation 10

$$Sensitivity = \frac{TP}{TP+FN} \tag{10}$$

C) Specificity

The specificity is defined as the ratio among the TN with the TP combined with FP, and it is mathematically given in Equation 11, and it is as follows.

$$Specificity = \frac{TN}{TN+FP} \tag{11}$$

4-2- Exploratory Data Analysis

EDA for audio signals aids in analyzing and visualizing the audio data to gain insights and comprehend the characteristics of the signals. EDA helps with determining the features that are relevant to the tasks and helps with understanding the patterns and structure of the audio data. Besides, EDA analyzes data distribution, identifies outliers, and understands the relationship between variables, thereby aiding the model for effective classification of ALS and non-ALS. Therefore, EDA is extremely important for examining the patterns of the signals within the audio data. Figures 13 to 15 depict the person suffering from ALS.

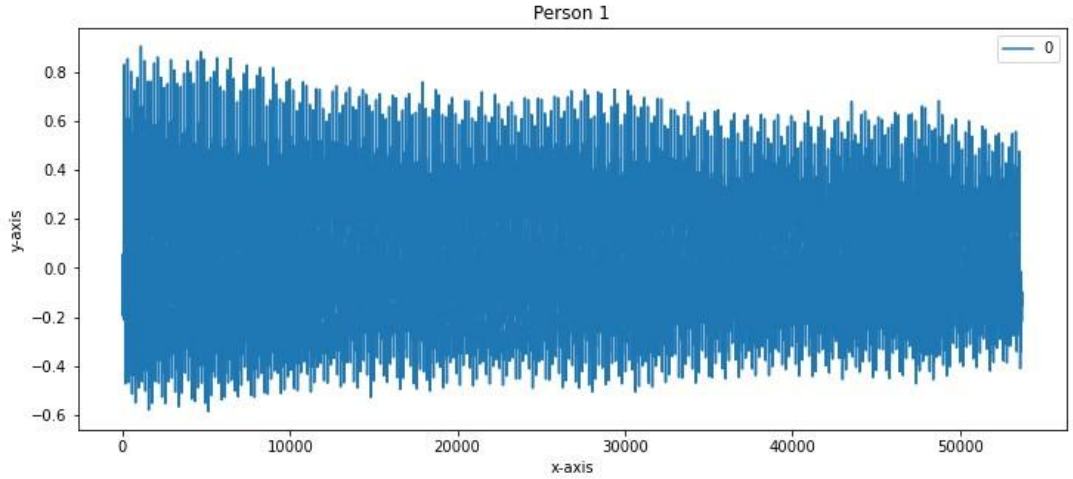

Figure 13. ALS affected Person 1

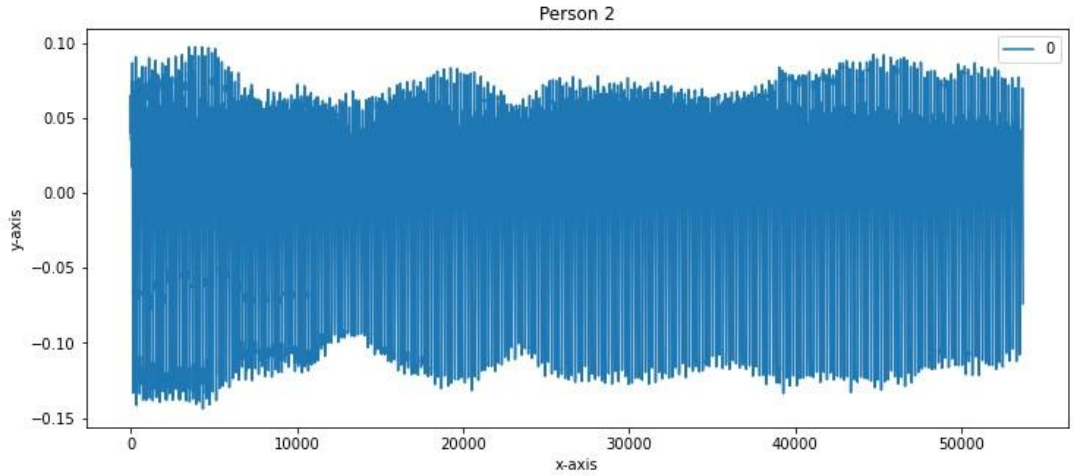

Figure 14. ALS affected Person 2

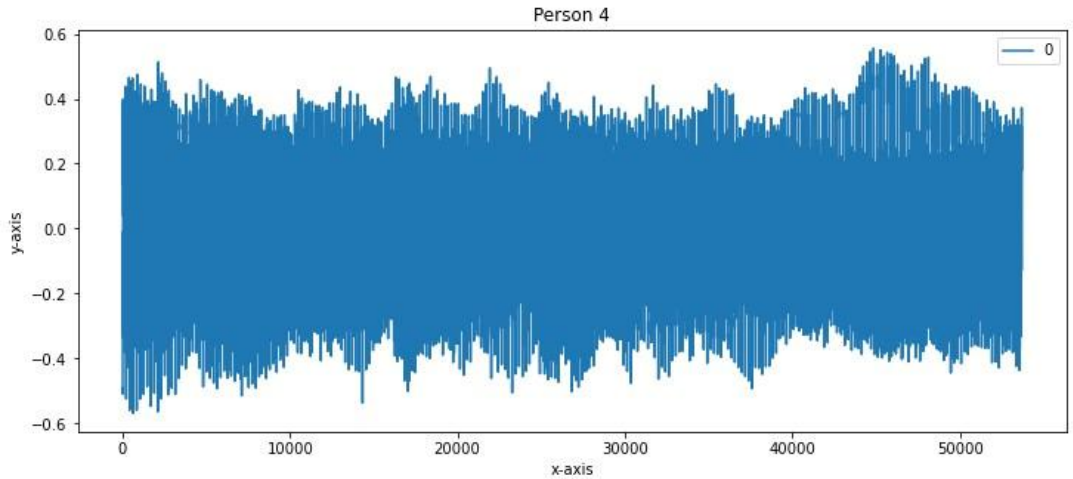

Figure 15. ALS affected Person 4

From the Figures 13 to 15 it is concluded that individuals 1, 2, and 4 have ALS because the data of these individuals probably include typical features of the condition, including muscular weakness, irregular electromyography (EMG) activity, or proof of motor neuron breakdown. These characteristics are typically used to diagnose ALS and reflect progressive deterioration in muscle control and function. On the other hand, Person 3, does not reflect these abnormalities in the respective figure, indicating typical muscle and nerve function without abnormalities of degeneration or weakness. Thus person 3 is classified as not being affected by ALS, and this illustrates clearly in the figures the difference in diagnostic markers at the affected and unaffected individuals (Figure 16).

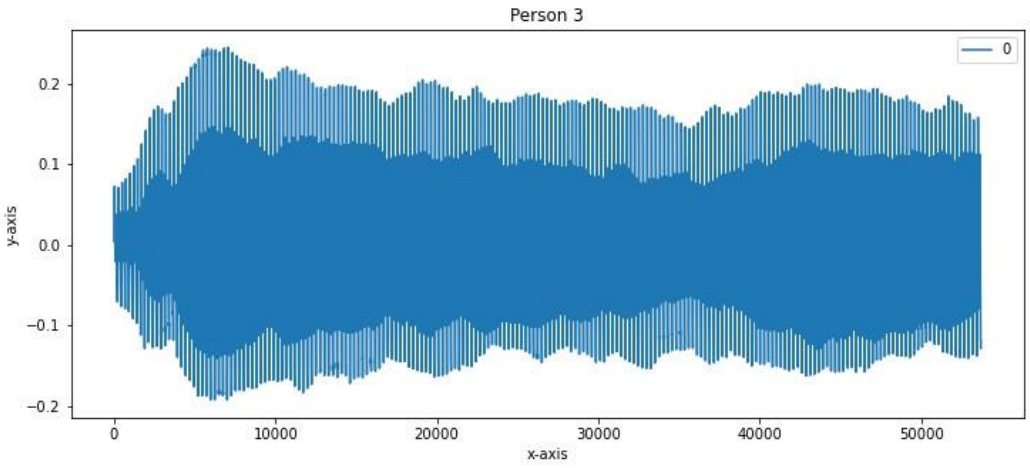

Figure 16. Normal Patient (person 3)

Similarly, the frequency power distribution of person 1, person 2, person 3, and person 4 is depicted in Figures 17 to 20. In signal processing, frequency distribution refers to how much of a signal's power is distributed across different frequency components. It helps to understand which frequencies are dominant or prevalent in the signal. Likewise, the power spectrum shows how the power of a signal is distributed across different frequencies. By calculating the FFT of the audio data, the signal can be converted from the time domain to the frequency domain, representing the signal as a combination of different sine waves of varying frequencies and amplitudes.

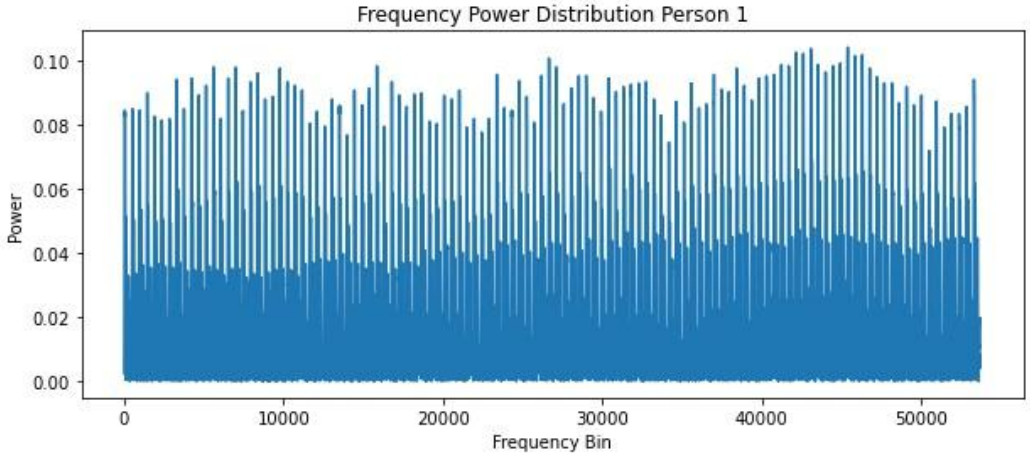

Figure 17. Frequency power distribution of person 1

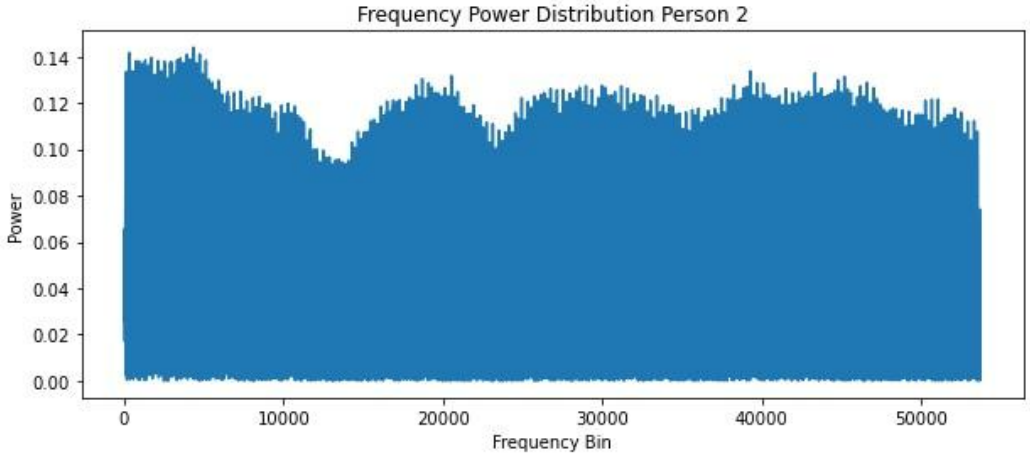

Figure 18. Frequency power distribution of person 2

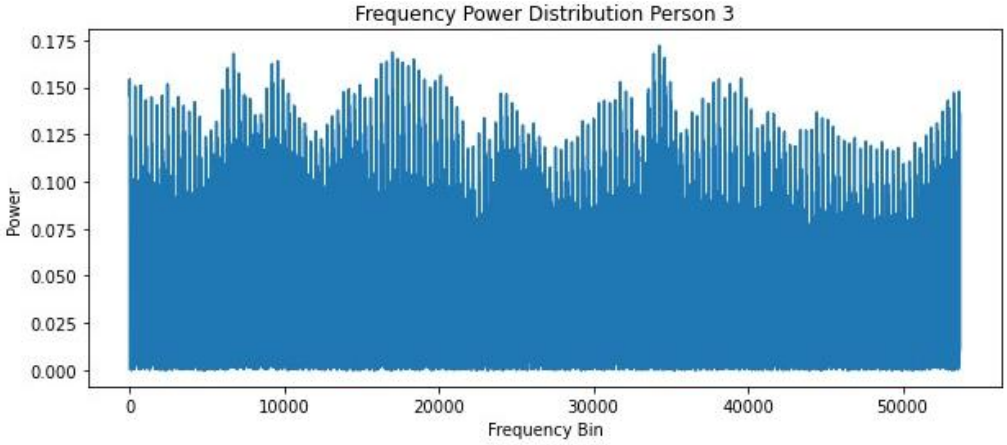

Figure 19. Frequency power distribution of person 3

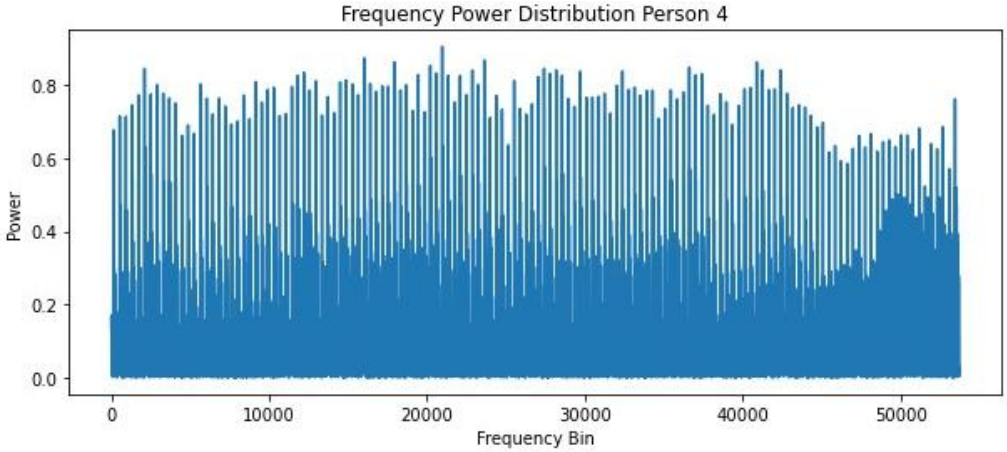

Figure 20. Frequency power distribution of person 4

Figures 17 to 20 show plots of the power distribution of a signal over different frequency bins, which is widely known as the power spectral density (PSD). The x-axis in these plots is discrete frequency bins for each covering a particular frequency band, while the y-axis is the power in each bin, reflecting how much energy of the total signal is occurring at each frequency. Every point on the plot thus indicates the strength of the signal's power at a certain frequency, and one can determine dominant frequencies or frequency bands in which the energy of the signal is pooled. This analysis is important for the understanding of the spectral properties of signals, as it indicates the global power and the power distribution along the frequency domain, which is essential for diagnosing system behaviors, detecting sources of noise, or characterizing biological signals.

The histogram in Figure 21 shows the distribution of audio sample values (amplitudes) over time. The x-axis represents the amplitude values of the audio signal, and the y-axis shows the frequency, or how often these amplitude values occur. The audio values are centered around 0, which is typical for audio signals that have been normalized. The distribution appears roughly symmetric around 0, indicating that the audio data oscillates between positive and negative values without a strong bias in either direction.

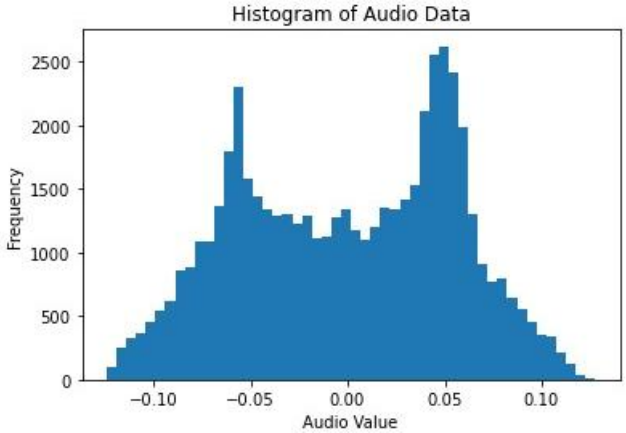

Figure 21. Histogram of Audio Data

Time-domain representation of the audio signal is depicted in Figure 22. The x-axis represents time (in sample points, likely corresponding to milliseconds or seconds), and the y-axis represents the amplitude of the audio signal at each point in time. The audio signal oscillates around zero, with fluctuations in amplitude. This reflects the oscillating nature of sound waves, which have both positive and negative phases as they oscillate. In terms of amplitude, there's a relatively steady range between approximately -0.1 and +0.1, with some variations, which aligns with the observations from the histogram.

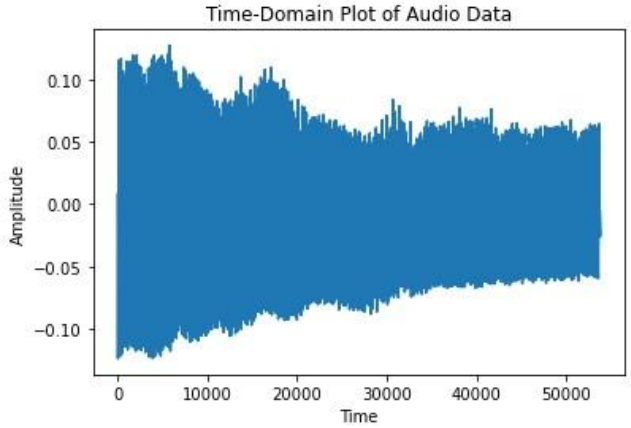

Figure 22. Time Domain plot of Audio Data

Spectrogram is showcased in Figure 23, where x-axis represents time in seconds, ranging from 0 to approximately 5.6 seconds. The y-axis represents frequency in Hertz (Hz), spanning from 0 Hz to 8192 Hz. The color intensity indicates the amplitude (or strength) of the frequencies at any given time. The scale on the right shows that lighter colors (yellow/white) represent higher intensity (close to 0 dB), while darker colors (purple/black) represent lower intensity (down to -80 dB). The evenly spaced, bright horizontal bands indicate strong frequency components around 128 Hz, 256 Hz, 512 Hz, and 1024 Hz, along with several higher harmonics up to around 4096 Hz. These patterns are consistent throughout the entire duration of the recording, suggesting a sustained tonal or harmonic sound that remains relatively stable over time.

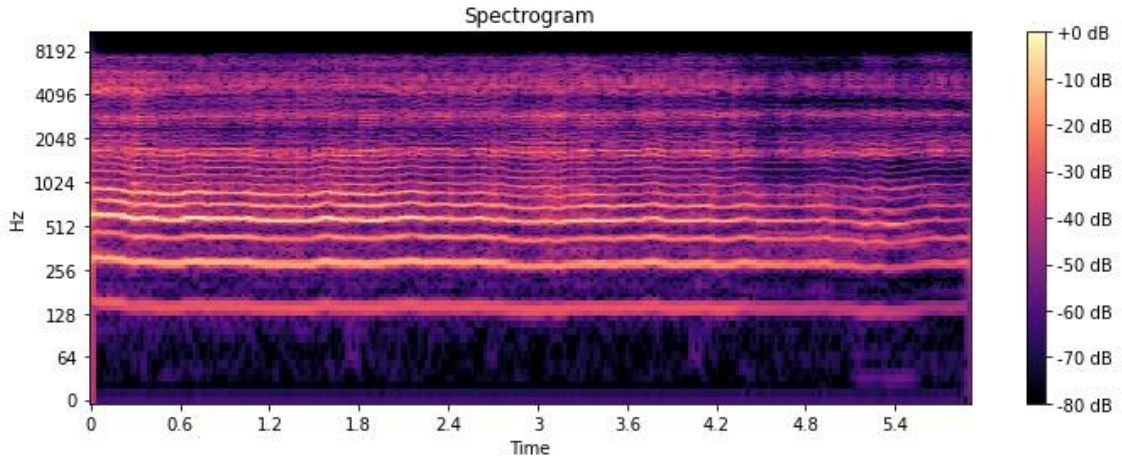

Figure 23. Spectrogram

4-3- Performance Analysis

Performance analysis focuses on results obtained by using the proposed model for the classification of ALS and non-ALS patients based on vowel phonation. Thus, Table 6 shows the test accuracy and test loss obtained by the model proposed model for different dilate rates used in the proposed model.

Table 6. Test accuracy and test loss

Dilation Rate	Test Accuracy	Test Loss
1	0.9965	0.023589
2	0.9958	0.0258796
4	0.9989	0.01189756
8	0.9996	0.011627456

Table 6 presents the test accuracy and test loss obtained for different dilation rates. When the dilation rate was set to 1, the model achieved a test accuracy of 0.996 with a test loss of 0.0235. Similarly, at a dilation rate of 2, the values were 0.9958 for test accuracy and 0.0258 for test loss. For a dilation rate of 4, the accuracy improved to 0.998, while the test loss decreased to 0.01189. Notably, when the dilation rate reached 8, the highest accuracy of 0.9996 was observed, with a corresponding test loss of 0.01162.

These variations indicate that increasing the dilation rate enables the proposed model to extract more detailed and complex features from the data, leading to more precise classification between ALS and non-ALS cases. A higher dilation rate also improves gradient flow, which contributes to achieving better optimization. Figures 24 and 25 provide graphical representations of the test accuracy and test loss. As shown, the lowest accuracy (0.995) and highest test loss (0.02587) occurred at a dilation rate of 2.

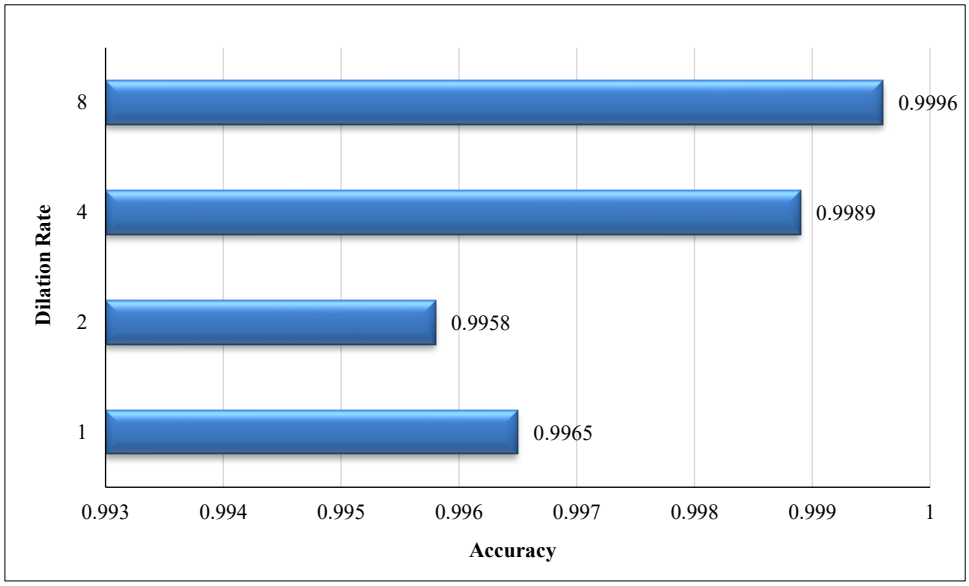

Figure 24. Graphical Representation Test accuracy

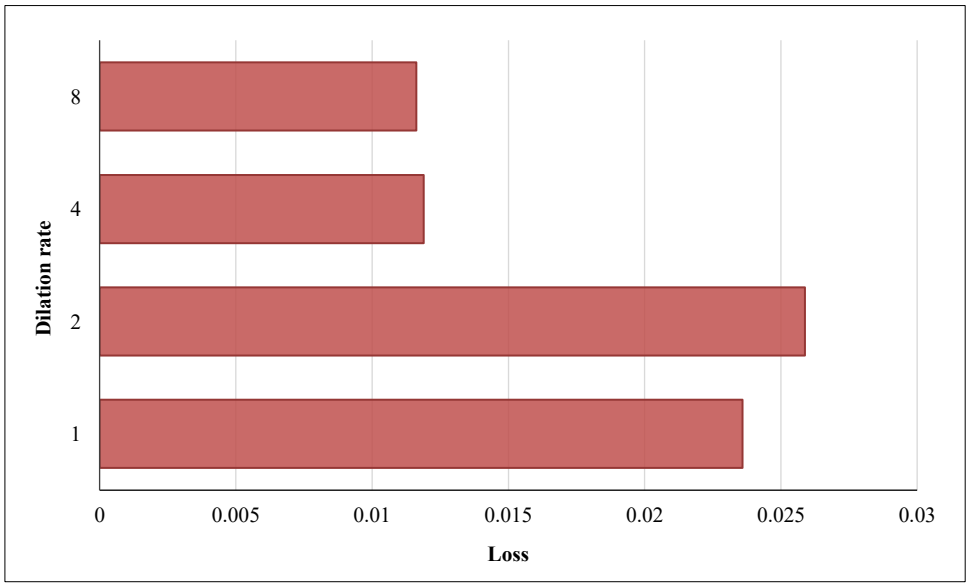

Figure 25. Graphical Representation Test Loss

Similarly, the model accuracy and model loss corresponding to different dilation rates are illustrated in Figures 26 and 27. Model accuracy reflects how effectively the proposed mechanism can correctly predict outcomes, and it is typically calculated by dividing the number of correct predictions by the total predictions made. Accordingly, Figure 25 depicts the training and validation accuracy for the various dilation rates. During training, the model is iteratively updated across multiple epochs, and after each epoch, both training accuracy and validation accuracy are evaluated to monitor the model's performance as it learns.

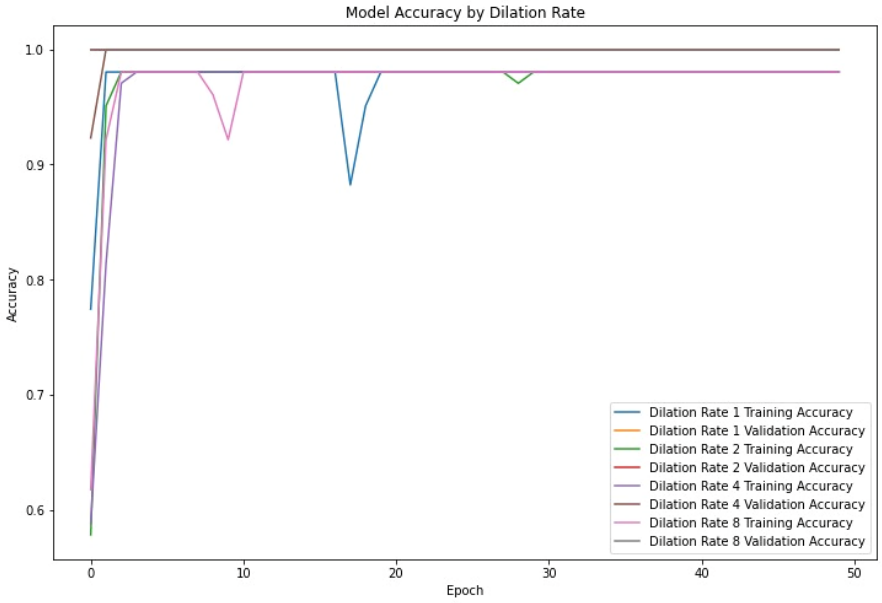

Figure 26. Model accuracy by dilation rate

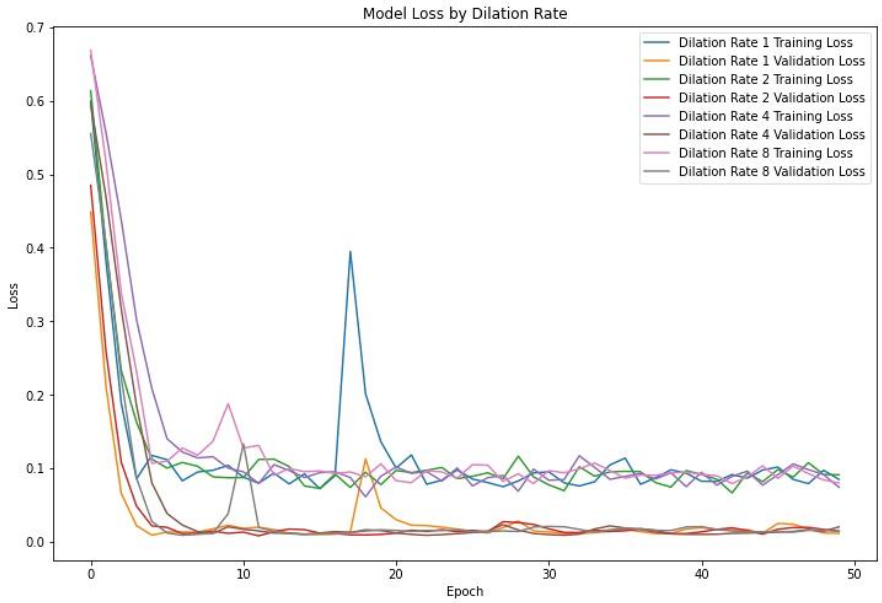

Figure 27. Model loss by dilation rate

Figure 26 illustrates the training and testing accuracy for different dilation rates. Training accuracy refers to the model's performance on the data it has already seen during training, while validation accuracy assesses how well the model generalizes to unseen data. The proposed model achieved higher training and validation accuracy, particularly at dilation rates of 4 and 8, as evidenced by the curves in Figure 26, demonstrating the effectiveness of the model in ALS classification.

Similarly, Figure 27 presents the training and validation loss for different dilation rates. Training loss measures how well the model fits the training data during the learning process, whereas validation loss is used to evaluate the model's performance on unseen data and to prevent overfitting. As shown in Figure 27, the training loss is relatively high when the dilation rate is 1, while it decreases significantly at a dilation rate of 8. This indicates that increasing the dilation rate improves the model's performance. Furthermore, Table 7 provides a comparison of models with and without dilation rates, showing both training accuracy and training loss.

Table 7. With and Without Dilation Rate

Model	Test Accuracy	Test Loss
With Dilation Rate	0.9996	0.011627456
Without Dilation Rate	0.9799	0.096587

Table 7 presents the test accuracy and test loss for models *with* and *without* dilation rate. The model with dilation achieved a test accuracy of 0.9996 and a test loss of 0.01162, whereas the model without dilation obtained a test accuracy of 0.9799 and a test loss of 0.096. These results clearly indicate that the model incorporating dilation outperformed the one without it. The superior performance is attributed to the ability of the dilation rate to reduce spatial information loss, capture long-range dependencies, and extract more relevant features for the proposed model. Consequently, the inclusion of the dilation rate significantly enhances classification performance. Figures 28 and 29 provide a graphical representation of the results summarized in Table 7.

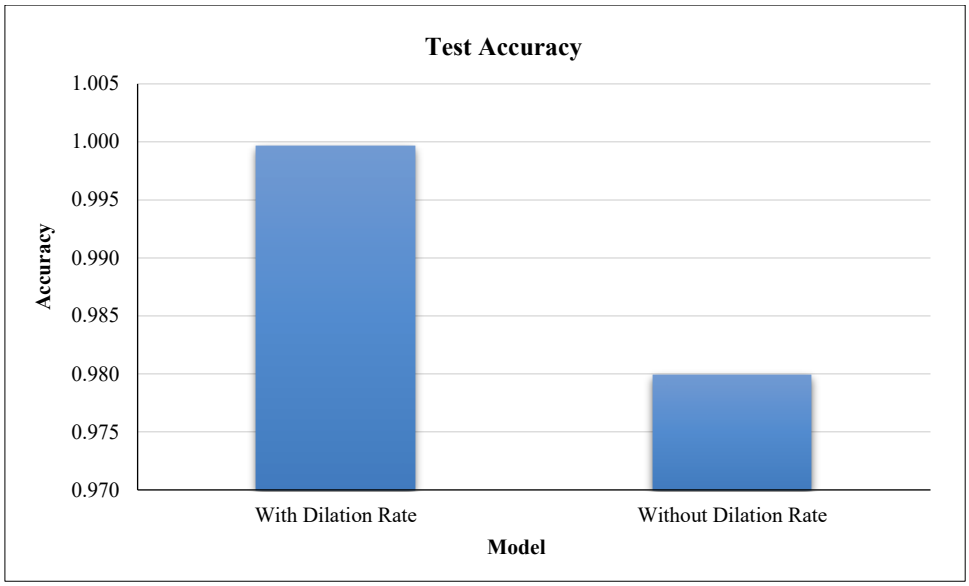

Figure 28. Test accuracy with and without dilation rate

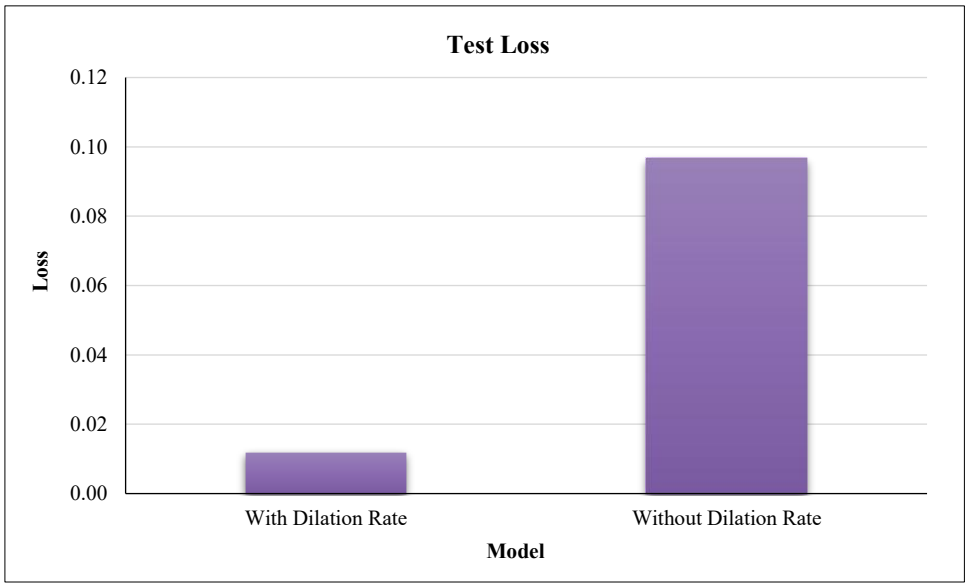

Figure 29. Test loss with and without dilation rate

Figure 28 shows that the accuracy difference between the 'with dilation rate' model and the 'without dilation rate' model is 1.9904%. Likewise, the difference of test loss between the 'with dilation' and' without dilation rate' models is 157.021% in Figure 29. Similarly, Table 8 depicts the models 'with PCA without SC', 'without SC without PCA', and' both present'.

Table 8. Standard scalar and PCA

Model	Test Accuracy	Test Loss
Without SC Without PCA	0.979657	0.0915863
With SC Without PCA	0.986985	0.058913
Both Present	0.9996	0.011627456

Table 8 highlights test accuracy and test loss. Test accuracy obtained for SC without, with SC without PCA, and both present is 0.9796, 0.9569, and 0.9996. Likewise, test loss attained by without SC without, with SC without PCA, and both present is 0.0915, 0.0589, and 0.01162. Higher accuracy and low-test loss are achieved by using both SC and PCA. Figures 30 and 31 show the Graphical Representation test accuracy and Graphical Representation test loss.

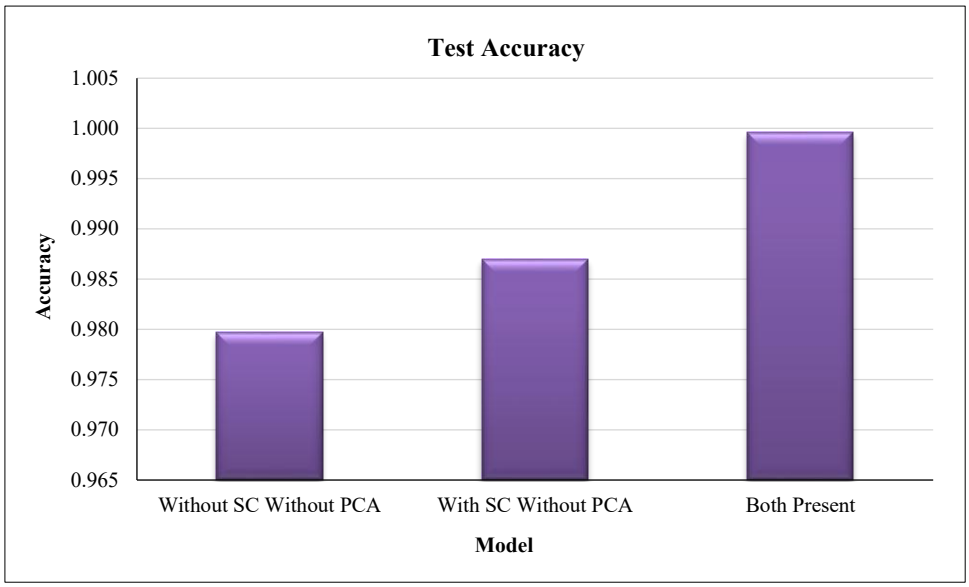

Figure 30. Graphical Representation Test accuracy

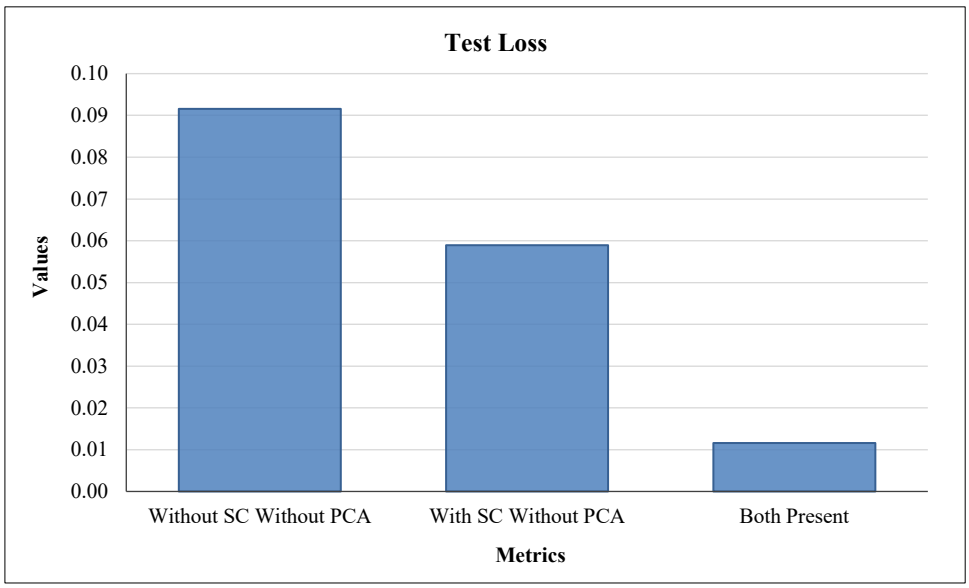

Figure 31. Graphical Representation Test Loss

Though the proposed model has delivered better performance for classifying ALS and non-ALS patients based on vowel phonations, it is important to assess the efficacy of the proposed work with other state-of-the-art approaches. Hence, a subsequent section compares the existing work with the proposed CNN-LSTM with a rapid dilatnet model.

4-4- Comparative Analysis

Comparative analysis is carried out by comparing the proposed CNN-LSTM with the rapid dilate model with the existing models. Thus, Table 9 showcases the accuracy, sensitivity, and specificity of the existing and proposed mechanisms [37]. From Table 9, it can be identified that accuracy, sensitivity, and specificity attained by the existing model are 99.7%, 99.3%, and 99.9%, whereas accuracy gained by the proposed model is 99.99%, sensitivity gained by the proposed work is 99.78%, and finally specificity accomplished by the proposed model is 99.999%. Differences show that the proposed model has gained 0.29045% more accuracy than the proposed work. Likewise, the gain of sensitivity and specificity of the proposed model is 0.48% and 0.09% higher than the existing model, and this is primarily due to the hybrid nature of the CNN-LSTM model, along with the inclusion of rapid dilatant function. Figure 32 shows the comparative analysis of the proposed work.

Table 9. Comparative analysis

Model	Accuracy	Sensitivity	Specificity
Existing Model	99.7	99.3	99.9
Proposed Model	99.99	99.78	99.99985

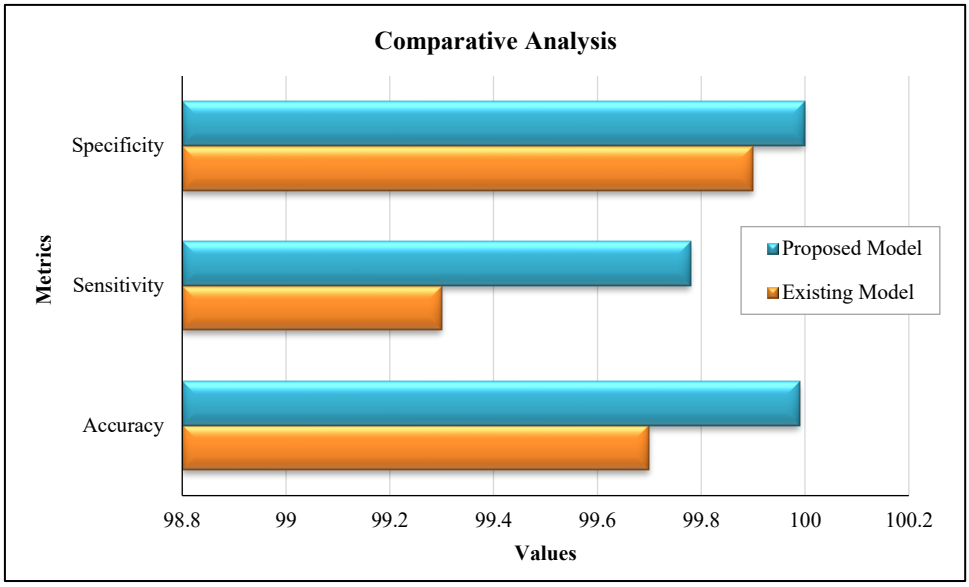

Figure 32. Comparative analysis

From the experimental outcome, it can be identified that better performance and accuracy for classification of ALS and non-ALS are based on the vowel phonation/a/ and /i/. Better performance was delivered due to the incorporation of the proposed CNN-LSTM with a rapid dilatenet model for classification. However, extraction of better relevant features has been primarily carried out using the MFCC model, and the slightest changes in the phonation of vowels are detected by maximizing the expansion/dilation rate and aiding the context information for interpreting and analyzing the sound of vowels accurately and correctly without any loss of information. Therefore, better accuracy is obtained by the proposed mechanism for ALS classification by also preventing overfitting drawbacks.

4-5-Statistical Analysis

Statistical analysis is enhanced significantly by incorporating confidence intervals. A confidence interval is an estimate derived from sample data that indicates a range within which the true population parameter is expected to lie. Figure 33 shows the confidence interval of the study.

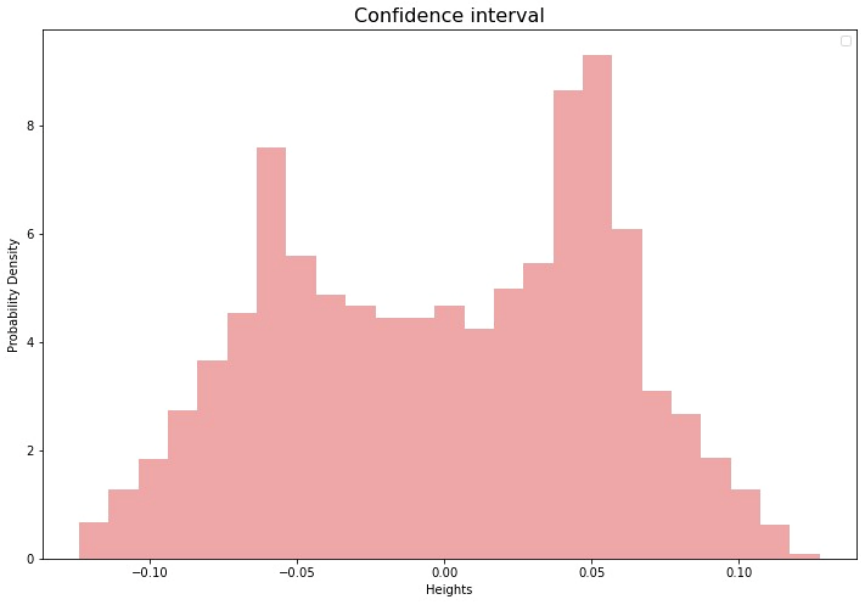

Figure 33. Confidence Interval

Figure 33 shows the distribution is roughly bimodal, with two main peaks such as 0.05 and -0.05. This shows that the dataset has two frequently occurring ranges of values around these points, with lower density of values between them, around 0. The distribution is relatively symmetric with similar patterns of density on both the negative and positive sides.

The proposed work concludes that voice analysis illustrates potential as a non-invasive and objective technique for characterizing motor speech deficits in ALS. The potential contribution of the research work using vocal characteristics to aid clinical progression facilitated better detection of ALS. Moreover, continued exploration in this realm of research can advance the comprehension of the pathophysiology of ALS, thereby enhancing the quality of life of individuals affected by this devastating disease. Though the proposed model has delivered an effective outcome, the limitation of the proposed work is that it only uses two vowels, such as /a/ and/i/, as the dataset encompasses two vowels, and moreover, the model has only opted for one dataset, which can be overcome in the future by using different datasets.

5- Conclusion

ALS is one of the fatal diseases that needs to be detected as early as possible; therefore, the proposed work focused on utilizing better methods for the effective classification of ALS and non-ALS patients. Thus, it was accomplished by using the Minsk2020 dataset, which consists of data collected from various patients using smartphones. As audio signals are tricky, it is important to extract features that are needed for the model. Thus, the proposed work utilized MFCC for extracting the relevant features needed for the model; after extracting features, a proposed CNN-LSTM with a rapid dilatant model was used. This ensured the identification of even the slightest changes from non-ALS patients and ALS patients by expanding the dilation rates, thereby increasing the receptive field for better classification using rapid DilateNet. Incorporation of these proposed CNN-LSTMs helped in detecting the changes in vowel phonation, in so doing identifying the patients with ALS and non-ALS effectively. This was assessed by using different metrics. The accuracy obtained by the proposed CNN-LSTM was 99.99%, and the sensitivity and specificity obtained by the proposed were 99.78% and 99.99%. Therefore, the promising outcome depicted by the proposed model showcases the efficacy of the proposed model for ALS classification. While the proposed model has produced effective results, a notable limitation is that it relies solely on two vowels, /a/ and /i/, as the dataset consists exclusively of these two vowel sounds.

The implications of utilizing vowel phonation, specifically the sounds /a/ and /i/, for detecting ALS are significant for clinical practice. Further, clinically, this approach offers a non-invasive, cost-effective method for early diagnosis and monitoring of bulbar involvement in ALS patients, which is crucial given the disease's progressive nature and the challenges associated with traditional diagnostic methods. By integrating automated acoustic analysis into routine assessments, clinicians can enhance their diagnostic accuracy and tailor intervention strategies based on individual speech characteristics, ultimately improving patient outcomes. However, the proposed model is speaker-dependent; then the dataset is split randomly, deprived of confirming speaker-level separation, and hence the model is limited for speaker-independence. As a result, it will be evaluated in a future study. Additionally, the research will focus on expanding the dataset to include a broader range of phonetic elements and exploring the application of these methodologies in mobile or web-based platforms. This would not only facilitate remote monitoring of speech changes over time but also enhance patient engagement in their care, paving the way for more personalized treatment plans as the disease progresses.

6- Declarations

6-1- Author Contributions

Conceptualization, H.A.D., M.P.R., and M.S.; methodology, H.A.D., M.P.R., and M.S.; validation, H.A.D.; formal analysis, H.A.D. and M.P.R.; visualization, H.A.D., M.P.R., and M.S.; writing—original draft preparation, H.A.D., M.P.R., and M.S.; writing—review and editing, H.A.D., M.P.R., and M.S. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding and Acknowledgements

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/03/272690).

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Zhou, L., & Xu, R. (2024). Invertebrate genetic models of amyotrophic lateral sclerosis. Frontiers in Molecular Neuroscience, 17. doi:10.3389/fnmol.2024.1328578.

[2] Adashek, J. J., Pandya, C., Maragakis, N. J., De, P., Cohen, P. R., Kato, S., & Kurzrock, R. (2024). Neuregulin-1 and ALS19 (ERBB4): at the crossroads of amyotrophic lateral sclerosis and cancer. BMC Medicine, 22(1), 74. doi:10.1186/s12916-024-03293-3.

[3] Moțățăianu, A., Andone, S., Stoian, A., Bălașa, R., Huțanu, A., & Sărmășan, E. (2024). A Potential Role of Interleukin-5 in the Pathogenesis and Progression of Amyotrophic Lateral Sclerosis: A New Molecular Perspective. International Journal of Molecular Sciences, 25(7), 3782. doi:10.3390/ijms25073782.

[4] Mead, R. J., Shan, N., Reiser, H. J., Marshall, F., & Shaw, P. J. (2022). Amyotrophic lateral sclerosis: a neurodegenerative disorder poised for successful therapeutic translation. Nature Reviews Drug Discovery, 22(3), 185–212. doi:10.1038/s41573-022-00612-2.

[5] Colasuonno, F., Price, R., & Moreno, S. (2023). Upper and Lower Motor Neurons and the Skeletal Muscle: Implication for Amyotrophic Lateral Sclerosis (ALS). Roles of Skeletal Muscle in Organ Development. Advances in Anatomy, Embryology and Cell Biology, vol 236. Springer, Cham, Switzerland. doi:10.1007/978-3-031-38215-4_5.

[6] Masrori, P., & Van Damme, P. (2020). Amyotrophic lateral sclerosis: a clinical review. European Journal of Neurology, 27(10), 1918–1929. doi:10.1111/ene.14393.

[7] Xu, R.-S., & Yuan, M. (2021). Considerations on the concept, definition, and diagnosis of amyotrophic lateral sclerosis. Neural Regeneration Research, 16(9), 1723. doi:10.4103/1673-5374.306065.

[8] Štětkářová, I., & Ehler, E. (2021). Diagnostics of Amyotrophic Lateral Sclerosis: Up to Date. Diagnostics, 11(2), 231. doi:10.3390/diagnostics11020231.

[9] Keon, M., Musrie, B., Dinger, M., Brennan, S. E., Santos, J., & Saksena, N. K. (2021). Destination Amyotrophic Lateral Sclerosis. Frontiers in Neurology, 12. doi:10.3389/fneur.2021.596006.

[10] Benatar, M., Wuu, J., McHutchison, C., Postuma, R. B., Boeve, B. F., Petersen, R., Ross, C. A., Rosen, H., Arias, J. J., Fradette, S., McDermott, M. P., Shefner, J., Stanislaw, C., Abrahams, S., Cosentino, S., Andersen, P. M., Finkel, R. S., Granit, V., Grignon, A.-L., … Wuu, J. (2021). Preventing amyotrophic lateral sclerosis: insights from pre-symptomatic neurodegenerative diseases. Brain, 145(1), 27–44. doi:10.1093/brain/awab404.

[11] Biolabs, c. (2020). Rodent Amyotrophic Lateral Sclerosis (ALS) Model. Creative Biolabs, Shirley, United States.

[12] Sindhu, I., & Sainin, M. S. (2024). Automatic Speech and Voice Disorder Detection Using Deep Learning—A Systematic Literature Review. IEEE Access, 12, 49667–49681. doi:10.1109/access.2024.3371713.

[13] Teplansky, K. J., Wisler, A., Green, J. R., Heitzman, D., Austin, S., & Wang, J. (2023). Measuring Articulatory Patterns in Amyotrophic Lateral Sclerosis Using a Data-Driven Articulatory Consonant Distinctiveness Space Approach. Journal of Speech, Language, and Hearing Research, 66(8S), 3076–3088. doi:10.1044/2022_jslhr-22-00320.

[14] Donohue, C. A. (2021). Proactive Dysphagia Management of Patients with Neurodegenerative Diseases: Early Identification and Intervention. Ph.D. Thesis, University of Pittsburgh, Pittsburgh, United States.

[15] Ghasemzadeh, H., Doyle, P. C., & Searl, J. (2022). Image representation of the acoustic signal: An effective tool for modeling spectral and temporal dynamics of connected speech. The Journal of the Acoustical Society of America, 152(1), 580–590. doi:10.1121/10.0012734.

[16] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., … Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. The Innovation, 2(4), 100179. doi:10.1016/j.xinn.2021.100179.

[17] Cebola, R., Folgado, D., Carreiro, A., & Gamboa, H. (2023). Speech-Based Supervised Learning Towards the Diagnosis of Amyotrophic Lateral Sclerosis. Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies. doi:10.5220/0011694700003414.

[18] Simmatis, L. E., Robin, J., Pommée, T., McKinlay, S., Sran, R., Taati, N., Truong, J., Koyani, B., & Yunusova, Y. (2023). Validation of automated pipeline for the assessment of a motor speech disorder in amyotrophic lateral sclerosis (ALS). Digital Health, 9. doi:10.1177/20552076231219102.

[19] Cebola, R. A. S. M. (2022). Towards the Automatic Diagnosis of Amyotrophic Lateral Sclerosis from Speech. Master Thesis, Universidade NOVA de Lisboan, Lisbon, Portugal.

[20] Dash, D., Ferrari, P., Hernandez-Mulero, A. W., Heitzman, D., Austin, S. G., & Wang, J. (2020). Neural Speech Decoding for Amyotrophic Lateral Sclerosis. INTERSPEECH 2020, 25-29 October, 2020, Shanghai, China.

[21] Rowe, H. P., Gutz, S. E., Maffei, M. F., & Green, J. R. (2020). Acoustic-Based Articulatory Phenotypes of Amyotrophic Lateral Sclerosis and Parkinson's Disease: Towards an Interpretable, Hypothesis-Driven Framework of Motor Control. INTERSPEECH 2020, 25-29 October, 2020, Shanghai, China.

[22] Deeb, O., & Nabulsi, M. (2020). Exploring Multiple Sclerosis (MS) and Amyotrophic Lateral Scler osis (ALS) as Neurodegenerative Diseases and their Treatments: A Review Study. Current Topics in Medicinal Chemistry, 20(26), 2391–2403. doi:10.2174/1568026620666200924114827.

[23] Ueha, R., Cotaoco, C., Kondo, K., & Yamasoba, T. (2023). Management and Treatment for Dysphagia in Neurodegenerative Disorders. Journal of Clinical Medicine, 13(1), 156. doi:10.3390/jcm13010156.

[24] Younger, D. S., & Brown Jr, R. H. (2023). Amyotrophic lateral sclerosis. Handbook of Clinical Neurology, 196, 203-229, MedlinePlus Health Information, Maryland, United States. doi:10.1016/B978-0-323-98817-9.00031-4.

[25] Shabber, S. M., Bansal, M., & Radha, K. (2023). A Review and Classification of Amyotrophic Lateral Sclerosis with Speech as a Biomarker. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–7. doi:10.1109/icccnt56998.2023.10308048.

[26] Zhang, J., & Singh, R. (2024). Vocal Fold Dynamics for Automatic Detection of Amyotrophic Lateral Sclerosis from Voice. ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 311–315. doi:10.1109/icassp48485.2024.10448151.

[27] Abdulmajeed, N. Q., Al-Khateeb, B., & Mohammed, M. A. (2023). Voice pathology identification system using a deep learning approach based on unique feature selection sets. Expert Systems, 42(1), 13327. doi:10.1111/exsy.13327.

[28] AL-Dhief, F. T., Latiff, N. M. A., Malik, N. N. N. Abd., Sabri, N., Baki, M. M., Albadr, M. A. A., Abbas, A. F., Hussein, Y. M., & Mohammed, M. A. (2020). Voice Pathology Detection Using Machine Learning Technique. IEEE 5th International Symposium on Telecommunication Technologies (ISTT), 99–104. doi:10.1109/istt50966.2020.9279346.

[29] Milella, G., Sciancalepore, D., Cavallaro, G., Piccirilli, G., Nanni, A. G., Fraddosio, A., D'Errico, E., Paolicelli, D., Fiorella, M. L., & Simone, I. L. (2023). Acoustic Voice Analysis as a Useful Tool to Discriminate Different ALS Phenotypes. Biomedicines, 11(9), 2439. doi:10.3390/biomedicines11092439.

[30] Tena, A., Clarià, F., Solsona, F., & Povedano, M. (2022). Detecting Bulbar Involvement in Patients with Amyotrophic Lateral Sclerosis Based on Phonatory and Time-Frequency Features. Sensors, 22(3), 1137. doi:10.3390/s22031137.

[31] Vengalil, S., Nashi, S., Preethish-Kumar, V., Polavarapu, K., & Nalini, A. (2024). Amyotrophic Lateral Sclerosis. Case-based Approach to Common Neurological Disorders. Springer, Singapore. doi:10.1007/978-981-99-8676-7_18.

[32] Iluț, S., Stan, A., Rahovan, I., Hapca, E., Strilciuc, S., & Muresanu, D. (2023). Variants of Amyotrophic lateral sclerosis and rehabilitation: an overview. Balneo and PRM Research Journal, 14(2), 559. doi:10.12680/balneo.2023.559.

[33] Tena, A., Clarià, F., Solsona, F., & Povedano, M. (2023). Voiceprint and machine learning models for early detection of bulbar dysfunction in ALS. Computer Methods and Programs in Biomedicine, 229, 107309. doi:10.1016/j.cmpb.2022.107309.

[34] Luptáková, I. D., Hanuliaková, J., Žido, L., & Bartoš, P. (2025). M-Learning and Experiential Learning in Vocational Education. Emerging Science Journal, 8, 298–310. doi:10.28991/ESJ-2024-SIED1-017.

[35] Tena, A., Claria, F., Solsona, F., Meister, E., & Povedano, M. (2021). Detection of Bulbar Involvement in Patients with Amyotrophic Lateral Sclerosis by Machine Learning Voice Analysis: Diagnostic Decision Support Development Study. JMIR Medical Informatics, 9(3), e21331. doi:10.2196/21331.

[36] Cave, R., & Bloch, S. (2021). The use of speech recognition technology by people living with amyotrophic lateral sclerosis: a scoping review. Disability and Rehabilitation: Assistive Technology, 18(7), 1043–1055. doi:10.1080/17483107.2021.1974961.

[37] Vashkevich, M., & Rushkevich, Yu. (2021). Classification of ALS patients based on acoustic analysis of sustained vowel phonations. Biomedical Signal Processing and Control, 65, 102350. doi:10.1016/j.bspc.2020.102350.

[38] Likhachov, D., Vashkevich, M., Azarov, E., Malhina, K., & Rushkevich, Y. (2021). A Mobile Application for Detection of Amyotrophic Lateral Sclerosis via Voice Analysis. Speech and Computer: Lecture Notes in Computer Science (SPECOM 2021), Springer, Cham, Switzerland. doi:10.1007/978-3-030-87802-3_34.

[39] Simmatis, L. E. R., Robin, J., Spilka, M. J., & Yunusova, Y. (2024). Detecting bulbar amyotrophic lateral sclerosis (ALS) using automatic acoustic analysis. BioMedical Engineering OnLine, 23(1), 15. doi:10.1186/s12938-023-01174-z.

[40] Kurmi, O. P., Gyanchandani, M., Khare, N., & Pillania, A. (2023). Classification of Amyotrophic Lateral Sclerosis Patients using speech signals. 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), 172–177. doi:10.1109/icsccc58608.2023.10176797.

[41] Vashkevich, M., Gvozdovich, A., &Rushkevich, Y. (2019). Detection of Bulbar Dysfunction in ALS Patients Based on Running Speech Test. Pattern Recognition and Information Processing. PRIP 2019. Communications in Computer and Information Science, 1055, Springer, Cham, Switzerland. doi:10.1007/978-3-030-35430-5_16.

[42] Lv, C., Fan, L., Li, H., Ma, J., Jiang, W., & Ma, X. (2024). Leveraging multimodal deep learning framework and a comprehensive audio-visual dataset to advance Parkinson's detection. Biomedical Signal Processing and Control, 95, 106480. doi:10.1016/j.bspc.2024.106480.

[43] Alqahtani, A., Alsubai, S., Sha, M., Dutta, A. K., & Zhang, Y.-D. (2024). Intellectual assessment of amyotrophic lateral sclerosis using deep resemble forward neural network. Neural Networks, 178, 106478. doi:10.1016/j.neunet.2024.106478.

[44] Mahum, R., El-Sherbeeny, A. M., Alkhaledi, K., & Hassan, H. (2024). Tran-DSR: A hybrid model for dysarthric speech recognition using transformer encoder and ensemble learning. Applied Acoustics, 222, 110019. doi:10.1016/j.apacoust.2024.110019.

[45] Rong, P., Heidrick, L., & Pattee, G. L. (2024). A multimodal approach to automated hierarchical assessment of bulbar involvement in amyotrophic lateral sclerosis. Frontiers in Neurology, 15. doi:10.3389/fneur.2024.1396002.

[46] Mehra, S., Ranga, V., & Agarwal, R. (2024). A deep learning approach to dysarthric utterance classification with BiLSTM-GRU, speech cue filtering, and log mel spectrograms. The Journal of Supercomputing, 80(10), 14520–14547. doi:10.1007/s11227-024-06015-x.