International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 618-635
https://doi.org/10.70917/ijcisim-2026-2102
Article

IMPLEMENTATION OF ROBUST SPEECH RECOGNITION TECHNIQUES FOR IMPAIRED SPEECH PATTERNS

Suryakant B. Kamble[1], Santosh C. Wagaj[2], Anil S. Shirsat[3]

[1] Department of Electronics & Telecommunication Engineering, JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune, India.suryakantbk@gmail.com
[2] Department of Electronics & Telecommunication Engineering, JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune, India.scwagaj@gmail.com
[3] Department of Electronics & Telecommunication Engineering, PES's Modern College of Engineering, Pune, India. asshirsat@gmail.com
Corresponding Author: Suryakant B. Kamble (suryakantbk@gmail.com)

Abstract: These Recent advances in speech recognition have made possible precise transcription and command recognition in a variety of application areas. Nevertheless, the current systems have a major problem in identifying speech of afflicted people, including the patients with dysarthria, post-stroke, or growth speech disorders. To overcome this deficiency, the current paper presents a new framework that will help enhance the performance of speech recognition among impaired speakers. The system first divides the speech into male and female to be able to use the specifics of the voice and the differences in acoustics. After that, a large data set of impaired and healthy samples of speech is prepared and processed. Noise reduction and acoustic data extraction, such as Mel-Frequency Cepstral Coefficients (MFCCs) and Mel-spectrograms, are part of the preprocessing phase. The core design concept in the effective use of CNNs and LSTM networks is their combination to achieve spatial and temporal speech patterns recording. Once the first transcription has been made, the identified text (impaired) is fixed with the help of an n-gram language model to strengthen its contextual interpretation and linguistic coherence. To verify the strength of the proposed model, the Word Error Rate (WER) is used to evaluate it and compare it with deep learning architectures. As per the experimental results, the proposed methodology is much better than the existing speech recognition methods in terms of accuracy and ability to withstand damaged speech. Therefore, the system improves human-computer interaction among affected people and leads to the creation of more inclusive and adaptive speech recognition systems..

Keywords: Speech Recognition, Impaired Speech, Deep Learning Models, Adaptive Algorithms, Customized Acoustic Features, Features Extraction

1. INTRODUCTION

The technology of speech recognition has developed significantly, and it has been widely used in the last few decades. Solutions that have been created in areas like automated customer support and voice-activated virtual assistants have fundamentally changed the way individuals engage with technology and how the society uses technology. But mostly, these systems share the fact that they were typically intended to identify speech input which is reasonably distortion free, continuous and defect-free. This premise is a big stumbling block to anyone having speech problems when discussing people with speech problems. The victims of stroke, individuals with cerebral palsy or other neurological disorders, and individuals with dysarthria are just some examples of conditions [1]. Normal speech recognition algorithms that do not handle impaired speech do not align well in handling abnormal articulation, abnormal prosody and changing fluency, which are typical of impaired speech.

As a gap exists in the state-of-the-art speech recognition systems, this study aims to fill it by developing techniques for accurate speech problem detection and interpretation. This is vital in order to ensure that people with speech difficulties have better access and communication and to make current voice-controlled products more inclusive. Traditional models of voice recognition, trained on massive amounts of high-quality audio recorded under controlled environments, struggle to handle deviant representations, or speech that does not adhere to a specific standard. Consequently, there is an immediate need to develop tailored algorithms and models that can acquire knowledge about the nuances of speech impairments.

This paper concentrates on the creation of speech classification methods, which involve ways of classifying speech of people with speech problems. The aim of this work is to increase the accuracy and consistency of the speech recognition algorithms in identifying degraded speech patterns. It is accomplished by providing feature extraction, model training and classification. The authors consider the challenges of speech impairments in the research, and they offer research and methodologies that are applicable. The work is a machine learning algorithm that may help quickly recognize the impaired speech patterns, and advanced ways of feature extraction that characterize the unique acoustics of impaired speech. It also analyses the use of MFCC features which are an accurate measure of speech signals that are damaged. The paper also examines the DL models, LSTM and CNN, which are effective in recognizing complex and non-linear categorical speech data, and thus helping identify speech disorders accurately.

The primary contribution of our work implies:

Conducting an extensive literature survey to determine the best feature extraction methods for deciphering damaged speech, as well as the classification algorithms used by researchers in the area of voice recognition.

The use of a deep learning classification model to differentiate male and female voices increases the ability of the model to support the unique vocal traits and improves the performance.

Utilizing the language model to enhance contextual understanding, hence improving speech pattern recognition for those with disabilities.

The remaining paper is organized as.; Section 2 – literature reviews related work Section 3 - presents a detailed review of previous studies. Section 4 - describes the proposed method. Section 5 - presents the results and their analysis using different performance metrics. Section 6 - concludes the paper.

2. RELATED WORK

Feature Extraction Techniques for Speech Recognition with Impaired Speech [1] [2]

Mel-Frequency Cepstral Coefficients (MFCC)

MFCC is among the commonly used feature extraction methods in the field of voice recognition. In order to extract the power spectrum of a voice stream, MFCC transforms it into the Mel scale, which is suitable to the human auditory perception. The initial phase involves partitioning the audio stream into smaller frames, subsequently acquiring the power spectrum by the Fourier Transform. The frequencies are then converted to the Mel scale using triangular filter banks. Thereafter, the logarithm of the Mel spectrum is calculated, succeeded by the use of the DCT to decorrelate the coefficients. The resultant coefficients are referred to as MFCCs, which fundamentally denote the spectral envelope of the speech signal. This results in a feature set that is both compact and informative. Since the method is highly effective in reproducing the timbral texture of the voice, it is a good solution when it comes to recognition of speech or some distorted speeches for that matter. It is possible to increase the precision of speech recognition in various situations, since MFCC's are insensitive to variations in speech signal characteristics, either frequency, pitch, and volume, when the voice is distorted.

Linear Predictive Coding (LPC):

One effective way to compress the spectral envelope of a voice signal is by linear predictive coding, or LPC. Estimating vocal tract resonances and modelling the speech signal are the two pillars upon which the LPC method rests. The principal role of the LPC is to minimize prediction error by calibrating the coefficients to achieve optimal fitting; hence, the current sample's prediction is derived from preceding samples with maximal accuracy. These coefficients allow for an accurate description of the formant structure of the speech signal, which importantly supports the determination of the phonetic content. Due to its characteristics, LPC is especially advantageous in speech recognition for severely damaged audio, since it can encapsulate the essential elements of the speech waveform even when the signal is erratic or distorted, making it crucial for effective speech recognition. LPC parameters are useful

in cases of damaged speech since they are relatively tolerant of the background noise that may cause difficulties to record proper speech. The coefficients produced by the LPC acting as features, used in the process of generating a spectral envelope, can be further used in other classification algorithms, thus providing a reliable representation of RMS signal's properties.

Discrete Wavelet Transform (DWT)

In contrast, the DWT is a method for feature extraction from voice signals that involves band separation. Therefore, the possibility of analysing the temporal information and spectral information is achieved simultaneously. The DWT facilitates multi-resolution analysis by employing wavelets that may capture both high-frequency and low-frequency components of the voice signal in this instance. This differs from the Fourier Transform, which solely conveys information about frequency. Thus, it becomes easier to do voice recognition since features that accommodate short-term changes in the voice signal can be extracted. Short-term changes can occur due to speaking changes or disfluency, among other factors. In the DWT, the voice message is processed via both a high-pass filter and a low-pass filter. Then, wavelet coefficients of the voice signal at different scales are generated. Therefore, this information is collected as the features for further analysis or classification. DWT provides a good way of analysing non-stationary signals such as voices – their features change with time. Therefore, it is most appropriate in recognizing enhanced voices which are associated with speech characteristics.

Mel-Spectrogram

A Mel-Spectrogram illustrates the frequency spectrum of a vocal signal represented on the Mel scale. The frequency spectrum varies throughout time, as illustrated in this picture. As a perceptual frequency scale, the Mel-scale shows that various human-heard pitches are equally distant from one another. This lines up with how people hear the pitch. The speech stream is initially separated into overlapping frames before a Mel-Spectrogram is generated. Subsequently, we employ the Short-Time Fourier Transform on each frame to extract its frequency components. Subsequently, the Mel scale is utilized to map the resultant power spectra using triangular filter banks. Besides providing a comprehensive feature set that is invaluable to deep learning models, the Mel-Spectrogram can describe the length and breadth of the speech signal across space and time. Standard features will not generate good results because they may not show the wide variability associated with impaired speech, including differences in articulation, pitch, and timing. It is a super effective method of capturing these variations from impaired speech. The Mel-Spectrogram is a firm that helps identify patterns associated with a variety of speech disorders. It accomplishes this by providing a detailed view of the speech signal over time.

Pitch Contour Analysis

Pitch contour analysis is the process of extracting pitch, also known as the fundamental frequency, from a speech signal over a certain period of time. Speech has many crucial features, one of which can be pitch. Pitch ranging has the potential to encode a wide range of infrequent devices, including the speaker's intonation, emotion, and stress. Perhaps because impairments frequently affect the reliant or consistent nature of pitch differences, the analysis of pitch contours may be critical in the situation of flawed and/or pathological speech production or for speech impairments. To describe the variations in pitch through the spoken signal, i.e., time series, pitch changes that have been calculated, a pitch contour has thus been produced. When specific acoustic features may be problematic when affected by impairments, this matrix may be calculated and incorporated as a speech recognition system to describe characteristics between various speech phones. This is beneficial in identifying utterances and terms suitable for persons with hearing loss who have experienced these impairments. Since it could be beneficial in describing tonal inconsistencies and variations subtleties of human speech, Pitch Contour Analysis is a necessary method in understanding and interpreting the speech of speech-impaired persons.

Spectral Subtraction

Spectral Subtraction is a prevalent noise reduction method, particularly effective as a pre-processing step for speech recognition systems when voice quality is compromised by background noise. It is especially effective in low signal-to-noise ratio conditions due to its simplicity and model-independent nature. Comparative studies in speech enhancement identify spectral subtraction as a classical baseline method alongside Wiener filtering; however, Wiener-based approaches typically require more complex noise estimation and adaptive filtering mechanisms, whereas spectral subtraction offers a computationally efficient and easily implementable solution for ASR front-end processing [3]. The essence of the method consists of the following steps: one has to first estimate the noise spectrum based on those parts of the signal that are silent or do not contain speech, then subtract this estimated noise spectrum from the complete signal's spectrum. As a result, the signal that is obtained as serial output contains less noise, which makes it

easier to capture the features that are necessary for speech-based models to do recognition. Spectral Subtraction should be preferred when conducting experiments on the recognition of degraded speech since the co-occurrence of background noise can make the recognition challenges even harder to overcome. Since the process cleans the signal before feature extraction, the features that are retrieved will be of significantly better quality, and this will enhance the performance of the models on identifying degraded speech. The technique's original intent was to mitigate the effect of background noise on speech, but it has since found a new home in making speech-based models more resilient in noisy environments.

Delta and Delta-Delta Coefficients

Thereafter, the Mel scale is employed to convert the resulting power spectra through triangular filter banks. Delta coefficients are used to characterize the features if the feature values are altering; Delta-Delta coefficients are used to reflect the acceleration or the rate of change. Dynamic characteristics in speech, such as modifications in forms and the linear interpolation of pitch, are crucial for the recognition of spoken words. These attributes of speech are thus helpful in the sense that they facilitate the capture of such dynamic changes. They are also useful when it comes to identifying the differences when a person has trouble speaking, which is common when they fail to articulate appropriately for a long period

Table 1. A Comparative Analysis of Audio Feature Extraction

Technique	Application	Benefits	Results
MFCC	General speech recognition, including impaired speech	Captures the spectral envelope, robust to pitch and loudness variations	Widely used in speech recognition systems with good accuracy; effective in capturing timbral textures of speech, even for impaired speech
DWT	Examination of non-stationary signals, including speech qualities that are transitory	Recordings spectral and temporal data; deals with non-stationarity	Effective in recognizing impaired speech by capturing transient and irregular signal features, leading to better classification of speech variations
Mel-Spectrogram	Deep learning-based speech recognition, pattern recognition in speech signals	Provides a detailed time-frequency representation; aligned with human auditory perception	Enhances recognition accuracy in speech with irregularities; especially beneficial for deep learning models handling impaired speech
Pitch Contour Analysis	Intonation, emotion detection, and speech impairment recognition	Captures pitch variation over time; identifies tonal and rhythmic irregularities	Useful for recognizing speech affected by pitch irregularities, improving system sensitivity to prosodic features of impaired speech
Gammatone Frequency Cepstral Coefficients (GFCC)	Robust speech recognition, noise-robust feature extraction	Models auditory system more accurately; captures critical auditory cues	Improves recognition in noisy environments or with impaired speech, offering robust features less affected by distortion
Spectral Subtraction	Noise reduction in speech recognition, preprocessing for clearer signal extraction	Reduces background noise effectively; enhances feature clarity	Improves the quality of extracted features, leading to better recognition accuracy, particularly in noisy environments where impaired speech is present
Delta and Delta-Delta Coefficients	Capturing dynamic changes in speech, temporal features in speech signals	Captures temporal variations; enhances feature set with dynamic information	Enhances recognition accuracy by capturing speech dynamics, useful in cases of impaired speech with irregular articulation and timing

3. LITERATURE REVIEW

Prashant G. Patil et al. [4] (2022) cantered their work on suggesting a novel approach to improving the aural perception of different types of background noise for those who wear hearing aids. The suggested approach uses the DCT to link a neuro-fuzzy classifier with an I-AMS algorithm, which improves the efficacy of voice augmentation (Figure 1). I-AMS algorithm breaks down noisy speech signals into time-frequency units to demodulate noise. Significant parameters such as modulation frequency, centre frequency (fc), and time-frequency units of these parameters (t-f units) are reclaimed from the de-noised signal. The speech environment in the background is then segmented into three parts by applying a neuro-fuzzy classifier to increase the systems adaptability to various acoustic environments, as shown in Figure 2. The result of the excellent function of the proposed I-AMS algorithm is verified by conducting an experiment, achieving a substantial improvement in speech recognition performance, as shown in Table 2. The suggested method improved recognition rates by 11.80% compared to the performance of the other techniques, sensitivity growth by 1.02%, and the effect of de-noising on overall recognition by 1.27%. According to these results, the I-AMS algorithm holds the promise of significantly increasing the speech recognition precision and clarity users with a hearing aid.

Abdusalomov A.B et al. [5] (2022) introduced an ML-based solution which pays attention to the voice signal pre-processing component. It is stressed out in this study that processing speed matters a lot for real-time systems; for speech recognition, fast-processing computing is vital. The authors' solution suggested in this idea is to select an optimal cache block size to significantly reduce the processing time that is needed for real-time implementation by block-mapping main memory blocks to cache because the optimal selection can address the difficulties of high computation requirements for speech signal processing and the effective suppression of overclocking problems in digital signal processing. The results of the experiment performed illustrate just how much better the proposed solution to a voice recognition problem performs over simple algorithms used in practice nowadays in a stable classification performance without hopping and features the extraction problem is faced. This paper is important because it demonstrates the need for advanced ML technology performance combined with efficient computation approaches to improve real-time voice recognition in complex environments such as smart cities.

In their article, B. A. Al-Qatab et al. [6] (2021) investigated the potential improvement of dysarthric speech categorization via measurement of the degree of speech impairment by analyzing different acoustic signals and feature selection techniques. There were four big sets of acoustic characteristics, and they were voice quality, prosodic, spectral, and cepstral. The methods of feature selection were as follows: The methods applied in the critical approaches. These significant decisions were used to improve the classification outcome and narrow down the number of elements that were used in the job. Such techniques were particularly necessary as they rendered it possible to identify cues that are of the utmost importance in order to perform successful classification. The ML classifiers were those that were used to assess the results of the selected characteristics. The performance differs dramatically between the features selected and classification: it is ranging between 40.41 percent and 95.80 percent. It is obvious that the discovery of the best features and, finally, algorithms have a substantial impact on the performance and the targeted objective in terms of speech classification.

Labied Maria et al. [7] (2021) presented a detailed comparison of the many methods of extracting features used in Automatic Speech Recognition systems. They also focus on compiling a list of matrices based on the standards of comparison of methods, including accuracy, computational efficiency, noise immunity, and language and dialect variability. The comparison includes in-depth feature analyses, such as Wavelet Transform and Linear Predictive Coding, as well as the most popular ones— MFCC. Each method is discussed outlining its strengths and weaknesses, as well as its suitability for a range of ASR applications. For instance, while MFCC has been praised for being able to capture the vital features of voice, at the same time, the method is sensitive to noise. While LPC is generally praised for its efficiency and simplicity, the method falls short when it comes to accurately capturing all subtleties of the speech, especially when the emission is high. Wavelet Transform is recommended for use in various acoustical environments, as the method can analyse non-stationary signals such as speech differently. To establish maximum ASR performance in any setting, the authors recommend a hybrid approach employing several feature extraction methods, as per their final recommendation. This paper can be of great use to students and developers in their quest to maximize the performance of ASR based on the needs of a particular application.

M. K. Reddy et al. [8] (2020) introduce an innovative approach for detecting children with Specific Language Impairment by analyzing speech waveforms, focusing particularly on glottal source attributes. This method employs glottal inverse filtering on the voice source signal to extract glottal features in both time and frequency domains.

Furthermore, glottal characteristics and acoustic features, assessed by MFCC and openSMILE, are extracted from spoken utterances for study. Two ML algorithms are trained separately using openSMILE, MFCC, and glottal data, and their performance is evaluated. Additionally, the purpose of their study is to examine how MFCC and openSMILE acoustic properties complement one another. The results show that glottal features can improve the accuracy of SLI identification, and that when utilizing the FFNN approach in an independent speaker, glottal features with the MFCC feature voter work well.

Vidyashree Kanabur et al. [9] (2019) ensured a comprehensive review of the methods used for Automatic Speech Recognition today, focusing on feature extractions. They analyse numerous methods for ascent, featuring that it is essential to focus on it for achieving accuracy about recording the sound's acoustic features. Since the novel extraction methods of MFCC, Linear Predictive Coding, and Perceptual Linear Prediction exceeded prior models' capabilities, the review sufficiently covered the existing difficulties with the extent methods. These problems include speaker variability, growing sensitivity, and the complexity of real-time processing. The authors also considered features that have recently appeared in the field of ASR, such as adjusting conventional methods of extraction with deep learning models for enhancing artifacts' stabilization and recognition accuracy. Such a perspective is beneficial for practitioners and researchers alike, as it gives an impression of future research possibilities in ASR and helps a better understanding of the pros and cons of different methods of feature extraction.

Raviraj Vishwambhar Darekar and colleagues [10] (2018). The author introduced an innovative adaptive learning architecture for artificial neural networks to enhance the accuracy of emotion recognition from speech data. The methodology relies on the integration of voice data from multiple sources. Thus, it is designated as a hybrid PSO-FF algorithm, which integrates the advantages of both Feed-Forward and Particle Swarm Optimization techniques to enhance the network's training process. The experimental findings, obtained using the supplied models, reveal a 10.85% enhancement in accuracy for both the Marathi speech database and the benchmark database, illustrating that the developed model significantly surpasses conventional methods. The obtained result confirms well the work of a hybrid PSO-FF algorithm and adaptive learning architecture in order to improve accuracy achieve speech signal emotion. Prajakta Rokade et al. [11] (2018) endeavoured to develop a system which assists people with speech and hearing impairments to communicate effectively with others. Although there are numerous sign languages spoken across various nations, this system is specifically designed for Indian Sign Language. This language is still in the process of standardization. Since only hand gestures are recognized in this paper, which can allow transmission of ideas, message, and thoughts between deaf and mute. Therefore, in this method International Sign Language gestures and phrases words are converting into the Marathi text and pictures are written on which vice-versa video are taken. In this method we perform three phases, it's described as first phases, feature extraction: recognition of hand moves from ISL. Second phases, classification is performing the classification between them and convert into characters. Third phases, pre-processing: how to convert ISL into written text uses a simple method that helps those hard of hearing and deaf can easily interact with regular people

Alim Sabur et al. [12] (2018) presented a good overview of the main algorithms used in voice feature extraction, which is a problem at the key point in a speech recognition system. They discussed several algorithms that are widely used, including Perceptual Linear Prediction, MFCC, and Linear Predictive Coding. The latter are pretty versatile in capturing the unique properties of speech signals since MFCC is capable of simulating the functioning of the human ear when trying to understand how sound is perceived. LPC creates a solid representation of the spectral envelope of speech, while PLP is designed to take into account the perceptual properties of human hearing. Algorithms' pros and cons are discussed in the context of how effectively they work with different types of voice recognition tasks. Based on that, the authors conclude regarding the relevance of the right choice of feature extraction methods which should be made depending on a specific goal of one's application. This work can be fairly useful to researchers and practitioners since which provides a strong knowledge base necessary for understanding and applying efficient speech feature extraction practices. Kishori R. Ghule et al. [13] (2015) focused on developing a database for speech recognition in the independent Marathi words. A list of 100 Marathi words was chosen to create the database, while the voice signals for all words in the database were sampled directly from a microphone. ASR, automatic speech recognition, was executed with the data of 100 speakers – each speaker produced three predetermined words each. The word was subsequently recognized using ASR after characteristics were extracted from the voice signals using the DWT. The rationale is that DWT facilitates multi-resolution and multi-scale analyses, which are fundamental for the concept's basis in the examination of non-stationary signals such as speech. The Artificial Neural Network was employed to categorize the feature vectors produced by DWT, addressing the multiclass classification challenge. In this experiment, the ANN is trained on a feature vector produced from the recognized speech pattern in the classification phase implementation, and its efficiency is tested with an independent dataset. This method utilizes the

feature extraction capabilities of DWT and the classification proficiency of ANN to improve the efficiency and accuracy of speech recognition for isolated Marathi words.

Magre, Smita et al. [14] (2013) presented a substantial overall picture of the strategies developed at each speech recognition phase. At the same time, the author gave valuable advice on the possibility and benefit of choosing one or another strategy based on the advantages and disadvantages of each methodology. A comparative analysis is conducted to examine the various tactics employed at different levels of speech recognition. Consequently, the merits and benefits of the method in question are evaluated. Finally, the findings led to recommendations. Among recommendations, the development of a research agenda is particularly noted, iterating future exploration avenues in creating human-computer interface systems using the Marathi language. Therefore, this paper's objective is to compare tactics in order to discover the best ones for improving the system capability of voice recognition technology. This approach leverages the feature extraction skills of DWT and the classification expertise of ANN to enhance the efficiency and accuracy of audio recognition for isolated Marathi words.

Gaikwad Santosh et al. [15] (2011) examined the efficiency of several feature extraction techniques concerning speech recognition. The paper aims to introduce two widely spread appearance-based methods of feature extraction, which are MFCC and Linear Discriminant Analysis. While MFCC's main role is the extraction of features, LDA is used to reduce extracted features' dimensionality. It eases the computational efforts and may result in the enhancement of classification's accuracy. One can present a novel approach that combines MFCC and LDA in a fusion model for feature extraction, and the experiments with MFCC as an independent feature extraction approach. The comparative analysis explains how voice recognition systems may benefit from both MFCC's and LDA's strengths in the extraction and dimensionality reduction areas. Sonal Yadav et al. [16] (2023) focused on feature extraction and classification techniques, doing a comprehensive examination of the most prevalent contemporary approaches for voice recognition. Along with a brief description of feature extraction methods, the authors also note the importance of timely formulating the inherent features of human speech. These are mainly MFCC, Linear Predictive Coding, and Gammatone frequency cepstral coefficients. In addition, in this study, well-known classification methods CNN, SVM, as well as ANN have been tested for high or poor performance in various voice recognition tasks. The authors presented a comparison and pointed to the pros and cons of each method, as well as described in which conditions they function best and weakest. Thus, this study is a necessary source for both researchers and practitioners to choose appropriate methods for enhancing performance and accuracy in speech recognition.

X. Hu et al. [17] (2022) The author presented a cutting-edge wearable target-location technology named StereoPilot, made for assisting visually impaired humans in their spatial cognition. The system's head-mounted RGB-D camera gathers and analyzes 3D spatial data of the environment to provide BVI with navigational cues. These cues were sent using spatial audio rendering to use BVI spatial sense of sound localization to identify where the sound is coming and their direction. Students with BVI and sighted human volunteers were involved in the studies, which included studies comparing SAR with multiple other more haptic and aural presentation modalities. The findings of the Fitts' law test shown that SAR significantly enhances spatial navigation, decreasing positional inaccuracy by 40% and increasing the information transmission rate threefold compared to verbal input. It was also found that the learning curve is preferable to other signification techniques such as vOICe. StereoPilot reduced the time until target grasping and enabled a precise localization of the object in desktop manipulation tasks compared to voice-based directions on the desktop. Overall, StereoPilot is a fast and efficient communication channel to inform BVI people about their surroundings, which helps them navigate and perceive space in real-world scenarios. Labied and Belangour [21] provide a multi-criteria review of the traditional ASR feature extraction techniques that reveal that features such as MFCC, PLP, GFCC and wavelet-based variations have differing merits in presence of noise features and application scenarios. Jalil et al. [22] introduce a hybrid feature extraction method that integrates wavelet based and cepstral features and show a better performance on text-independent speaker recognition in the presence of noise

Table 2. Comparative Analysis of Literature Review

Reference	Methodology Used	ML / DL Algorithms Used	Dataset	Results Achieved
Prashant G. Patil et al. (2022)	Noise-robust speech enhancement for hearing aids	Neuro-Fuzzy Inference System, DCT-based I-AMS Algorithm	Noisy speech signals (real acoustic environments)	Accuracy ↑11.80%, Sensitivity ↑1.02%, Overall gain ↑1.27%

Abdusalomov A. B. et al. (2022)	Fast speech preprocessing for real-time systems	ML-based Cache Optimization, Feature Mapping Models	Real-time speech data	Reduced processing time with stable classification
B. A. Al-Qatab et al. (2021)	Dysarthric speech classification using acoustic feature selection	SVM, k-NN, ANN, Decision Tree	Dysarthric speech dataset	Accuracy ranged from 40.41% to 95.80%
Labied Maria et al. (2021)	Comparative study of ASR feature extraction methods	MFCC-based ASR models, LPC, Wavelet-based classifiers	Standard ASR benchmark datasets	Hybrid feature extraction improves robustness
M. K. Reddy et al. (2020)	SLI detection using glottal and acoustic features	SVM, FFNN	LANNA SLI speech corpus	Improved SLI detection using glottal + MFCC features
Vidyashree Kanabur et al. (2019)	Review of ASR feature extraction techniques	MFCC, LPC, PLP with Deep Learning hybrids	Review-based (multiple datasets)	Identified challenges & future research trends
Raviraj V. Darekar et al. (2018)	Emotion recognition using adaptive learning	Hybrid PSO-FFNN (Particle Swarm Optimization + FFNN)	Marathi & benchmark emotion datasets	Accuracy improvement of 10.85%
Prajakta Rokade et al. (2018)	Assistive communication via Indian Sign Language	Gesture Recognition, ANN-based Classification	ISL gesture dataset	Enabled communication for deaf and mute users
Alim Sabur et al. (2018)	Survey of speech feature extraction algorithms	MFCC-based ML models, LPC, PLP	Review-based	Emphasized task-specific feature choice
Kishori R. Ghule et al. (2015)	Marathi word recognition system	DWT + ANN	Marathi speech database (100 speakers)	Improved accuracy for isolated word recognition
Magre, Smita et al. (2013)	Comparative ASR strategy analysis	HMM-based ASR, ANN-based ASR	Marathi speech systems	Recommended optimal ASR strategies
Gaikwad Santosh et al. (2011)	Feature fusion for speech recognition	MFCC + Linear Discriminant Analysis (LDA) + Classifiers	Speech datasets	Improved accuracy with reduced dimensionality
Sonal Yadav et al. (2023)	Feature extraction & classifier evaluation	CNN, SVM, ANN with MFCC, LPC, GFCC	Standard speech datasets	Identified best methods per condition
X. Hu et al. (2022)	Wearable assistive navigation using spatial audio and RGB-D sensing	3D Spatial Mapping, Spatial Audio Rendering (SAR), Computer Vision Pipelines	Custom dataset (BVI students & sighted users)	40% reduction in positional error; 3× higher information transfer rate

The comparative analysis of literature at hand shows that in challenging speech recognition and impairment-related tasks, deep learning models often outperform classic machine learning approaches. Classical machine learning algorithms such as SVM [24], k-NN, Decision Trees, and ANN show promising results when carefully designed features (e.g., MFCC, LPC, DWT) are used; however, their performance is rather unpredictable in terms of accuracy variability across datasets and speech environments. Independent neural network models, e.g. FFNN, ANN, CNN were found to perform better with complementary acoustic features. Based on the reviewed literature, the effectiveness of speech recognition and classification systems largely depends on the choice of feature extraction algorithms,

classification algorithms, and the complexity of the speech conditions, particularly when dealing with impaired, noisy, or fluctuating speech signals. Out of all feature extraction methods, MFCCs become the most popular and universal method of feature extraction in all studies because they can reflect characteristics of speech that are perceptually relevant.

4. METHODOLOGY

A developed methodology can be used to create a speech recognition system that can withstand speech impairment and it is composed of the following stages: Data Collection, Pre-processing, Feature Extraction, Classification Model Design, Acoustic Modelling, Impaired Speech Recognition and Model Evaluation. In the Data Collection stage, corpora of different speech impairments were collected through publicly available corpora. Each audio file was carefully annotated with its transcription to enable supervised learning

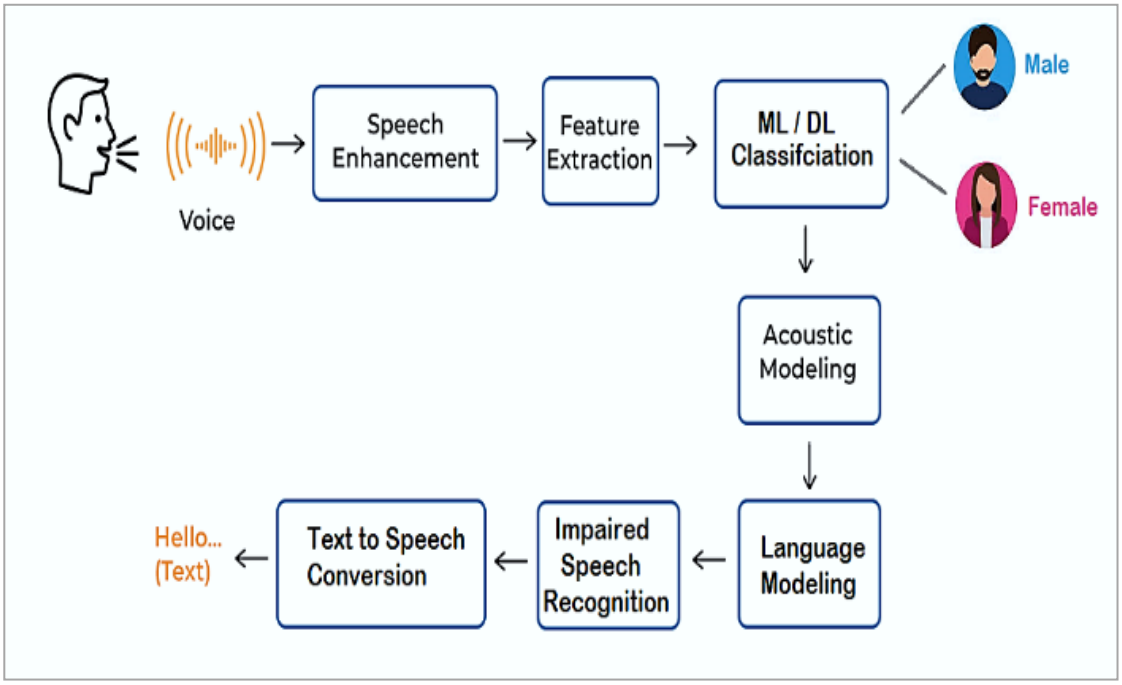


Fig. 1 System architecture of impaired speech recognition

Pre-processing involved normalization of features extracted, and a number of noise reduction techniques, including Spectral Subtraction. In Feature Extraction, MFCCs, Mel- Spectrograms, DWT, and GFCC were used to convert raw audio into meaningful features. Designing Classification Model involved the extraction of spatial features of spectrograms via CNNs. Acoustic Modelling and Impaired Speech Recognition consisted of using LSTM networks for learning temporal dependencies in speech. Another model that was incorporated to enhance recognition by the use of context of the recognized words sequence was a language model that incorporated the n-gram model. Model Evaluation was done by relying on performance Words Error Rate, Character Error rate and accuracy/loss curve to evaluate the recognition performance. This kind of methodology can be applied to create a speech recognition system that is specific to impaired speech patterns. The system overview is depicted in figure 1. The description of the methodology is detailed below with several important steps of data collection to deployment of the models;

Data Collection

Dataset Acquisition: The Hindi Speech Classification dataset that was downloaded through Kaggle [23] is comprised of about 4,000 Hindi speech audio files, half of which were women and the rest were men. The data is categorized into two major groups (male and female), which makes it apt in gender-based speech classification problems. All the recordings are in the standard audio formats and include natural spoken Hindi with different pitches, tones and speaking style.

Data Annotation: Ensure that each audio file is accurately labelled with the corresponding transcription to facilitate supervised learning

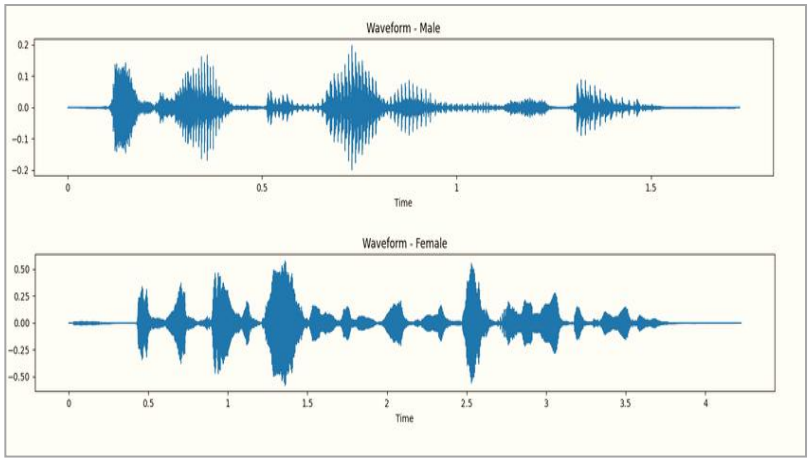


Fig. 2 Input Audio Signal Files

Pre-processing

Normalization

Normalization refers to normalizing the input feature values' ranges in such a way that all features' values influence the training proportionally. The values of the extracted features may take different scales in speech recognition systems such as Mel-Spectrograms, MFCC, DWT coefficients, and Gammatone Frequency Cepstral Coefficients. The phenomenon may influence the pattern-learning effectiveness in the modelling training process. In the absence of normalization, the ones with a larger range could affect the learning process, which could result in inefficient gradient descent optimization and relatively noisy training behaviour. The input audio signals were normalized with the help of the *librosa.util.normalize* function that introduces L2 norm (Euclidean) normalization that equalizes the total signal energy and minimizes the variation in amplitude across samples.

$$xnorm = \frac{x}{||x||}$$

Noise Reduction

Employ noise filtering methods, such as Spectral Subtraction, to enhance the intelligibility of the speech input.

Feature Extraction

Convert raw audio into meaningful features. Feature extraction techniques include:

Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs are computed through a series of steps:

Pre-emphasis: In order to improve feature detection in the higher frequency ranges, the input audio signal is filtered.

Framing: The continuous audio stream is segmented into brief overlapping frames, usually lasting 20-40 milliseconds, to capture the transitory features of speech.

Windowing: Each frame is processed using a window function, like the Hamming window, to reduce the impact of signal discontinuities at the frame boundaries.

Fast Fourier Transform (FFT): To create a signal spectrum, it is necessary to transform the windowed signal from time domain to frequency domain.

Mel Filter Bank: A series of triangular filters, arranged according to the Mel scale, are employed to process the power spectrum in a manner analogous to human auditory frequency perception.

Logarithmic Compression: The logarithmic scale is applied to the Mel-filtered signal so that it more faithfully represents the audible volume to the human ear.

Discrete Cosine Transform (DCT): A collection of compact, uncorrelated coefficients called MFCCs is produced by applying the DCT to decorrelate the log Mel spectrum.

Using these procedures, raw audio can be compressed into a form that speech recognition tasks can understand. Following Figure 3 shows output as:

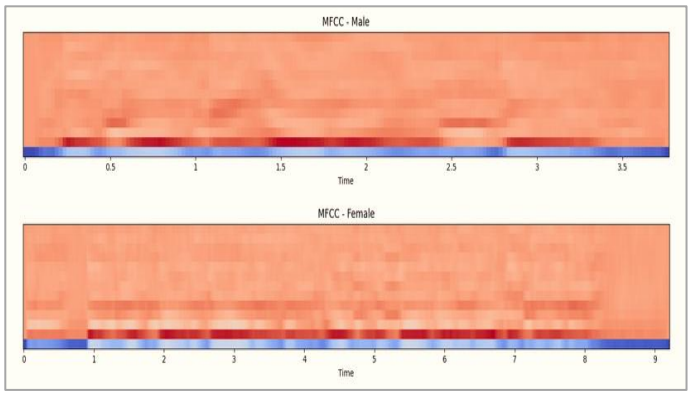


Fig. 3 MFCC Feature Extraction

Mel-Spectrograms

The power spectrum of an audio signal represented on a Mel scale is termed a Mel-spectrogram.Figure 4 displays the outcome of the retrieved mel-spectrogram.

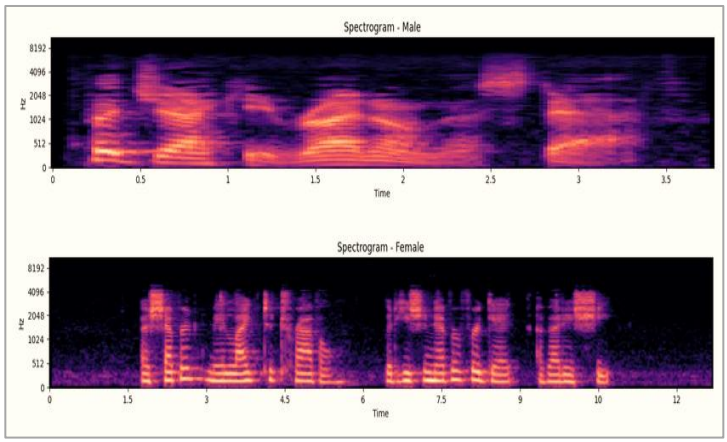


Fig. 4 Mel-Spectrograms Feature Extraction

Classification Model

Convolutional Neural Networks (CNNs)

To extract spatial features from spectrograms CNN model is used. Figure 5 shows the 1D CNN architecture.

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 98, 64)	256
max_pooling1d (MaxPooling1D)	(None, 49, 64)	0
conv1d_1 (Conv1D)	(None, 47, 128)	24,704
max_pooling1d_1 (MaxPooling1D)	(None, 23, 128)	0

flatten (Flatten)	(None, 2944)	0
dense (Dense)	(None, 128)	376,960
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Fig. 5 CNN Model Architecture

This 1D CNN [25] [26] model for speech recognition processes 1D audio feature inputs (e.g., MFCCs) through a series of layers. Firstly, the input data is processed by two Conv1D layers using convolution operations to extract local patterns. Then, the data is reduced in dimensionality by MaxPooling1D layers, which preserve key features while minimizing computational expense. Finally, the process ends. Subsequent to transmitting the 1D vector, derived from the Flatten layer's conversion of the 2D feature maps, through a fully linked Dense layer including 128 neurons, the data undergoes additional processing. A Dropout layer is utilized to randomly eliminate neurons from the training set to prevent overfitting. The final Dense layer categorizes input into two classes with a softmax activation function.

Acoustic Model and Impaired Speech recognition

Long Short-Term Memory (LSTM)

LONG short-term memory (LSTM) networks programmed to record speech's temporal relationships. You can see the LSTM model's main points in Figure 6.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, None, 128)	67,072
dropout (Dropout)	(None, None, 128)	0
lstm_1 (LSTM)	(None, 64)	49,408
dropout_1 (Dropout)	(None, 64)	0
dense (Dense)	(None, 1)	65

Fig. 6 LSTM Model Architecture

Input Layer: It is necessary to provide the initial LSTM layer with input in the format (timesteps, num_features), where timesteps is the number of time steps in the sequence and num_features is the number of features at each time step (e.g., MFCCs, Mel-spectrogram).

LSTM Layers: The first 128-unit LSTM layer is set up to return sequences, so it will provide different outputs at each time step. Since it does not return sequences, the second 64-unit LSTM layer only provides the final output.

Dropout Layers: Between LSTM layers, dropout is employed to avoid overfitting by training with a randomly selected fraction of input units set to 0.

Output Layer: The output probabilities for binary classification are generated by a 1-unit Dense layer that uses a sigmoid activation function.

Table 1. Hyper parameter Setup Table

Hyper-parameter	Value
Hidden Layer 1 Units	128
Hidden Layer 2 Units	64
Activation Function	ReLU
Dropout Rate	0.2
Output Activation	Softmax

Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Evaluation Metric	Accuracy
Batch Size	32
Number of Epochs	30

Language Model Integration

The language model is integrated as a post-processing module to refine the output of the acoustic speech recognition system. Figure 7 shwos the language model integration workflow. The complete procedure is carried out in the following steps:

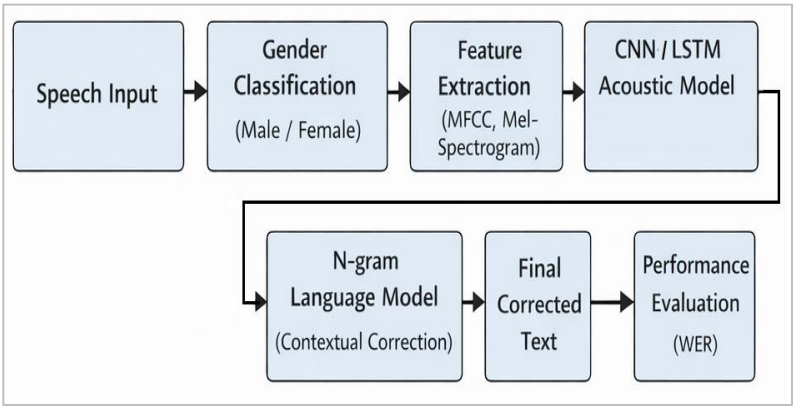


Fig. 7 Language Model Integration Workflow

Step 1: Gender-Based Text Segregation: After speech input is classified into male and female categories, the corresponding gender-specific acoustic model generates an initial text transcription. This segregation helps reduce acoustic ambiguity but does not fully eliminate pronunciation-induced errors.

Step 2: Generation of Initial ASR Output: The CNN model processes the extracted MFCC and Mel-spectrogram features to produce a preliminary word sequence. Due to impaired pronunciation, this output may contain substitution, deletion, or insertion errors.

Step 3: Tokenization of Recognized Text: The recognized text is tokenized into individual words to form candidate word sequences. These sequences serve as inputs to the language modeling stage.

Step 4: N-Gram Probability Estimation: Based on the previous n−1n words, an n-gram language model is used to estimate the conditional probability of each word. The model computes likelihood scores for possible word sequences using statistical frequency information learned from the training corpus.

Step 5: Contextual Error Correction: Low-probability or linguistically inconsistent word sequences are replaced with more probable alternatives. This step corrects acoustically misrecognized words by enforcing contextual and grammatical constraints.

Step 6: Optimal Word Sequence Selection: The final transcription output is determined by taking the word sequence with the highest likelihood according to the language model.

Step 7: Performance Evaluation: The revised transcription is assessed using WER, facilitating a quantitative analysis of the influence of language model integration on recognition precision.

Model Evaluation

Performance Metrics: To evaluate the model's recognition accuracy, employ standard measures including WER, accuracy, macro F1-score, and loss curve.

Word Error Rate (WER) is calculated as;

$$WER = \frac{S + D + I}{N}$$

Where:

S = Substitutions

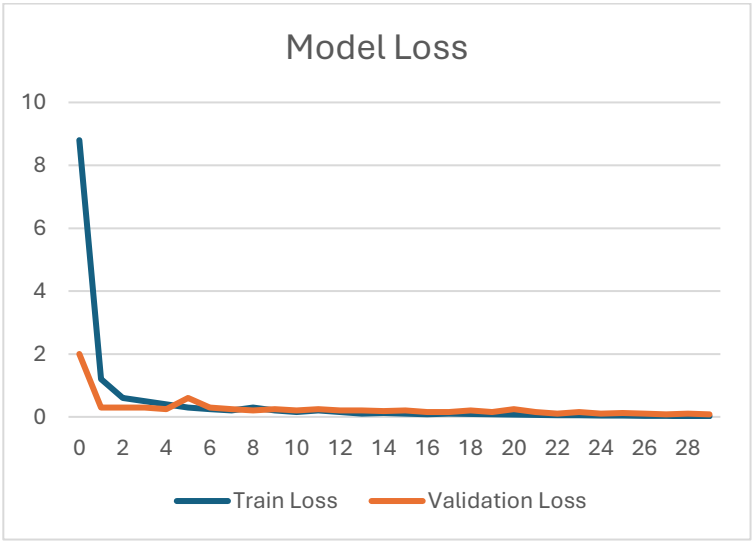
D = Deletions (missing words)

I = Insertions

N = Number of words in the actual (reference) text

Results Analysis

The analysis section on impaired speech recognition performance delineates the efficacy of models designed for impaired speech recognition. The impairment performance is analysed based on CNN's effectiveness in discriminating male and female voices model and impairment recognition using a language model. The CNN model had an accuracy of 96% which was a good indication the model can be used to classify voice samples from speech impaired individuals. The training and validation accuracy curves had an increase which was steady and predictably maintained at a high accuracy which was a good indication of learning and generalization to new data. This means the training and validation loss for impaired speech recognition had a trend that decreased which showed there was some learning but there's no correct overfitting. Figure 8 shows the accuracy and loss curve, which is included in this section.



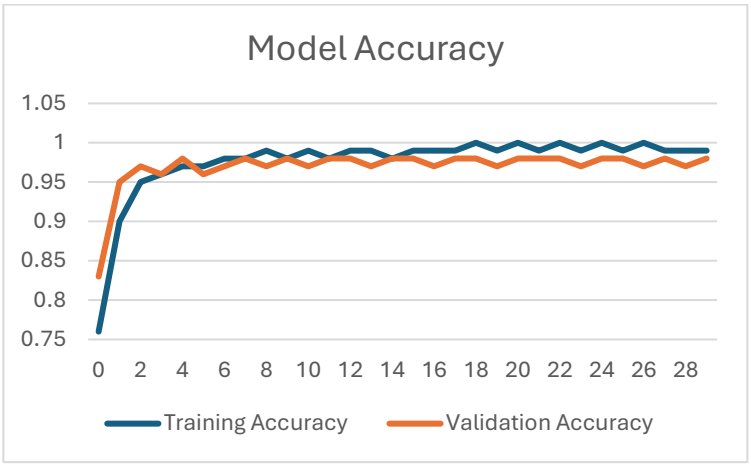


Fig. 8 Accuracy and loss curve of CNN model

To further understand how well the model worked, we constructed the confusion matrix. It was found that approximately 1281 male samples and 402 female samples were accurately predicted. The small number of errors confirms the model's practical applicability for real-life voice-controlled systems. The confusion matrix is presented in Figure 9 below.

Actual \ Predicted	Female	Male
Female	402	120
Male	175	1281

Fig. 9 Confusion Matrix

Figure 10 shows a performance comparison between various models in terms of accuracy. It was discovered that the proposed method considerably outperforms other approaches with its 96% accuracy. Specifically, [20] FF-ATT-BILSTM achieves an accuracy of 92.1%, [18] BI-GRU + MULTI-HEAD ATTENTION is 88.93%, and [19] RFID + RANDOM FOREST represents the weakest performance of 80%. This substantial enhancement proves that the suggested approach outperforms other models created in earlier studies with regard to the accuracy of predictions

.

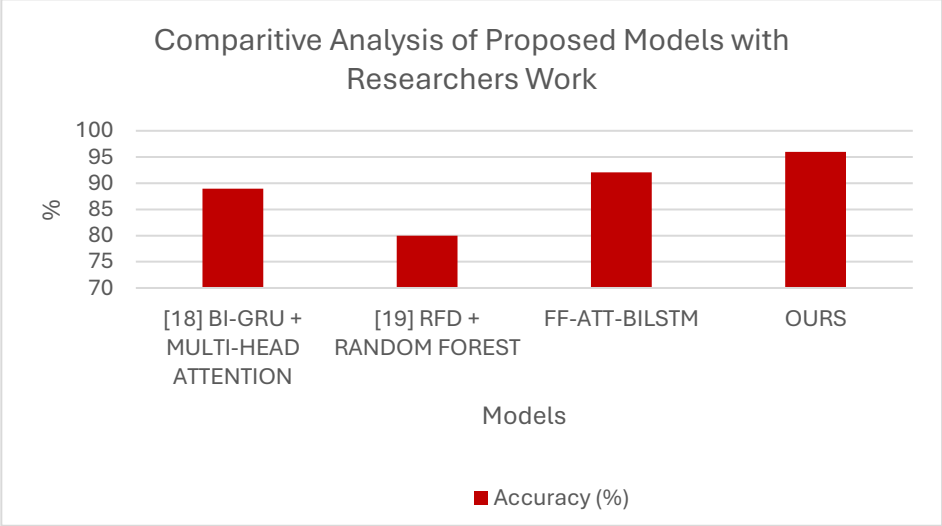


Fig. 10 Comparative Analysis of Researchers Work with Proposed Model

In terms of accuracy, the speech recognition system's performance analysis shows that different audio samples are not all created equal. When evaluating the efficacy of identified words, WER score measures are employed. Figure 11 shows the original filename along with predicted text and actual text and WER Score while Table 3 shows the WER Score Calculation of Input Audio.

Filename	Predicted Text	Actual Text	WER Score	Missing Words	Reconstructed Text
common_voice_hi_24758397.mp3	Gujarat	Gujarat Aatankee hamleki aashanka samudri Seema para high alert Jarry	0.900	Seema, high, aashanka, samudri, Jarry, para, hamleki,	Gujarat Aatankee hamleki aashanka samudri Seema para

				Aatankee, alert	high alert Jarry
common_voice_hi_25285588.mp3	merely tennis	mere liye tennis khel na bahut aasaan hai	0.875	aasaan, liye, hai, bahut, na, khel, mere	mere liye tennis khel na bahut aasaan hai
common_voice_hi_23937568.mp3	Dukan Aaj Coulee nahi	Dukan Aaj khulee nahi Hai	0.40	Hai, khulee	Dukan Aaj khulee nahi Hai

Fig. 11 Predicted Text and Reconstructed Text along with WER Score

Table 3. WER Score Calculation of Input Audio

Filename	**Actual Text**	**Predicted Text (incorrectly pronounced)**	**Reconstructed Text**	**N**	**S**	**D**	**I**	**WER**
24758397.mp3	Gujarat Aatankee hamleki aashanka samudri Seema para high alert Jarry	Gujarat	Gujarat Aatankee hamleki aashanka samudri Seema para high alert Jarry	10	0	9	0	0.90
25285588.mp3	mere liye tennis khel na bahut aasaan hai	merely tennis	mere liye tennis khel na bahut aasaan hai	8	0	7	0	0.875
23937568.mp3	Dukan Aaj khulee nahi Hai	Dukan Aaj Coulee nahi	Dukan Aaj khulee nahi Hai	5	1	1	0	0.40

5. CONCLUSION

Using a CNN as its focal point, this study concluded with a detailed performance evaluation of damaged voice recognition systems. The CNN demonstrated an extraordinary classification accuracy of 96%, making it an efficient method to classify male and female speech patterns. In addition to that, this model utilized the LSTM networks to incorporate temporal dependencies that enable the system to learn how to process speech pattern sequences over time. Further integration of N-gram techniques significantly improved the model performance by providing more contextual meaning to the recognition process increasing the possibility of recognizing an impaired speech. The N-gram models help in identifying and retreating false recognized words considering the neighbour words in the speech, hence making the speech recognition more accurate in case of speaker impairment such as dysarthria and stroke. The combination of CNNs to analyze the space, LSTMs to analyze the time, and N-grams to analyze the contextual accuracy is a feasible solution to enhance and fix the damaged speech recognition. This study's results indicate that the combination of CNN and LSTM with the N-gram technique is effective and adaptable in identifying and rectifying speech patterns of individuals with impairments. This research is crucial to scholars because it provides a sound basis of approaches and studies in the future of speech recognition technology, which would likely result in the development of a comprehensive disabled voice recognition system. The integrated deep learning system, the adaptive algorithm, and the personalized adaptive acoustic feature is pointing out a new direction and path to where speech recognition will go in the future..

References:

1. Brahmi, Z., Mahyoob, M., Al-Sarem, M., Algaraady, J., Bousselmi, K., & Alblwi, A. (2024). Exploring the role of machine learning in diagnosing and treating speech disorders: A systematic literature review. Psychology Research and Behavior Management, 17, 2205–2232. https://doi.org/10.2147/PRBM.S460283
2. Tobin, J., Nelson, P., MacDonald, B., Heywood, R., Cave, R., Seaver, K., Desjardins, A., Jiang, P.-P., & Green, J. (2024). Automatic speech recognition of conversational speech in individuals with disordered speech. Journal of Speech, Language, and Hearing Research. https://doi.org/10.1044/2024_JSLHR-24-00045
3. Upadhyay, N., & Karmakar, A. (2015). Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study. Procedia Computer Science, 54, 574–584. https://doi.org/10.1016/j.procs.2015.06.066
4. Patil, P. G., Jaware, T. H., Patil, S. P., Badgujar, R. D., Albu, F., Mahariq, I., Al-Sheikh, B., & Nayak, C. (2022). Marathi speech intelligibility enhancement using I-AMS based neuro-fuzzy classifier approach for hearing aid users. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3223365
5. Abdusalomov, A. B., Safarov, F., Rakhimov, M., Turaev, B., & Whangbo, T. K. (2022). Improved feature parameter extraction from speech signals using machine learning algorithm. Sensors, 22(21), 8122. https://doi.org/10.3390/s22218122
6. Al-Qatab, B. A., & Mustafa, M. B. (2021). Classification of dysarthric speech according to the severity of impairment: An analysis of acoustic features. IEEE Access, 9, 18183–18194. https://doi.org/10.1109/ACCESS.2021.3053335
7. Labied, M., & Belangour, A. (2021). Automatic speech recognition features extraction techniques: A multi-criteria comparison. International Journal of Advanced Computer Science and Applications, 12(8). https://doi.org/10.14569/IJACSA.2021.0120821
8. Reddy, M. K., Alku, P., & Rao, K. S. (2020). Detection of specific language impairment in children using glottal source features. IEEE Access, 8, 15273–15279. https://doi.org/10.1109/ACCESS.2020.2967224
9. Kanabur, V., Harakannanavar, S. S., & Torse, D. (2019). An extensive review of feature extraction techniques, challenges and trends in automatic speech recognition. International Journal of Image, Graphics and Signal Processing, 5, 1–12. https://doi.org/10.5815/ijigsp.2019.05.01
10. Darekar, R. V., & Dhande, A. P. (2018). Emotion recognition from Marathi speech database using adaptive artificial neural network. Biologically Inspired Cognitive Architectures, 23, 35–42. https://doi.org/10.1016/j.bica.2018.01.002
11. Rokade, P., Kadam, A., Shinde, D., Yadav, S., & Sali, N. (2018). Indian sign language recognition system in Marathi language text. International Journal of Computer Applications, 182(30), 19–22. https://doi.org/10.5120/ijca2018918202
12. Ajibola Alim, S., & Khair Alang Rashid, N. (2018). Some Commonly Used Speech Feature Extraction Algorithms. In From Natural to Artificial Intelligence - Algorithms and Applications. IntechOpen. https://doi.org/10.5772/intechopen.80419
13. Ghule, K. R., & Deshmukh, R. R. (2015). Automatic speech recognition of Marathi isolated words using neural network. International Journal of Computer Science and Information Technologies, 6(5), 4296–4298. DOI:10.35940/ijitee.L2651.1081219
14. Magre, S., Deshmukh, R., & Shrishrimal, P. (2013). A comparative study on feature extraction techniques in speech recognition.
15. Gaikwad, S., Gawali, B., Yannawar, P., & Mehrotra, S. (2011). Feature extraction using fusion MFCC for continuous Marathi speech recognition. In Proceedings of INDCON. https://doi.org/10.1109/INDCON.2011.6139372
16. Yadav, S., Kumar, A., Yaduvanshi, A., & Meena, P. (2023). A review of feature extraction and classification techniques in speech recognition. SN Computer Science, 4(6). https://doi.org/10.1007/s42979-023-02158-5
17. Hu, X., Song, A., Wei, Z., & Zeng, H. (2022). StereoPilot: A wearable target location system for blind and visually impaired using spatial audio rendering. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, 1621–1630. https://doi.org/10.1109/TNSRE.2022.3182661
18. Xu, C., Liu, Y., Song, W., Liang, Z., & Chen, X. (2024). A new network structure for speech emotion recognition research. Sensors, 24(5), 1429. https://doi.org/10.3390/s24051429
19. Hameed, H., Lubna, Usman, M., et al. (2024). Artificial intelligence enabled smart mask for speech recognition for future hearing devices. Scientific Reports, 14, 30112. https://doi.org/10.1038/s41598-024-81904-y
20. Feng, Y. (2024). Intelligent speech recognition algorithm in multimedia visual interaction via BiLSTM and attention mechanism. Neural Computing and Applications, 36, 2371–2383. https://doi.org/10.1007/s00521-023-08959-2
21. Labied, M., & Belangour, A. (2021). Automatic speech recognition features extraction techniques: A multi-criteria comparison. International Journal of Advanced Computer Science and Applications, 12(8). https://doi.org/10.14569/IJACSA.2021.0120821
22. Jalil, A., Hasan, F., & Alabbasi, H. (2020). Robust hybrid features based text independent speaker identification system over noisy additive channel. Journal of Engineering and Sustainable Development, 24(4), 56–70. https://doi.org/10.31272/jeasd.24.4.7
23. Vivmankar. (n.d.). Hindi speech classification dataset. Kaggle. https://www.kaggle.com/datasets/vivmankar/hindi-speech-classification
24. Mulmule, Pallavi V., Rajendra D. Kanphade, and Dhiraj M. Dhane. "Artificial intelligence-assisted cervical dysplasia detection using papanicolaou smear images." The Visual Computer 39.6 (2023): 2381-2392.

25. Shimbre, Nivedita, and Ram Kumar Solanki. "Activation heatmap-guided FT-MultiCNN: advancing skin cancer classification through transfer learning." Ingenierie des Systemes d'Information 30.5 (2025): 1349.
26. RG, Hemanth Kumar, et al. "Engineering Data-Driven Approaches for Classifying Tumor Types." Proceedings of the 6th International Conference on Information Management & Machine Intelligence. 2024.