MARATHI ISOLATED SPEECH RECOGNITION FOR DISEASES USING HTK IN HEALTHCARE SECTOR

RUTUJA RAVINDRA VYAPARI[1], DR. SUNIL S. NIMBHORE[2]

Dr. Babasaheb Ambedkar Marathwada University, ch. Sambhaji Nagar (Aurangabad) - 431005 Maharashtra

ABSTRACT: *Marathi, as a language, holds significant importance in Maharashtra, India, with a vast number of speakers. This review paper focuses on exploring the use of the Hidden Markov Model Toolkit (HTK) for Marathi isolated word recognition specifically tailored for medical terms and disease diagnosis. The paper provides an overview of the existing literature, methodologies, challenges, and future directions in this domain. Talking has an inherent and effective method of human communication. Speaking is the most essential form of communication, and speaking is its primary medium. The diction of every individual is unique. This paper provides an overview of recent advancements in Marathi speech recognition. In this approach, speech samples from three distinct speakers are compiled for the database. The first phase of building an autonomous speech detection System is Speech Database Development. This paper seeks to develop a Marathi-based voice recognition system for diseases.*

Keywords: HMM, ASR (Automatic Speech Recognition), Speech Recognition (SR), etc.

1. INTRODUCTION:

Speaking is an especially natural form of communication. Since childhood, everyone has been conversant with his native dialect. It also promotes effective communication between humans and machines. In order for machines to fathom conversations between people, computers can act as an intermediary between human specialists and machines so that machines may react accurately and consistently to people dialects. This is possible by means of the use of a speech identification structure, which allows data processing equipment to distinguish and convert into written text the phrases individual talks into a headset or telephone. Speech is input into a machine known as a Computerized Speech Recognition System as a whole Artificial Voice Recognition, Computer Speech Awareness, or Speech to

Text system, which receives the command and converts it into text format. Information is typically transmitted between machines and human beings via mouse, try mouse, etc. However, humans may speak faster than computers can transcribe. Speech input offers details with a high data transfer rate and relative ease of use. It also enables someone's fingertips and focus to be employed through an activity, which can be especially useful when the user is engaged in exercise or in a natural setting [1]. When compared to written output, speech is both more stunning and intelligible. Vocal interfaces provide solutions for these issues. The components of speech interfacing are speech creation and recognition. A speech synthesizer functions as a computers converter by converting the input of text into speech output. Speech recognition software converts conversations to text [2]. This paper seeks to develop and put into action a Marathi speech recognition system.

1.1 History of speech recognition [4]:

In Audrey, who's a system that recognized speech developed at Bell Laboratories in 1952, was the first to discern a single individual's spoken numbers. IBM debuted a product capable of recognizing sixteen English words in 1962. Together with the USA, England, and Japan, the Soviet Union created a device that detects four vowels and 9 consonants. The "Harpy" speech recognition engine at Carnegie Mellon identified 1,011 utterances between 1971 and 1976. Bell Labs and Thresholds Science are the first commercial speech recognition companies capable of interpreting the utterances of multiple people. In 1980, The Hidden Markov Model (HMM) is a novel statistical method that has dramatically improved the accuracy with which of terms it could identify from 100 at first to multiple million and then to a limitless quantity. In 1987, toddlers could teach the Worlds of Wonder Julie marionette to reply to their voice. IBM establishes a 5,000-word vocabulary for Kurzweil's speech-to-text system in 1985. Dragon Dictate, the company's first commercial speech recognition product, identified a hundred phrases every minute and needed forty-five minutes of training. Bell South's Voice Activated Link (VAL) speech recognition technology, which allowed for dial-in cooperation and delivered information depending on the speaker's voice was first deployed in 1996 phone conversation. In 2001, the accuracy of voice recognition systems reached 80%.[1].

1.2 About Marathi language:

The Marathi population from the western and central regions of India speak the Indo-Aryan tongue of Marathi. Marathi language is akin to the nation's Hindi. Both languages are derived from the Sanskrit language. The script known as Devanagari is the writing script used. Marathi has among the four highest native speaker population in India. Marathi is spoken throughout the entire province of Maharashtra. The writing direction is left to right. The form of Devanagari utilized to write Marathi differs marginally from that of Hindi and other languages. It employs both consonants and vowels that are that are absent in other dialects that use the Devanagari script. [32]

2. LITERATURE REVIEW:

Vidyashree Kanabur et.al (2019) "An Extensive Review of Feature Extraction Techniques, Challenges and Trends in Automatic Speech Recognition" This task is achieved by Automatic Speech Recognition (ASR) system which is typically a speech-to-text converter. In order to recognize the areas of further research in ASR, one must be aware of the current approaches, challenges faced by each and issues that needs to be addressed. Therefore, in this paper human speech production mechanism is discussed. The various speech recognition techniques and models are addressed in detail. The performance parameters that measure the accuracy of the system in recognizing the speech signal are described.

Trishna Barman et.al (2017) "State of the Art Review of Speech Recognition using Genetic Algorithm" In recent years, speech recognition has become of the most emerging topic for many researchers. Speech is one of the most important tools for communication between human and his environment. People are so easeful with speech that they would also like to intercommunicate with computers via speech, rather than having to apply to ancient adapters. This paper covers mainly two topics firstly, what are the different speech recognition techniques and secondly, how genetic algorithm helps in speech recognition.

Sheena Christabel Pravin et.al (2016) "Isolated Word Recognition for Dysarthric Patients" In this paper, an HMM based speech recognition system is proposed for the Dysarthric patients. The speech samples recorded from patients with Spastic Dysarthria with mid to high intelligibility are taken from the UA Research database. The speech samples are de-noised using Discrete Wavelet Transform (DWT), after which the MFCC, LPC features are extracted. A comparative study on speech recognition using MFCC and LPC are presented. The recognition efficiency is found to be 68.50% using MFCC features and 66.54% using LPC features.

Namrata Dave et.al (2013) "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition" The automatic recognition of speech, enabling a natural and easy to use method of communication between human and machine, is an active area of research. Speech processing has vast application in voice dialing, telephone communication, call routing, domestic appliances control, Speech to text conversion, text to speech conversion, lip synchronization, automation systems etc. Nowadays, Speech processing has been evolved as novel approach of security.

Ravindra Parshuram Bachate et.al (2019) "Automatic Speech Recognition Systems for Regional Languages in India" Speech recognition systems has made remarkable progress in last ¬few decades such as Siri, Google assistant, Cortana. For improving the automation in services of all sectors including medical, agriculture, voice dialling, directory services, education, automobile etc., ASR systems must be built for regional languages as most of the Indian population in not familiar with English. Lots of work is done for English language but not for regional languages in India. Developing ASR and ASU systems will change the scenario of current service sector.

Shital S. Joshi et.al (2019) "A Review on Marathi Speech Recognition"This paper is an attempt to summarize the work done in the area of Speech Recognition for the Marathi language. Lot of work is being done in Asian languages. However, in the case of Indian languages, there is a scope for development. Work done in Indian languages is negligible. Researchers are working in Tamil, Telugu, Hindi, Bengali and Marathi. Marathi is a versatile language. It has a variety of effects. The work in Indian languages, in terms of application-oriented is very less. The language technology can help to bridge the gap between technically illiterate people to join the mainstream & be part of Digital India & avail the benefits of it.

Gaurav D. Saxena et.al (2022) “Extricate Features Utilizing Mel Frequency Cepstral Coefficient in Automatic Speech Recognition System” In this paper, we introduce the process to extricate the feature from the signal utilizing Mel-frequency cepstral coefficients. Mel-frequency cepstral coefficients are a genuinely far wide and proficient methodology for feature extraction from a sound file. This technique improved the speech recognition process and removes the distortion in the voice. In this manuscript we applied the Mel-frequency filtration process to improve speech and remove the background noise.

Kayte CharansingNathoosing et.al (2012) “Isolated Word Recognition for Marathi Language using VQ and HMM” This paper describes the implementation of Marathi Swar, an experimental, speaker-dependent, real-time, isolated word recognizer for Marathi. The motivation and the advantages of choosing Marathi as the language for recognition are discussed here. The results obtained with Marathi Swar for tests conducted on a vocabulary of Marathi digits for 2 male and 2 Female speakers were presented in the end. The rest of the paper discussed the implementation of the system.

Somnath Hase et.al (2023) “Speech Recognition Techniques in Agriculture Sector: A Comparative Study” In this research, various techniques related to speech recognition in field agriculture have been studied in detail. Most of the speech recognition systems are based on isolated keywords specifically designed for accessing the price of commodities. Some are designed to assist farmers based on crop symptoms. There is an agricultural voice-based speech recognition system through oral description using isolated speech databases. This study helps to compare different tools and techniques of speech recognition that are used in agriculture.

Leena R Mehta et.al (2013) “comparative study of MFCC and LPC for Marathi isolated word recognition system” This Paper presents Marathi database and isolated word recognition system using Mel-frequency cepstrum coefficients (MFCC) and vector quantization (VQ) technique. It also compares the performances of MFCC and LPC features under VQ environment. Marathi speech database is recorded in noisy environment aiming language learning tool as an application. The database consists of simple Marathi words starting with both vowels and consonants.

Sai Sawant et.al (2018) "Isolated Spoken Marathi Words Recognition using HMM" This paper presents a speaker independent word recognition system for Marathi language. Hidden Markov Model Toolkit (HTK) is used to implement the system. Mel Frequency Cepstral Coefficients (MFCCs) of phonetically rich 20 Marathi words collected from ten native speakers are used to train HMMs. Viterbi algorithm is used to recognize test word utterances. The performance of phoneme based HMMs and their recognition accuracy are discussed.

Gajanan Pandurang Khetri et.al (2012) "Automatic Speech Recognition for Marathi Isolated Words" In this paper, speech recognition is one of the techniques of biometrics, to collect the words convert to an acoustic signal. The use of isolators Marathi word feature extraction, to identify and verify each spoken word is a field that is being actively researched. The results from MFCC and VQ algorithm apply on IWAMSR databases, found easy access of end users.

Ganesh B. Janvale et.al (2014) "Recognition of Marathi Isolated Spoken Words Using Interpolation and DTW Techniques" This paper contains a Marathi speech database and isolated Marathi spoken words recognition system based on Mel-frequency cepstral coefficient (MFCC), optimal alignment using interpolation and dynamic time warping. Initially, Marathi speech database was designed and developed though Computerized Speech Laboratory. The database contained Marathi isolated words spoken by the 50 speakers including males and females.

Mayur Babaji Shinde et.al (2013) "Speech processing for isolated Marathi word recognition using MFCC and DTW features" After studying the history of speech recognition we found that the very popular feature extraction technique Mel frequency cepstral coefficients (MFCC) is used in many speech recognition applications and one of the most popular pattern matching techniques in speaker dependent speech recognition is Dynamic time warping (DTW). The signal processing techniques, MFCC and DTW are explained and discussed in detail and these techniques have been implemented in MATLAB.

Rohini B. Shinde et.al (2012) "Isolated Word Recognition System based on LPC and DTW Technique" In this paper, we present the steps involved in the design of a speaker-independent speech recognition system for Marathi Language. We focus mainly on the pre-processing stage that extracts salient features of a speech signal and a technique called Dynamic Time Warping commonly used to compare the feature vectors of speech signals. These techniques are applied for recognition of isolated as well as connected words spoken. In his case the experiment is conducted in MATLAB to verify these techniques.

3. Methodology:

The following example illustrates the development of SVASR with the proposed procedure. The application that is exemplified is building a recognition system for the Indonesian voice based electrical switcher (IDSwitch) and planting it into home automation electrical circuit switches.

A. Prerequisites softwares

Besides preparing computer and Bluetooth microphones, supporting software, libraries and applications must be installed as seen in TABLE I. HTK is downloaded and then compiled using Cygwin-gcc or Visual Studio. A slight modification of HTK source prior compilation is required to support raw databases recording, volume scaling and Principle Component Analysis, PCA[11].

Table 1. Softwares and Development Libraries

A. ***No.***	B. ***Software, Libraries***		
	C. ***Software***	D. ***Version***	E. ***License***
F. *1*	G. *Speech filling System*	H. *Release 4.9, SFSWin 1.9*	I. *UCL*
J. *2*	K. *HTK*	L.	M.

N. *3*	O. *Visual Studio*	P. *2008*	Q. *Microsoft*
R. *4*	S. *Praat*	T. *6.0.43*	U. *Paul*
V. *5*	W. *Windows OS*	X. *Pro 10.0.1439*	Y. *Microsoft*
Z. *6*	AA. *Sygwin-gee*	BB. *2.11.2(-.329/5/3)- i686*	CC. *GPL*

B. Grammar:

Grammar for IDSwitch is as shown in Fig.2. In this figure the first 4 natural numbers in Indonesian language represent the number of electrical switches for home installation. The bar '|' means logical 'OR'

```
SWITCH_DGT = NUM_0 | NUM_1| NUM_2 | NUM_3 | NUM_4 ;
SILENCE = SIL ;
SWITCH_W = BATHROOM | GARDEN | AIR_COND;
ON_OFF = ON| OFF|HIDUP|MATI;
CONT = < SWITCH_DGT > ON_OFF | SWITCH_W  ON_OFF;
SIL_NOISE = <SIL | NOISE> ;
SENTENCE = { <$CONT SIL_NOISE>  } ;
(SILENCE SENTENCE SILENCE)
```

Fig. 2. Grammar for IDSwitch

In addition, for the dictionary of the spelling of the sub-word pronunciation dictionary following the phoneme list in [6] is quoted in part as in Fig.3:

```
NOISE     t sp
NOISE     d  sp
NUM_0     k oh s oh ng
NUM_1     s ah t uh
NUM_2     d uw  ah
NUM_3     t ih g ah
NUM_4     ah m p ah t
```

Fig. 1.Pronunciation dictionary untukIDSwitch

So, on the dictionary at Fig.3 above the word is placed in the first column and the pronunciation is placed in the next columns. Sub-word sp at the end of some entries in the dictionary above stands for short pause, which is if silence occurs, but in a short time. The word NOISE represents noise that might occur during testing. For example, in Fig.3 noise that appears is defined as a pattern of certain phoneme sequences such as the sequence of two consonants without vowels between. In Indonesian some words appear sequentially as contraction which is the suffix of the word is the prefix of the next word. An example is if the sentence NUM_3 NUM_4 is pronounced then /ah/ is the phoneme suffix of the word NUM_3 at the same time become the prefix to the next word NUM_4. In this case HVite often only recognizes one word, which means that word deletion error occurs. The solution used is to create new words in the dictionary. An example is for the case of NUM_3 NUM_4 the solution is like in Fig. 4

NUM_3_4 t ih g ah m p ah t

Fig. 4. Contraction modeling in dictionary

Contraction modelling with dictionary entry is effective enough for SVASR, but could be a problem when applied to LVASR because of huge entries in order to accommodate all possibilities of contractions.

C. Speech database:

- **Voice recording:**

Sound is recorded using HSLab and HVite software with bluetooth microphone input devices. Other recording softwares such as Audacity and CoolEdit is possible, but these softwares are not used, because the desired database is really the one that equal with the test conditions. For bluetooth microphones used are common branded ones and are easily found on the market. This device can be obtained at a low price but without the noise cancellation function. This microphone also contains only one single sensor as according to various literature is less good than those containing an array of censors. An example of a microphone used can be seen in Fig. 5 The advantage of using a wireless microphone is that the speaker is free to move and not tied to an ASR processor, computer or smartphone

- **Transcribing (annotating):**

Sound is transcribed using Praat [8], then converted into SFS [7] format and converted again into HTK format. This is because Praat cannot convert directly to the HTK format. To label training data, the subword (phoneme) sequences follow the grammar generated in section B. In Figure 7 a sample example of subword with Praat is shown for the order of words NUM_2 NUM_1.

- **Noise data:**

In this phase, noise datas are included into training database. Noise present is a must in online testing. Example of noise is shown in SFS display as shown in Figure 8.

- **Feature Extraction**

Feature extraction to convert waveform to MFCC (Mel Frequency Cepstral Coefficient) format. In this case, the raw waveform sound is recorded from the replay buffer of HVite or using HSLab program recording. An HTK program called HCopy, which is to convert raw PCM sound files with extension. htk to feature files with extension. mfc, does feature extraction. The specific MFCC format used is with the HTK code MFCC_0_D_A. This format is the MFCC format with spectral coefficients to zero, (zeroth coefficient), delta coefficient and acceleration coefficient [8]. Previously preprocessing was done, namely emphasizing with coefficients of 0.97 with windowing of 250 ms (400 samples) consisting of framing, striding of 100ms (160 samples) and overlapping of 150 ms (240 samples). After obtaining MFCC_0_D_A, then proceed with PCA [11], where based on the results of PCA, the number of the required coefficients can be reduced.

- **Acoustic model creation**

In HTK, acoustic models are made using HInit and HRest and HERest programs. HInit initializes parameters using the Viterbi Extraction algorithm. HRest estimates HMM-GMM parameters using the Baum-Welch algorithm for small vocabulary and fixed variance version [1]. HRest is outperformed HERest here because of better performance in noisy environment and complete dataset. This is because HERest is mainly intended to train acoustic speech where the exactly sub-word time boundary transcription is not completely present[1] but only complete sentence patterns that are available. So, for complete labelled data, only HInit and HRest is used as shown in figure 9.

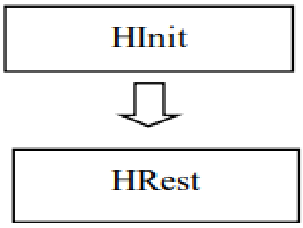

Fig. 2. HTK programs for creating HMM acoustic model

- **Testing:**

Tests were carried out with the HVite program. Hvite can perform testing offline using a recorded database or online directly from the microphone. In this case Bluetooth microphone is used.

- **Refining:**

Repair is done by looking at the results of online testing. If in online testing an error occurs in the sentence that is spoken, then the sentence is taken and labelled and added to the database, then retest the sentence again. After that, it is examined whether the previous sentence can be recognized correctly now. The results showed that this method succeeded in correcting the errors of online testing.

Sometimes even though errors data have been entered in the database but still cannot eliminate errors, repairs can be done by looking at the phoneme sequence that is wrongly detected then adding the entry to the dictionary as a word explicitly. At Fig. 10 we show a multi-version entry for spelling NUM_3, whenever there is an entry that is not explicitly added mainly will cause an error

```
NUM_3        t+ih ih-g+ah g-ah+ah
NUM_3        ah-t+ih t-ih+g ih-g+ah uw-ah
```

Fig.3. Multiple word pronunciation accommodation in dictionary

- **Deployment:**

Planting is done on computers with Windows operating systems within the same noise environments between testing and training stages. The software that is planted is HVite from HTK. The speaker will pronounce the digits repeatedly and randomly via the Bluetooth

ISSN: 2278-6252
Impact Factor: 7.388

microphone input. HVite will record and recognize the speech based on the trained HMM-GMM acoustic model. The records kept can be used to refine the system whenever the error occurs.

An example of deploying SVASR application is presented below. This configuration will enable the toggle of electrical switches using voice commands. The main components are:

1) HTK Speech Recognition server, this is a personal computer or smartphone in which a modified HTK is installed and bluetooth microphone is connected.

2) Arduino Uno-Wifi R3 Robodyn, contain Espressif ESP 8266 connected with Atmel Atmega 328p through serial pins. Atmega328p[14] is used to control electrical switch while the ESP 8266[15] is used to communicate with speech recognition server. The appearance of this board as in Figure-11

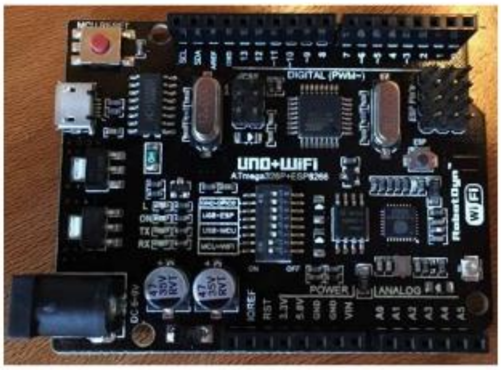

Fig. 4 Embedded board with Atmega 328p and ESP 8266

D. Dataset for Speech Recognition:

सामान्य रोगांची यादी

1. सर्दी
2. अतिसार
3. डोकेदुखी
4. पोटदुखी
5. मधुमेह
6. नैराश्य
7. चिंता
8. मूळव्याध
9. सोरायसिस
10. नागीण
11. न्यूमोनिया
12. खरुज
13. हृदयचा झटका
14. कर्करोग
15. क्षयरोग
16. खोकला
17. थंडी ताप
18. गोवर
19. पीतज्वर
20. गालगुंड
21. मलेरिया
22. कॉलरा
23. मळमळ
24. डांग्या खोकला
25. विषबाधा
26. मूतखडा
27. काचबिंदू
28. कंपवात
29. संधिवात
30. धनुर्वात
31. मोतीबिंदू
32. कांजिण्या
33. रेबीज
34. डेंग्यू
35. विषमज्वर
36. कुष्ठरोग
37. हत्तीरोग
38. चिकुनगुन्या
39. हिवताप
40. घटसर्प
41. काविळ
42. उच्चरक्तदाब
43. दृष्टिदोष
44. अंधत्व
45. पोलिओ

Fig. 5 Dataset for speech recognition

The structure of this essay is as follows: The second chapter on the Marathi language. The third section explains how the text corpora were acquired. The final part discusses the methods of speech data collection. The fifth section details the process of recording and the necessary recording details. In the sixth paragraph, the result and next steps are described.

Outcomes from study:

The ability to recognize and interpret words in Marathi has been an emerging area of research. Developing an accurate speech recognition system for Marathi, particularly for medical terminology, holds great promise in improving healthcare delivery to the local population. In this review, we explore the advancements made in Marathi isolated word recognition using the HTK toolkit, with a specific focus on diseases.

Datasets and Corpus:

An overview of the datasets and corpora used in Marathi isolated word recognition for diseases is provided. This section highlights the availability of medical speech databases in Marathi and their suitability for training and evaluating speech recognition systems.

Pre-processing Techniques:

Pre-processing is a crucial step in any speech recognition system. Various techniques, such as noise reduction, normalization, and feature extraction, are discussed in this section, emphasizing their importance in Marathi disease word recognition.

Language Modeling for Medical Terms:

Since medical terms can be complex and domain-specific, language modeling plays a critical role in improving recognition accuracy. This section discusses the techniques employed to develop language models tailored to medical terminologies in Marathi.

Performance Evaluation Metrics:

The evaluation of a speech recognition system requires well-defined metrics. This section outlines the standard metrics employed to assess the performance of the HTK-based Marathi isolated word recognition system for diseases.

Challenges and Limitations:

Recognizing medical terms in Marathi poses specific challenges, including dialectal variations, speaker diversity, and limited medical speech databases. This section provides an in-depth analysis of the challenges faced in the development of a robust disease word recognition system.

State-of-the-Art Approaches:

In this section, the most recent state-of-the-art approaches using HTK for Marathi isolated word recognition are reviewed. It covers the advancements made to address the challenges and improve system accuracy.

4. Suggestion for Future Work:

This review paper summarizes the current state of Marathi isolated word recognition for diseases using HTK. It underscores the importance of accurate speech recognition in the healthcare sector and calls for further research to overcome existing challenges and improve system performance.

The review concludes by highlighting potential research directions and innovations in the field of Marathi isolated word recognition using HTK. It emphasizes the need for continued efforts in developing accurate and robust speech recognition systems for the benefit of the healthcare domain.

To enhance the system's recognition capabilities, the dataset can be expanded to include a broader variety of disease-related terms. In addition, fine-tuning the parameters of the Hidden Markov Model Toolkit (HTK) and investigating other advanced machine learning techniques, such as deep learning with neural networks, could potentially result in improved disease word recognition accuracy and robustness. In addition, the efficacy of the system can be evaluated on a larger scale using real-world healthcare data to assess its applicability and generalizability. In addition, efforts can be focused on developing a user-friendly interface or integrating the technology with existing healthcare applications to make it accessible and beneficial for medical professionals and patients. Overall, these prospective endeavours will contribute to the advancement of Marathi speech recognition for medical purposes, expediting diagnosis and enhancing healthcare outcomes for Marathi-speaking populations.

References:

[1] N. Dave, “Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition,” *Int. J. Adv. Res. Eng. Technol.*, vol. 1, no. Vi, pp. 1–5, 2013.

[2] S. Sawant and M. Deshpande, "Isolated Spoken Marathi Words Recognition Using HMM," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–4, 2018, doi: 10.1109/ICCUBEA.2018.8697457.

[3] T. Barman and N. Deb, "State of the art review of speech recognition using genetic algorithm," *IEEE Int. Conf. Power, Control. Signals Instrum. Eng. ICPCSI 2017*, pp. 2944–2946, 2018, doi: 10.1109/ICPCSI.2017.8392264.

[4] C. Munteanu, G. Penn, R. Baecker, and Y. Zhang, "Automatic speech recognition for webcasts," no. December, pp. 39–42, 2006, doi: 10.1145/1180995.1181005.

[5] G. D. Saxena, N. A. Farooqui, and S. Ali, "Extricate Features Utilizing Mel Frequency Cepstral Coefficient in Automatic Speech Recognition System," *Int. J. Eng. Manuf.*, vol. 12, no. 6, pp. 14–21, 2022, doi: 10.5815/ijem.2022.06.02.

[6] V. Kanabur, S. S. Harakannanavar, and D. Torse, "An Extensive Review of Feature Extraction Techniques, Challenges and Trends in Automatic Speech Recognition," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 5, pp. 1–12, 2019, doi: 10.5815/ijigsp.2019.05.01.

[7] K. C. Nathoosing, "Isolated Word Recognition for Marathi Language using VQ and HMM," *Online) Sci. Res. Report.*, vol. 2, no. 2, pp. 161–165, 2012.

[8] R. B.Shinde and V. P. Pawar, "Isolated Word Recognition System based on LPC and DTW Technique," *Int. J. Comput. Appl.*, vol. 59, no. 6, pp. 1–4, 2012, doi: 10.5120/9549-4005.

[9] S. Christabel, A. Chellu, and P. Kannan, "Isolated Word Recognition for Dysarthric Patients," *Commun. Appl. Electron.*, vol. 5, no. 2, pp. 14–17, 2016, doi: 10.5120/cae2016652219.

[10] S. S. Joshi, V. D. Bhagile, and R. R. Deshmukh, "A Review on Marathi Speech Recognition," *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3419224.

[11] S. C. Satapathy, P. S. Avadhani, S. K. Udgata, and S. Lakshminarayana, *ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol I*, vol. 248 VOLUME I. 2014.

[12] B. Robert and E. B. Brown, "A Review on Speech Recognition technique," no. 1, pp. 1–14, 2004.

[13] R. P. Bachate and A. Sharma, “Automatic speech recognition systems for regional languages in India,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 3, pp. 585–592, 2019, doi: 10.35940/ijrte.B1108.0782S319.

[14] L. R. Mehta and A. S. Dabhade, “Comparison Study of MFCC and LPC for Marathi Isolated Word,” *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 2, no. 6, pp. 2133–2139, 2013.

[15] S. S. Gangonda and P. Mukherji, “Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features,” *Int. J. Eng.*, vol. 3, no. 1, pp. 218–222, 2013.

[16] S. Hase and S. Nimbhore, “Speech Recognition Techniques in Agriculture Sector : A Comparative Study,” 2023, doi: 10.4108/eai.24-3-2022.2318772.