International Journal of Management, IT & Engineering
Vol. 13 Issue 07, July 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

MARATHI ISOLATED DISEASE WORDS SPEECH RECOGNITION USING HTK

Rutuja Ravindra Vyapari[1], Dr. Sunil S. Nimbhore[2]

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

Abstract: In computer science, speech recognition is commonly used to facilitate organised communication between humans and machines. The regional language of the Indian state of Maharashtra, Marathi, is the topic of this research on voice recognition. Maharashtra has a large Marathi-speaking population. The Hidden Markov Model (HMM) is used in conjunction with the Viterbi approach to identify unfamiliar sentences. The system was trained using 45 isolated Marathi words from the dataset. For the acoustical study of a spoken stream, a Mel frequency cepstral coefficient (MFCCs) is utilised as a feature extraction method. The proposed Marathi automated speech recognition system (DASR) employs a speaker-independent word model. The isolated words in the training and test datasets were spoken by eight native Marathi speakers. 89.77% of the time, the DASR system accurately detected the Marathi words. Recognise performance excellence.

Keywords: Marathi, Mel frequency cepstral coefficient (MFCCs), DASR, Viterbi method, Hidden Markov Model (HMM), etc.

1. Introduction:

The study of creating a computer system that can recognise human speech is known as speech recognition in the area of computer science. There are several strategies used for categorization as well as recognition, including template-based approaches, statistical approaches, learning approaches, knowledge-based approaches, artificial intelligence approaches, etc. Among them, using algorithms built around Hidden Markov Models is one of the most effective techniques. The same was used in the construction of a Marathi isolated digit recognition system.[1]

A. Marathi Language:

Marathi is an Indo-Aryan language spoken by the Marathi people who inhabit western as well as central India. Hindi and Marathi are both national languages. The two languages employ the Devanagari writing system and are descended from Sanskrit. Due to the fact that Marathi is spoken across the whole state of Maharashtra, which spans a large geographic region and is made up of 34 distinct districts, Marathi speakers make up the fourth largest group in India. Standard Marathi is the state's official language. This essay will go like this. The overview of the literature on voice recognition systems created in Indian languages using HTK is presented in Section 2; The definition of HTK (Hidden Markov Model) Toolkit is provided in Section 3; The specifics of the speech database utilised for the study are described in Section 4; Section 5 discusses the Acoustical Analysis, Section 6 the Training Phase, and Section 7 the Recognition Process The performance study outlined in Section 8; the system's conclusion and its roadmap for the future are presented in Section 9; and Section 10 contains the appendices.[2]

B. HTK:

HTK is a software toolkit created by the Cambridge University Engineering Department's Speech Group for creating and modifying systems that use Hidden Markov Models. The HTK software consists of a software library including an assortment of tools (programmes) that may be used to code data, train HMMs in a variety of ways, including one that embeds Baum-Welch re-estimation, decode Viterbi signals, and change HMM definitions. HTK is mostly used for research in voice recognition. HTK was additionally employed in a variety of other projects, including as studies into DNA sequencing, character recognition, and voice synthesis, among others. In HTK, there are four crucial processing stages: data preparation, training, testing/recognition, as well as analysis.[3]

2. Literature Review:

Suvarnsing Bhable et.al (2021): discusses the main goal of Using a device or microphone, the ASR system recognises a speech, which is then converted into text to carry out the required action. In this study, we employed (MFCC), Gaussian Mixture Model (GMM), and Vector quantization (VQ) for feature extraction in order to recognise words that were isolated from Hindi. We've conducted useful research for It was created a Hindi word spoken dataset of different men and men.[4]

Sheena Christabel Pravin et.al (2016): examines an HMM-based speech recognition tool for dysarthric patients is suggested in this research. The UA Research database contains speech samples of Spastic Dysarthria patients with moderate to high intelligibility. The MFCC as well as LPC characteristics are retrieved after the discrete wavelet transform (DWT) denoising of the speech samples. For voice recognition, a comparison of MFCC and LPC is provided. When utilising MFCC features, the efficiency of recognition is 68.50%, whereas when using LPC features, it is 66.54%.[5]

Devyani S. Kulkarni et.al (2016): uses the HTK technique to suggest a solution for the Marathi language's isolated digit recognition. The used database includes 800 utterances from 40 different people. There are twenty males and twenty females. The MFCC approach, also known as Mel frequency cepstral coefficients, is used to train a database's acoustic properties. To identify the isolated Marathi numerals, we employed a word-level model. The system's performance study shows a 99.75% recognition rate with a 48.75% accuracy.[6]

Kayte Charansing Nathoosing (2012): studied the article explains how Marathi Swar, an experimental, real-time, speaker-dependent, isolated Marathi word recognizer, was put into practise. In this post, the rationale and advantages of choosing Marathi as the language for recognition will be discussed. At the end, the findings from tests given to two male + two female speakers on a vocabulary of Marathi numbers with Marathi Swar were presented. The system's implementation was covered in the remaining sections of the study. The noise detection/elimination and HMM training algorithms have been changed in the suggested scheme in comparison to the usual implementation. The results of the experiments demonstrated that the overall accuracy of the shown system was 94.63 percent.[7]

Kishori R. Ghule et.al (2015): suggested method is used with single Marathi words. A list of 100 Marathi words is available. Three 100-word utterances are produced using ASR for 100 speakers. owing to its aptitude for processing non-stationary signals like speech owing to their multi-resolution as well as multi-scale analytic capabilities, Discrete Wavelet Transforms (DWT) are utilised to extract signal attributes. The categorization issue of speech recognition has numerous classes.[8]

3. Methodology:

The speech signal passes through three basic processing phases after being obtained from a microphone. The signal is subjected to pre-processing in the preliminary stages. The second step involves creating MFCC function vectors based on the speech signal utterance. The pattern recognition algorithm then compares these produced feature vectors to the database features. The system's flowchart for isolating words is shown in Fig. 1.

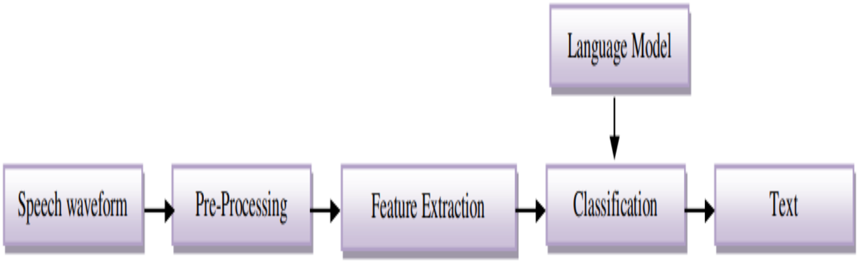


Fig. no.1 process flow of isolated words

A. Speech Waveform:

Individual words that need to be understood and shown on the computer will be sent via a microphone attached to a computer.[9]

B. Signal Pre-Processing:

By locating the border of the uttered word, end point detection is used to separate the words. We first establish a threshold value, and if the energy rises over the threshold or falls below the threshold, we determine the start of the duration and the end of the term, respectively. Lower frequencies are stressed while higher frequencies are muted during speech development. As a result, signal information is lost. The High Pass FIR filter is used in the speaker verification as well as speech recognition system prior to feature extraction in order to avoid data loss and maintain the features of the spoken signal. This method is known as pre-emphasis.

Before the is defined as—

$y(n) = x(n) - a.x(n-1)$; where $0.9 \leq a \leq 1$

C. Mel frequency cepstral coefficients (MFCC) Feature Extraction:

By locating the border of the uttered word, end detection is used to separate the words. We first establish a level of threshold, and then when the energy rises over the threshold or falls below the threshold, we determine the start of the duration and the end of the term, respectively. Lower frequencies are stressed while higher ones are muted during speech development. As a result, signal information is lost. The High Pass FIR filter is used in the speaker verification as well as speech recognition system prior to feature extraction in order to avoid data loss and maintain the features of the spoken signal. This method is known as pre-emphasis.[10]

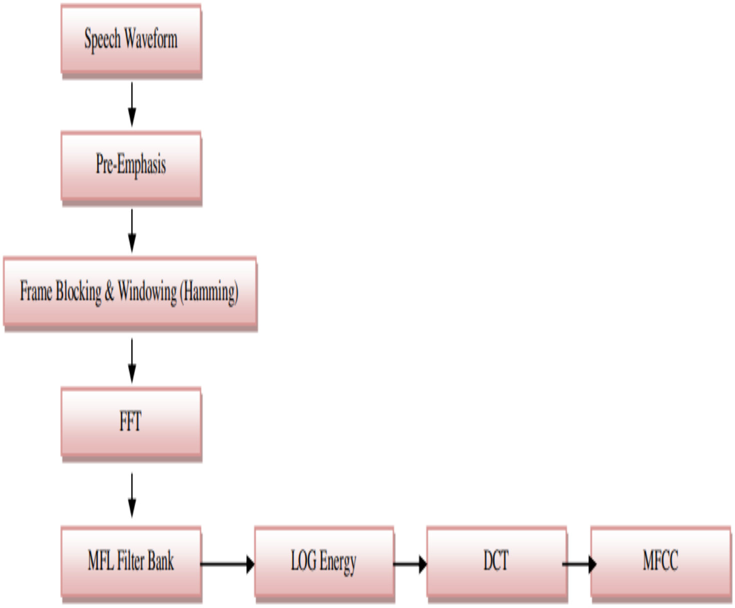


Fig. no.2: Block Diagram of MFCC

4. Isolated Marathi Spoken Word Recognition Using MFCC and HTK:

The employment of techniques to extract features, such as LPC, PLP, LPCC, as well as MFCC, among others, has been reported in the literature on speech recognition. Feature extraction has shown to be more successful with MFCC than other methods. Consequently, MFCC is used as a method for extracting features by the ASR system covered in this study. Acoustic coefficient vectors called Mel-frequency Cepstral Coefficients (MFCC).[11]

A. Features Extraction:

The voice power spectrum, which is determined by early on, in the planning phases. In the second stage, MFCC function vectors are constructed using, or is known as the MFCC. The suggested system's optimum parameters include overlapping frames with a 25millisecond length. A pounding window multiplied by ten frames is the difference between two consecutive frames. The output of 26 Mel filters was used to create 12 cepstral coefficients. With an increased sine and a sine window length of 24, the cepstral coefficients were elevated. A frame of speech is represented by the 12 MFCC that were taken from the feature vector. The features extraction process is shown in Figure 3.

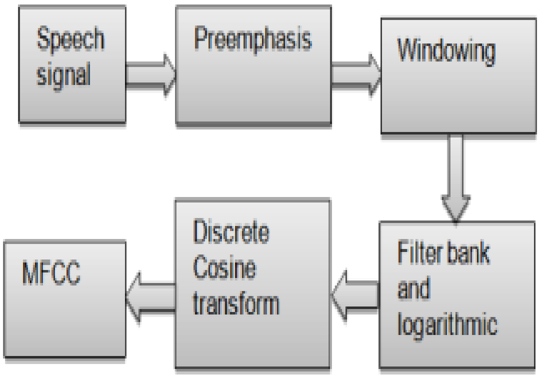


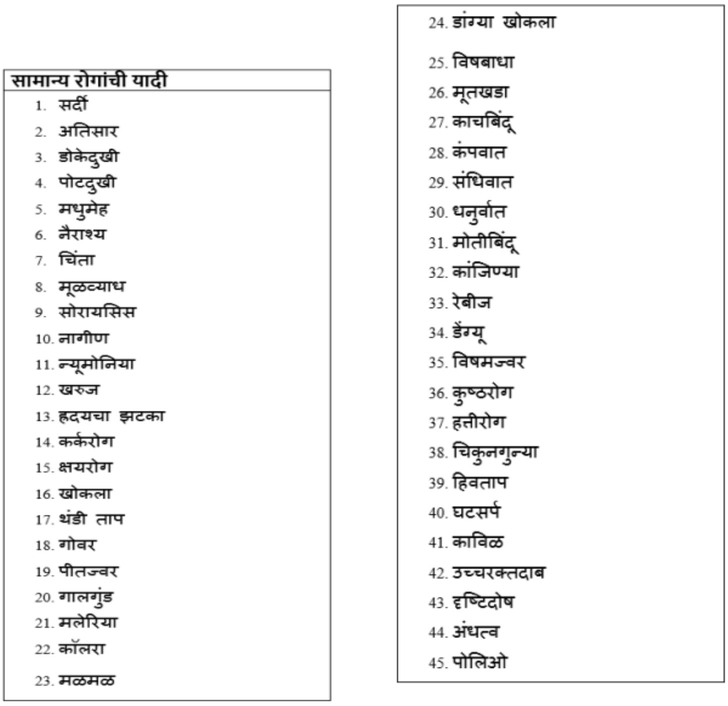
Fig. no. 3. Feature extraction process of MFCC

B. Hidden Markov Model Toolkit (HTK):

The Engineering Department at Cambridge University (CUED) created the Hidden Markov Model (HTK). Upon registering, this programme is made accessible for free. In the domains of mobile apps, IVR applications, pattern recognition, and speech recognition in particular, HTK is widely employed.[11]

5. Data Collection for Marathi Language:

The creation of a Marathi language standard database is the project's main goal. Eight natural speakers of different ages provide instruction to the DASR speaker-independent system using words. This system includes marathi 45 isolated disease words for data collection and we were acquiring 135 speech files. Following being recorded with the wavesurfer-1.8.8p5 toolbox in Audacity 1.3 Beta Unicode sound editing precisely identifies each word. Recording at 16MHz is possible with the Sennheiser PC-350's integrated microphone.



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 no. 4 Marathi Isolated Word

6. Acoustical Model and Task Grammar:

The Viterbi algorithm compares undocumented utterances using an acoustic word model as a foundation. The word model as well as phoneme model are the two types of acoustical models. HMM is used to initialise the word model, and the HMM Proto is established as well as used.[12]

A. Acoustical Model Generation:

There are 84 HMM Protos specified in the present system. Thirteen mel-cepstral coefficients form the feature vector.

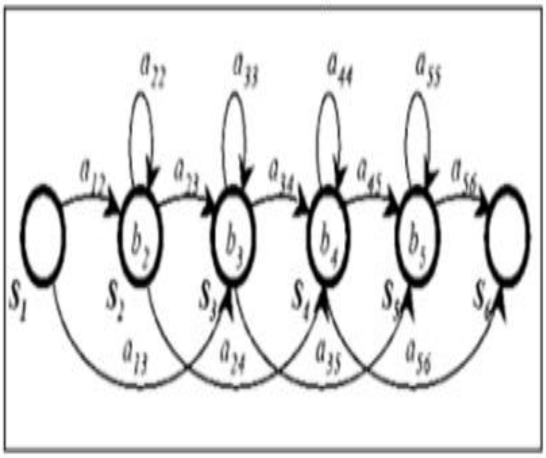


Fig. no. 5 HMM topology using GMM for each word

The MFCC as well as their time-dependent acoustical phones with six emission states were represented using HMM. The Gaussian Mixture Characterization (GMM) approach was used to describe the geographical distribution of each state's features. Figure 2 shows the HMM Proto's states. It is possible to estimate the best HMM proto values with the use of the HTK tool throughout the training phase. This procedure should be performed as often as is required. Three iterations are taken into account in the present approach, necessitating the creation of three HMM for each word in the dataset.

B. Task Grammar & Dictionary of V-ASR:

Work on a pronunciation dictionary is essential for defining the meaning of words. Grammar and a Marathi word dictionary are defined for the DASR (Marathi) job using text files depending Extended Back-Naur Form (EBNF). Network model (.slf) generation is done using the HParse tool. The correspondence also includes the names of the labels that match the symbols that the recognizer will produce.[12]

C. Testing of DASR:

The system that creates transcriptions of undocumented utterances is now in the testing phase. Using the HTK function HCopy, testing signals are now transformed into a string of acoustic vectors (.mfcc format). HVite is a software in the HTK library that generates output based on a combination of input data, HMM definitions, a Marathi word dictionary, a task network (.slf), as well as the names of the HMMs contained in the HMMs list build an output transcription (.mlf) file. With the help of the HVite tool, which analyses speech signals applying the Viterbi algorithm as well as compares test utterances to reference transcriptions found in dictionaries, testing is conducted.

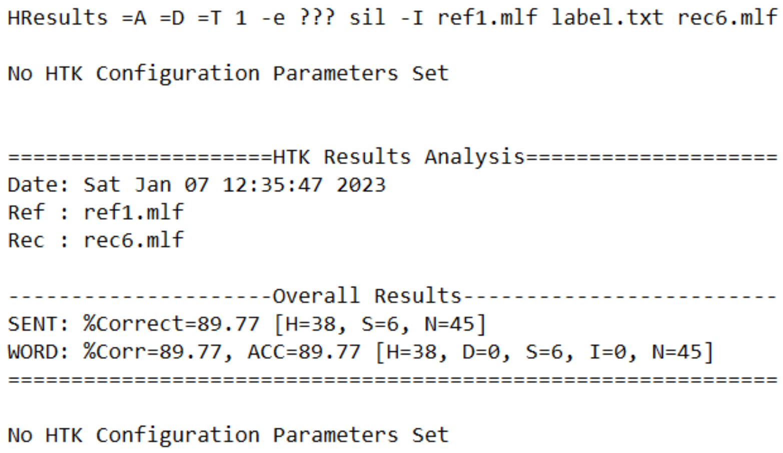
7. Result Analysis:

Word-level evaluations of the DASR system's performance [24]. The HResult tool from the HTK is used to measure the system's efficiency. The DASR system's word correcting rate is shown in Figure 6 and Table 1. Use equations 1, 2, and 3 to get the word accuracy rate (WAR), the word correction rate (WCR), and the word error rate (WER).

$$\textit{Word Correction Rate} = \frac{N - D - S}{N} * 100 \ldots eq.1$$

$$\textit{Word Accuracy Rate} = \frac{N - D - S - I}{N} * 100 \ldots eq.2.$$

When referring to the test set, where N is the total number of words, S stands for the total number of substitutions, D stands for the total number of deletions, as well as I stand for the total number of insertions. The phrase (3) provides an example of word error rate calculation.
$WordErrorRa(WER) = 100 - WAR \dots .eq. 3.$
The word correction rate may be calculated using Equation (1). Word accuracy may be computed using equation (2). Finally, we use equation (3)s to get the percentage of incorrect words.



```
HResults =A =D =T 1 -e ??? sil -I ref1.mlf label.txt rec6.mlf

No HTK Configuration Parameters Set


====================HTK Results Analysis===================
Date: Sat Jan 07 12:35:47 2023
Ref : ref1.mlf
Rec : rec6.mlf

---------------------Overall Results------------------------
SENT: %Correct=89.77 [H=38, S=6, N=45]
WORD: %Corr=89.77, ACC=89.77 [H=38, D=0, S=6, I=0, N=45]
============================================================

No HTK Configuration Parameters Set
```

Fig. no. 6 Recognition accuracy of 45 Marathi Isolated Words

Table 1: Recognition Performance of DASR:

Recognition Accuracy				
Spoken words for testing	Recognized spoken words	W.C.R	W.A.R	W.E.R
		Recognition Accuracy	Percentage Accuracy	Word error rate
45	38	89.77	89.77	10.23%

8. Conclusion:

In conclusion, the Hidden Markov Model Toolkit (HTK) is used to demonstrate a remarkable achievement in Marathi isolated word recognition for maladies. The devised DASR system, customised for the Marathi language, has demonstrated a remarkable recognition rate of 89.77 %. The successful establishment of a standard database for Marathi speech paves the way for future developments in speech recognition technology for this language. In addition, the potential expansion of this system to larger vocabularies in Interactive Voice Response Systems (IVRS) for disease words speech recognition purposes offers promising prospects for rural community improvement. This study contributes substantially to the advancement of Marathi speech recognition and its potential applications in a variety of domains.

References:

[1] Somnath Hase; Sunil Nimbhore "Speech Recognition: A Concise Significance" https://ieeexplore.ieee.org/document/9697255/authors#authors

[2] Siddharth S More1, Prashant kumar L. Borde, Sunil S Nimbhore "Isolated Pali Word (IPW) Feature Extraction using MFCC & KNN Based on ASR" Volume 20, Issue 6, Ver. II (Nov - Dec 2018), PP 69-74

[3] Siddharth S More1, Prashant kumar L. Borde, Sunil S Nimbhore "A Review on Automatic Speech Recognition System in Indian Regional Languages" International Journal of Computer Applications (0975 –8887) Volume 181 –No.4, July 2018 38

[4] Sunil S. Nimbhore, Ghanshyam Digambar Ramteke Rakesh Ramteke "Pitch estimation of Marathi spoken numbers in various speech signals"

[5] L. R. Bahl, P. V. de Souza, R. L. Mercer, M. A. Picheny, and P. F. Brown, "A Method for the Construction of Acoustic Markov Models for Words," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 4, pp. 443–452, 1993, doi: 10.1109/89.242490.

[6] S. Basak *et al.*, "Challenges and Limitations in Speech Recognition Technology: A Critical Review of Speech Signal Processing Algorithms, Tools and Systems," *Comput. Model. Eng. Sci.*, vol. 135, no. 2, pp. 1053–1089, Oct. 2022, doi: 10.32604/CMES.2022.021755.

[7] Aryana and S. J. Zafarmand, "GLOCAL PRODUCT DESIGN: A SUSTAINABLE SOLUTION FOR GLOBAL COMPANIES IN REGIONAL AND / OR LOCAL MARKETS," *IASDR07*, pp. 1–18, 2007.

[8] Valmlfdamir Zwass, "Speech recognition | Voice Recognition, AI & Machine Learning | Britannica." https://www.britannica.com/technology/speech-recognition (accessed Jul. 27, 2023).

[9] HTK, "HTK Speech Recognition Toolkit." https://htk.eng.cam.ac.uk/ (accessed Jul. 27, 2023).

[10] Zeidan., "Marathi language | Definition, History, Alphabet, & Facts | Britannica." https://www.britannica.com/topic/Marathi-language (accessed Jul. 27, 2023).

[11] L. Lewandowski, W. Wood, and L. A. Miller, "Technological Applications for Individuals with Learning Disabilities and ADHD," *Comput. Web-Based Innov. Psychol. Spec. Educ. Heal.*, pp. 61–93, 2016, doi: 10.1016/B978-0-12-802075-3.00003-6.

[12] 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.

[13] K. R. Ghule and R. R. Deshmukh, "Automatic Speech Recognition of Marathi isolated words using Neural Network," vol. 6, no. 5, pp. 4296–4298, 2015.

[14] S. Kulkarni, R. R. Deshmukh, V. L. J. Patil, P. Pukhraj, S. D. Waghmare, and A. M. Oirere, "Marathi Isolated Digit Recognition System using HTK," pp. 42–45, 2016.

[15] 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.

[16] S. Bhable, "Automatic Speech Recognition (ASR) of Isolated Words in Hindi low resource Language," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. 2, pp. 260–265, 2021, doi: 10.22214/ijraset.2021.33011.