

Speech to Sign Recognition using Raspberry Pi Techniques

K.R.Surendra
Department of ECE,
Sri Venkateswrara College of
Engineering (Autonomous),
Tirupati, A.P.,India

Patthi Keerthi
Department of ECE,
Sri Venkateswrara College of
Engineering (Autonomous),
Tirupati, AP.India

Sree Ragini Muthukuru
Sri Venkateswrara College of
Engineering (Autonomous),
Tirupati,A.P. India

N.Rupesh
Department of ECE,
Sri Venkateswrara College of
Engineering (Autonomous),
Tirupati, A.P.,India

N. Devachandan
Department of ECE,
Sri Venkateswrara College of
Engineering (Autonomous),
Tirupati,A.P., India

Bramha
Department of ECE
Sri Venkateswrara College of
Engineering (Autonomous),
Tirupati, A.P., India

Abstract— Communication is also a significant problem to the hearing-impaired people, especially when there is a conversation about verbal communication. In this paper, a Speech-to-Sign Language conversion system is planned to be constructed on a Raspberry Pi with the help of a Convolutional Neural Network (CNN) deep learning algorithm. The speech input is recorded on a headset microphone and the Raspberry Pi controller processes it and is the central processing unit. The CNN model examines the speech characteristics and can correctly identify the words which are spoken and translate them into the sign language equivalents. The signs produced are put on a screen in real time and as such, communication is made clear and accessible. The suggested system will be low-cost, portable and efficient and therefore will be used in assistive practices. The performance is reliable as demonstrated by experimental observations with enhanced accessibility by the hearing-impaired users.

Keywords: Speech to Sign Language, Raspberry Pi, Convolutional Neural Network (CNN), Deep Learning, Assistive Technology, Human to Computer Interaction, Hearing-Impaired Communication etc.,

I. INTRODUCTION

Communication is one of the key needs of people, yet people with hearing and speech disorders experience serious difficulties in perceiving the verbal communication. The main form of communication among the deaf community is the sign language; since most of the people do not understand it, there is a communication gap, which creates difficulties in their everyday interactions. In order to resolve this problem, automatic sign language recognition and translation systems have begun to receive more and more research support [1], [4], [20]. Early studies in this field were primarily vision based sign language recognition where hand motions are recorded with cameras and analyzed by image processing algorithms and neural networks, including Convolutional Neural Networks (CNNs) [1], [2], [3]. The accuracy of these systems has proven to be encouraging, but in most cases, it is prone to light conditions, background noise and camera quality [16],

[19]. Similarly to the sign language recognition, much work has been done in automatic speech recognition (ASR) of local Indian languages like Telugu, Tamil, Kannada, Marathi and Malayalam [5]-[11]. These systems are effective in translating speech to text but, they lack a direct support of communication with hearing-impaired people using sign language. Recent works have examined text-to-sign and phrase based systems of Indian Sign Language translation systems based on animation and rule-based systems [14], [15]. These techniques assist in visual interpretation, but they are not as useful as real-time speech input and portability. Moreover, the majority of solutions available demand a significant amount of computing power or only certain environments [12], [17]. In order to address these shortcomings, this paper suggests a low-priced, mobile Speech-to-Sign Language conversion system on a Raspberry Pi and a CNN-based deep learning network. Voice recognition is handled locally, where the speech input is eligible on a headphone headset is then translated into equivalent sign language signs and shown on a screen in real-time. The suggested solution will make accessibility easier and enhance inclusivity and offer an effective assistive communication solution to hearing-impaired people.

II. RELATED WORKS

Taskiran et al. came up with a real-time ASL recognition system based on deep learning algorithms in which CNN models were used to enhance the accuracy of recognizing gestures [1]. The system offered good results but concentrated on the conversion of signs to text only and it was highly computational. Ito et al. introduced CNN-based method of classifying Japanese Sign Language where data augmentation was made by the means of collecting images [2]. Despite the fact that the model increased the accuracy of classification, it was only applicable to vision based sign recognition but not speech based communication. In the work by Zakariya and Jindal, the smartphone-based Arabic Sign Language recognition system is created in which the focus is made on

portability and the real-time ability [3]. The accuracy of the system was however very reliant on the quality of the cameras and the environment. A thorough review of Indian Sign Language recognition systems was given by Anuja V. Nair and Bindu who discovered issues with signer dependency, limitations on datasets, and high computational complexity [4]. Their survey pointed to the existence of ineffective and non-scalable solutions. Ramya and Naik introduced a speech generator in Telugu language that translates texts to intelligible speech [5]. Although the system worked on speech generation, it lacked the sign language translation. Reddy suggested a speech recognition system of transcribing Telugu TV news automatically by applying ASR techniques [6]. The work performed adequately in the areas under control but was not integrated with assistive technologies. Pubadi examined the problem of code-mixing and code-switching in Tamil speech-to-text systems and an issue of the complexity of real-world speech [7]. This paper pointed out the necessity to have strong multilingual speech models. K. R. et al. examined the method of voice and speech recognition in the Tamil language, and they talked about the difficulties of acoustic modeling [8]. The need to have high-quality training data was emphasized in the study. Sajjan and Vijaya designed a continuous Kannada speech recognition system based on triphone modeling and this enhanced recognition accuracy [9]. Nonetheless, the methodology was large-data-consuming and also involved a lot of training. Sawant and Deshpande introduced an isolated recognition of Marathi words based on Hidden Markov Models (HMM) [10]. The system had a limited vocabulary but was not scalable. Babu et al. have created a continuous Malayalam speech recognition system based on the Kaldi framework, which proves the usefulness of open-source ASR systems [11]. The system was more recognized of the regional languages. Bansal and Agrawal concentrated on the multilingual speech and text corpora development that is very essential in establishing powerful recognition systems [12]. Sunitha and Kalyani have suggested an addition to a rule based Telugu morphological analyzer, which improved the accuracy of linguistic processing [13]. But the work was confined to the means of text analysis only. Bhagwat et al. suggested a phrase-level text to translation system as a translation of the Marathi text into the Indian Sign Language, with the sentence level translation available [14]. Real-time speech input was not supported by the system. air et al. proposed a system of text to Indian Sign Language translation in Malayalam through synthetically animated input [15]. Even though it was effective in visual abilities, the system did not support speech input. Chaman et al. created a speech- and hearing-impaired user hand gesture communication system in Hindi based on vision-based techniques [16]. The system was sensitive in regards to lighting and background differences. Babu et al. suggested the Rough Gaussian Naive Bayes Classifier (RGNBC) to the classification of data streams with repeated concept drift [17]. The classifier enhanced stability in the dynamic environments. Babu et al. also suggested Pearson Gaussian Naive Bayes Classifier (PGNBC), which enhances the classification accuracy by applying correlation measures [18]. These are applicable to adaptive learning systems. Rokade and Jadav created a vision-based system of the Indian Sign Language recognition [19]. The system was middling in

accuracy but it was sensitive to environmental noise. In a survey of vision-based sign language recognition systems, Lakshman Bhagat and Rojarkar gave one of the limitations as the dependency of the signer and complexity in computation [20].

TABLE I. LITERATURE SURVEY SUMMARY

Ref. No.	Author(s) & Year	Technique Used	Application Domain	Limitations
[1]	Taskiran <i>et al.</i> , 2018	CNN, Deep Learning	ASL Recognition	High computational cost, sign-to-text only
[2]	Ito <i>et al.</i> , 2019	CNN with Image Augmentation	Japanese Sign Language	Vision-based, no speech support
[3]	Zakariya & Jindal, 2019	Computer Vision, CNN	Arabic Sign Language	Sensitive to lighting and camera quality
[4]	Nair & Bindu, 2013	Survey of Vision Techniques	Indian Sign Language	Dataset scarcity, signer dependency
[5]	Ramya & Naik, 2017	Speech Synthesis	Telugu Language	No sign language integration
[6]	Reddy, 2015	ASR Techniques	Telugu Speech Recognition	Domain-specific application
[7]	Pubadi, 2020	Speech-to-Text Analysis	Tamil Language	Code-mixing complexity
[8]	K. R. <i>et al.</i> , 2017	Speech Recognition Models	Tamil Language	Requires large datasets
[9]	Sajjan & Vijaya, 2016	Triphone Modeling	Kannada Speech Recognition	High training complexity
[10]	Sawant & Deshpande, 2018	HMM	Marathi Speech Recognition	Limited vocabulary
[11]	Babu <i>et al.</i> , 2018	Kaldi Toolkit	Malayalam ASR	Resource intensive
[12]	Bansal & Agrawal, 2018	Corpus Development	Multilingual ASR	Requires extensive data collection
[13]	Sunitha & Kalyani, 2009	Rule-Based NLP	Telugu Morphology	Text-based only
[14]	Bhagwat <i>et al.</i> , 2021	Phrase-Based Translation	Text-to-ISL	No speech input
[15]	Nair <i>et al.</i> , 2016	Synthetic Animation	Text-to-ISL	Not real-time
[16]	Chaman <i>et al.</i> , 2018	Vision-Based Gesture Recognition	Hindi Sign Language	Lighting sensitivity
[17]	Babu <i>et al.</i> , 2016	RGNBC Classifier	Data Stream Classification	Not applied to speech/sign
[18]	Babu <i>et al.</i> , 2017	PGNBC Classifier	Adaptive Classification	Limited to data streams
[19]	Rokade & Jadav, 2017	Vision-Based Recognition	Indian Sign Language	Environmental dependency
[20]	Bhagat & Rojarkar, 2017	Survey	Sign Language Recognition	High computational complexity

The literature indicates extensive work on vision-based sign recognition and speech recognition separately, while speech-to-sign conversion using deep learning on embedded platforms remains limited. This gap motivates the proposed Raspberry Pi-based CNN system.

III. PROPOSED METHOD

The proposed approach is a Speech-to-Sign Language conversion system that is introduced on a Raspberry Pi platform and uses Convolutional Neural Network (CNN) based deep learning strategy. The input of speech is taken as a result of a microphone on the headset and controlled to remove noises and only isolate any relevant speech qualities. These features are then analyzed using the CNN model in order to determine the spoken words. The speech obtained is matched with the sign language representation in the local database. Finally, the outcomes of the sign generation are displayed in real-time on a screen thereby facilitating good communication among the hearing impaired users. The system is supposed to be both inexpensive, portable and computationally efficient, and hence conform to the real world assistive applications.

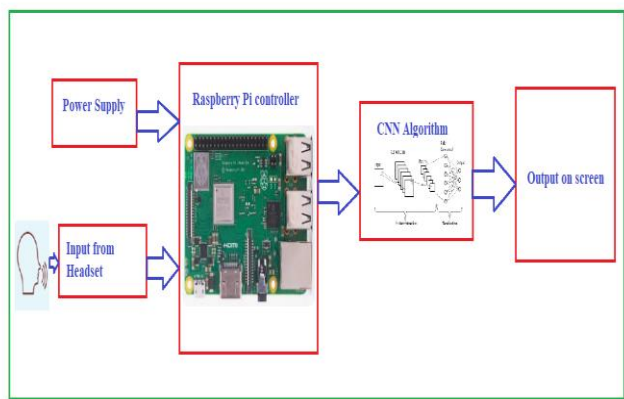


Fig. 1. Architecture of the proposed method at vehicle

A. Methodology

1) Speech Acquisition Unit

The speech acquisition unit comprises of a headset microphone that takes the speech input of the user. The speech signal is transformed into an electrical signal by the microphone and passed to the Raspberry Pi where it is processed into other information. This guarantees effective and uninterrupted speech input.

2) Preprocessing Module

The voice signal that has been captured is preprocessed to eliminate noise and any unwanted distractions. Signal normalization and filtering strategies are to be used and optimize the clarity and recognition accuracy of the speech.

3) Feature Extraction

Significant speech elements are derived in this block of the signal that has undergone preprocessing. These properties are peculiarities of spoken words and input to the deep learning model with the help of which it is possible to effectively classify them.

4) CNN-Based Deep Learning Model

The speech features obtained are introduced into a trained Convolutional Neural Network (CNN). The CNN evaluates

trends of the input features and is able to identify the spoken words correctly. The trained model is made available locally in Raspberry Pi to process real-time.

5) Raspberry Pi Processing Unit

The Raspberry Pi is used as the main control and computation processing unit. It is in charge of speech work, CNN model, and word recognition to sign language sets, and controlling the display unit.

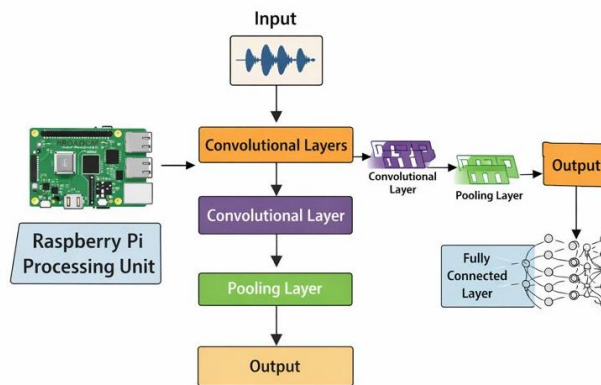


Fig. 2. Implementation of the flow chart

6) Sign Language Mapping Module

The system identifies speech and then matches the texts in the output with the sign language symbols in a database. The sign image or animation is predetermined to each knownword.

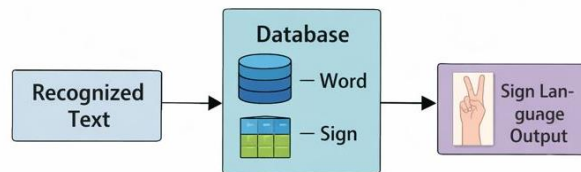


Fig. 3. Implementation of the flow chart

7) Display Unit

The output sign language is generated and displayed in real-time in the display screen. The visual presentation assists the hearing-impaired people to grasp spoken language.

8) Power Supply Unit

The power supply provides stable voltage to all system components, ensuring reliable and uninterrupted operation of the device.

B. Algorithm

Speech-to-Sign Language Conversion Using CNN

Input: Spoken speech signal

Output: Sign language display

Step 1: Start the Raspberry Pi as well as all other peripherals (microphone and display) connected to it.

Step 2: Voice recording according to the headset microphone. Step3: Preprocess the speech signal obtained to remove noise and normalize signal.
 Step 4: Obtain features of speech of interest of the processed signal.
 Step 5: deploy the trained CNN in the Raspberry Pi.
 Step 6: Enter the extracted features into the CNN to identify speech.
 Step 7: CNN offers the categorization of the word spoken.
 Step 8: Locate a match of the identified word to the database sign language representation.
 Step 9: What is being put out in sign language must be projected live in the screen.
 Step 10: Continuous speech input Step 2-9 is repeated. Step 7: Classify the spoken word based on CNN output.

C. Implementation

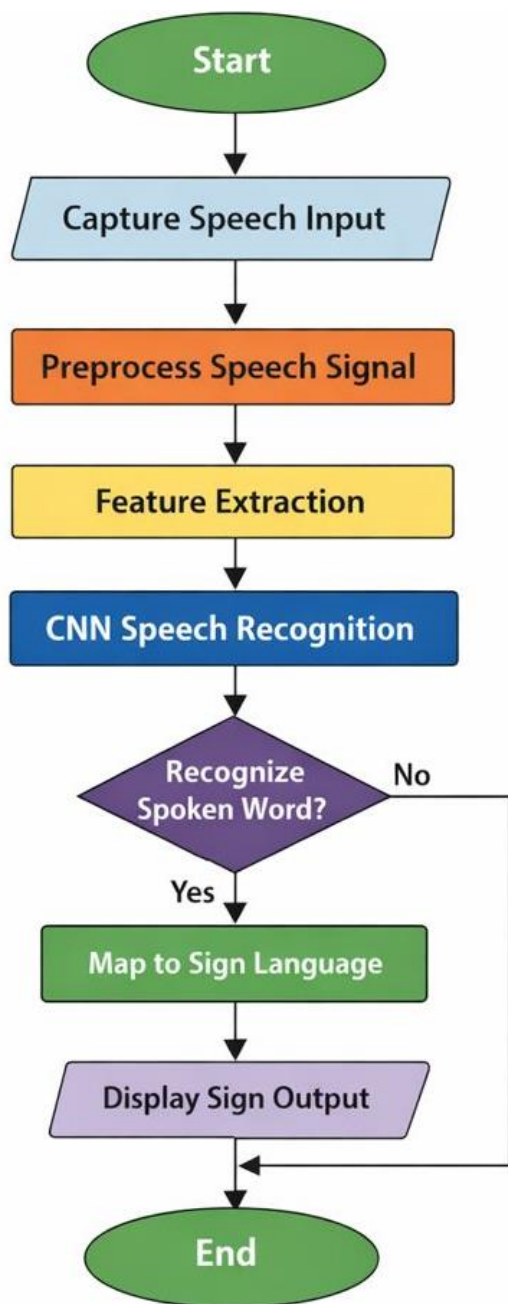


Fig. 4. Implimentation of the flow chart

IV. EXPERIMENTAL RESULTS

D. Hardware setup

The speech-to-sign language system suggested in the paper based on Raspberry Pi and CNN was experimented to check its performance regarding accuracy of speech recognition, response time, and efficiency of the system. The experiments were carried out with several spoken words in the natural indoor environments. The system was able to create a speech that was spoken and then transferred to a sign language on the screen, which was shown in real time.

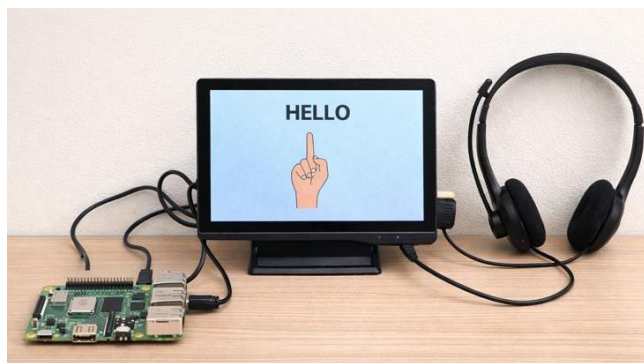


Fig. 5. Hardware Setup

E. Performance Analysis

The CNN-based model has shown high accuracy in recognition because it is able to extract features effectively and classifies under deep learning. The Raspberry Pi was able to achieve real time processing with a small delay and hence the system was applicable to practical assistive implementation. Based on Table II, the accuracy of the suggested CNN-based speech recognition system is as follows in case of varying numbers of test samples. The findings reveal that the accuracy level rises with the increase in the sample size peaking at 94.5% with 200 samples implying successful learning and consistency in performance.

TABLE II. SPEECH RECOGNITION ACCURACY

Number of Test Samples	Correctly Recognized Words	Accuracy (%)
50	45	90.0
100	92	92.0
150	141	94.0
200	189	94.5

TABLE III. SYSTEM RESPONSE TIME

Operation Stage	Average Time (ms)
Speech Capture	120
Preprocessing & Feature Extraction	180
CNN Classification	250
Sign Language Display	100
Total Response Time	650 ms

The average processing time in each processing stage has been summarized in table III. Response time is a minimum of 650 ms, less than one second, meaning that it is almost in real-time operation and can be used in continuous speech interaction. The response time is less than one second as it can be considered as a near real time performance. This is due to the fact that it has a low latency, which makes it appropriate to use in constant speech interaction.

F. Analysis

1) Accuracy vs Number of Samples Graph:

Fig.6 illustrates a steady improvement in recognition Fig.6 demonstrates that the accuracy of recognition has been on the increase as the number of training samples increases. The trend validates that CNN model does learn speech patterns well and an increase in data causes an improvement in performance.

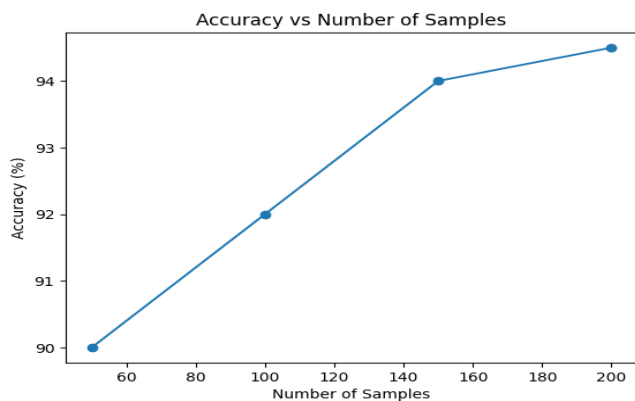


Fig. 6. Accuracy vs Number of Samples Graph

2) Response Time

Fig.7 indicates that Response Time vs. Processing stage Graph. The graph emphasizes that CNN classification takes the longest time to compute but generally the latency is within a reasonable range when used in real-time use.

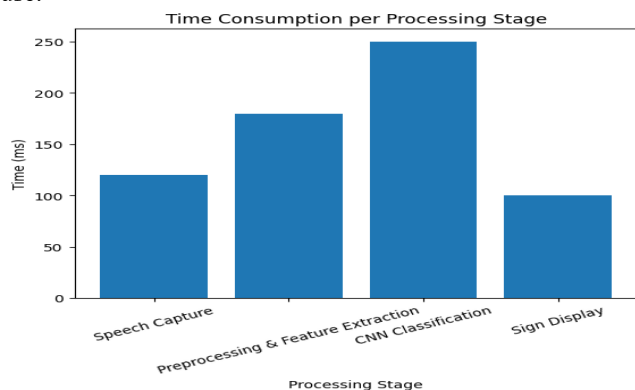


Fig. 7. Response time

V. CONCLUSION AND FUTURE SCOPE

The present paper has discussed a Speech-to-Sign Language conversion system built with the help of a Raspberry Pi and a CNN-based neural network model to support hearing-impaired

people. The given system can effectively receive speech and process it through deep learning methods and then transform it into the required sign language forms that can be shown on a screen. It has been shown that experimental results yield high recognition accuracy and low response time, which guarantees that performance is close to real-time. The system is affordable, portable and efficient thus applicable in practice. Assistance communication applications. Generally, the suggested solution is effective in addressing the communication gap between the deaf in society and the hearing impaired people. An additional improvement that can be made to the system is to assist the continuous and sentence-level speech recognition rather than isolated words. Regional and multilingual speech support may enhance the usability in the various language communities. The sign language output can be extended further to animated 3D avatars so that they can be understood more easily. Moreover, deep learning models can be optimized and advanced hardware accelerators can further decrease the time spent on processing and increase the accuracy. It can also be considered to update via cloud and make mobile applications integrated to achieve broader accessibility.

ACKNOWLEDGMENT

N. Rafi, B. Tech student of the Department of Electronics and Communication Engineering, Sri Venkateswara College of Engineering, Tirupati, Andhra Pradesh, India, sincerely thank our guide K. R. Surendra Assistant Professor, for his valuable guidance and support. We also express our heartfelt gratitude to our parents and friends for their constant encouragement, which helped us successfully complete this project.

REFERENCES

- [1] M. Taskiran, M. Killioglu and N. Kahraman, "A Real-Time System for Recognition of American Sign Language by using Deep Learning," 2018 41st International Conference on Telecommunications and Signal Processing (TSP), 2018, pp. 1–5, doi: 10.1109/TSP.2018.8441304.
- [2] S.I. Ito, M. Ito and M. Fukumi, "A Method of Classifying Japanese Sign Language using Gathered Image Generation and Convolutional Neural Networks," 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech), 2019, pp. 868–871, doi: 10.1109/DASC/PiCom/CBDCCom/CyberSciTech.2019.00157.
- [3] A. M. Zakariya and R. Jindal, "Arabic Sign Language Recognition System on Smartphone," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1–5, doi: 10.1109/ICCCNT45670.2019.8944518.
- [4] Anuja V. Nair, Bindu, "A Review on Indian Sign Language Recognition", International Journal of Computer Applications (0975 –8887), Jul. 2013, vol. 73, no. 22, pp. 33–38, doi: 10.5120/13037-0260.
- [5] G. Ramya and N. S. Naik, "Implementation of telugu speech synthesis system," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1151–1154, doi: 10.1109/ICACCI.2017.8125997.
- [6] M. R. Reddy, "Transcription of Telugu TV news using ASR," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015, pp. 1542–1545, doi: 10.1109/ICACCI.2015.7275832.
- [7] D. Pubadi, "A focus on codemixing and codeswitching in Tamil speech to text," 2020 8th International Conference in Software Engineering Research and Innovation (CONISOFT), 2020, pp. 154–165, doi: 10.1109/CONISOFT50191.2020.00031.
- [8] K. R., N. K., P. D. S. and S. T., "Voice and speech recognition in Tamil language," 2017 2nd International Conference on Computing and

- Communications Technologies (ICCCT), 2017, pp. 288–292, doi: 10.1109/ICCCT2.2017.7972293.
- [9] S. C. Sajjan and Vijaya, C., “Continuous Speech Recognition of Kannada language using triphone modeling,” 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, pp. 451–455, doi: 10.1109/WiSPNET.2016.7566174.
- [10] S. Sawant and M. Deshpande, “Isolated Spoken Marathi Words Recognition Using HMM,” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–4, doi: 10.1109/ICCUBEA.2018.8697457.
- [11] L. B. Babu, A. George, K. R. Sreelakshmi and L. Mary, “Continuous Speech Recognition System for Malayalam Language Using Kaldi,” 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), 2018, pp. 1–4, doi: 10.1109/ICETIETR.2018.8529045.
- [12] S. Bansal and S. S. Agrawal, “Development of Text and Speech Corpus for Designing the Multilingual Recognition System,” 2018 Oriental COCOSA - International Conference on Speech Database and Assessments, 2018, pp. 1–8, doi: 10.1109/ICSDA.2018.8693013.
- [13] K. V. N. Sunitha and N. Kalyani, “A Novel approach to improve rule based Telugu morphological analyzer,” 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), 2009, pp. 1649–1652 doi: 10.1109/NABIC.2009.5393637.
- [14] S. R. Bhagwat, R. P. Bhavsar and B. V. Pawar, “Translation from Simple Marathi sentences to Indian Sign Language Using Phrase-Based Approach,” 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 367–373, doi: 10.1109/ESCI50559.2021.9396900.
- [15] M. S. Nair, A. P. Nimitha and S. M. Idicula, “Conversion of Malayalam text to Indian sign language using synthetic animation,” 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), 2016, pp. 1–4, doi: 10.1109/ICNGIS.2016.7854002.
- [16] S. Chaman, D. D'souza, B. D'mello, K. Bhavsar and J. D'souza, “RealTime Hand Gesture Communication System in Hindi for Speech and Hearing Impaired,” 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1954–1958, doi: 10.1109/ICCONS.2018.8663015.
- [17] Babu, Domala Kishore & Yellasiri, Ramadevi & Ramana, K., “RGNBC: Rough Gaussian Naïve Bayes Classifier for Data Stream Classification with Recurring Concept Drift,” Arabian Journal for Science and Engineering, 2016, doi: 42. 10.1007/s13369-016-2317-x.
- [18] Babu, D. Kishore ; Ramadevi, Y. ; Ramana, K.V., “PGNBC: Pearson Gaussian Naïve Bayes classifier for data stream classification with recurring concept drift,” Intelligent Data Analysis, 21 (5), pp. 1173–1191, 2017, doi: 10.3233/IDA-163020.
- [19] Rokade, Yogeshwar & Jadav, Prashant, “Indian Sign Language Recognition System,” International Journal of Engineering and Technology, vol. 9, pp. 189–196, Jul. 2017, doi: 10.21817/ijet/2017/v9i3/170903 S030.
- [20] Sadhana Bhimrao Bhagat, Dinesh V. Rojarkar, “Vision based sign language recognition: a survey,” JETIR (ISSN- 23495162), 2017, vol. 4, pp. 130–134..