

Intelligent System for Translating Text and Audio into Indian Sign Language Gestures

Malapriya S¹ and Dr Manju N²

¹Research Scholar, JSS STU, Mysuru

²Associate Professor, JSS STU, Mysuru

Abstract— The Indian Sign Language Recognition System (ISLRS) is designed to address communication barriers between hearing-impaired individuals and the broader community through precise sign language interpretation. In India, approximately 26.8 million people experience some form of hearing impairment, highlighting the importance of effective communication solutions for the deaf population. The Indian Sign Language Recognition System employs innovative technology to enhance interaction for deaf individuals, utilizing a Speech-to-Sign Translation system. This involves a multi-step process that includes recording spoken words, transcribing them into text format, and then mapping this text to corresponding ISL gestures through animations using Convolution Neural Network. This approach enables the communication between the deaf and hearing population. By adopting standardized Indian Sign Language signs, the system makes communication easier, leading to better understanding and interaction.

Keywords—Sign Language, Natural Language Processing, Audio, Text, CNN.

I. INTRODUCTION

Sign language plays important role in communication among the hearing impairment community using face, hands, and eye motions in lieu of vocal communication. Sign language recognition tool has the major function to decipher the signals in order to allow the communication between the mute and deaf population. Indian Sign Language is a sign language that is employed by the deaf population of India as a means of communication. It makes use of hand signs, body gestures and facial expressions, to express meaning and emotions. Unlike spoken languages, ISL does not depend on regional spoken dialects but provides a standard way of communication for the hearing-impaired. Considered as a major factor in empowering the deaf community, ISL is an important factor in educational, social, and work-related interactions. However, the lack of widespread understanding of ISL among the general population creates communication barriers, highlighting the need for accessible tools and technologies to bridge this gap [1].

The proposed system bridges the communication gap by converting speech into text and subsequently displaying corresponding ISL images. Speech input is recognized using the Google Speech API, processed through Natural Language Processing (NLP), and translated using a dictionary-based machine translation approach. This application aims to empower the deaf community by enabling seamless daily interactions and access to information. The scope of this research focuses on developing an automated system to enhance communication and learning for the deaf community. This paper aims to provide comprehensive learning materials and interactive tools tailored to enhance language skills among hearing-impaired people.

By integrating visual aids, interactive exercises, and Indian Sign Language (ISL) videos, the system is designed to facilitate a deeper understanding of basic language concepts. The primary objective of this research paper is to develop an automated system for converting text or audio into Indian Sign Language gestures.

II. RELATED WORKS

The author presents a system designed by considering voice as an input into Indian Sign Language (ISL), solving the lack of text-to sign conversion tools and sign language dictionaries. The system translates voice input into English text, then parses sentences, by applying ISL grammar, stems words, and sends them to a video dictionary of ISL signs. It highlights the significance of sign language in bridging the communication gaps in India. Challenges such as emotive capture and processing ambiguous sentences still exist but the system delivers real-time operations and uses effectively human video representations [2].

The authors focused on D-Talk, an AI-powered application designed to assist individuals with hearing and speech impairments by interpreting hand gestures for effective communication. Utilizing machine learning techniques like skin color extraction and contour tracking, the system translates gestures into actionable commands, promoting accessibility through innovative technology [3].

The authors gave the review developments in sign language translation, exploring conventional computer vision and recent machine learning approaches. They present gesture recognition, natural language processing, and multi-modal integration while dealing with issues such as grammar variation, restricted datasets, and real-time system requirements [4].

The paper describes a system to convert regional language to Indian Sign Language (ISL) to help hearing-impaired individuals, with focus on its technical deployment. The system uses speech recognition technology to transcribe verbal speech to text. This content is processed employing NLP-based methods to translate it into ISL syntax and grammar in order to bridge the linguistic disparity between ISL and spoken languages. ISL gesture synthesis modules then produce ISL signs in most cases using pre-recorded gesture databases or 3D avatar animation, preserving the correctness of ISL signs [5]. The papers concentrated to enable communication for the hearing impaired by translating spoken words into Indian Sign Language (ISL). The system relies on voice input to transcribe speech into text that is subsequently mapped to ISL signs via graphics or GIFs. It incorporates Natural Language Processing (NLP) to guarantee accurate interpretation [6]. The researchers use a carefully selected dataset of high-quality YouTube videos, featuring strong speech and facial expressions, to train a conditional Generative Adversarial Network (GAN).

This model, called WAV2PIX, generates face images directly from raw voice signals, such as WAV files, by learning to map voice features to corresponding facial expressions. This approach demonstrates a novel application of GANs in generating realistic facial visuals from audio inputs [7].

The authors concentrated on employing the use of a hardware glove with flex sensors, but it was not accurate enough. To enhance it, it was trained on a diversified data set using a convolutional neural network model, which made it more efficient and accurate in detecting an assortment of sign language hand movements. The system is bidirectional, both converting speeches to sign language and vice versa. It employs text conversion as an intermediary to mediate between communications, with scope for special gesture additions and language-to-language expansion [8].

This paper introduces a system that translates English speech into Indian Sign Language (ES2ISL) animations to enhance communication for hearing-impaired people in India. Through the combination of Natural Language Processing (NLP), the Google Cloud Speech API, and a pre-defined sign language database, the system is 77% accurate and processes input within 0.85 seconds [9].

The author proposes an Android application that converts speech into Indonesian Sign Language using morphological analysis and a modified Enhanced Confix Stripping (ECS) stemming algorithm. The system integrates the Android Speech API for speech recognition, achieving 94.06% accuracy, while the morphological analysis method reaches 95% accuracy [10].

The author utilizes a hybrid approach combining machine learning and single-shot detection with convolutional neural networks to recognize and translate Ethiopian Sign Language into speech or text [11]. This paper proposes a deep neural network model for continuous sign language recognition, with a 3D residual convolutional network (3D-ResNet) employed to extract features and a stacked dilated convolutional network along with Connectionist Temporal Classification (CTC) to learn sequences. To handle training challenges caused by CTC loss, a process of iterative optimization is adopted. This process alternatively fine-tunes the 3D-ResNet on pseudo-labels computed by the sequence model and optimizes both the parts. The method shows success on the RWTH-PHOENIX- Weather benchmark for continuous sign language recognition [12].

III. METHODOLOGY

The following figure illustrates the proposed system by splitting the into two parts. Initially, it takes an audio as an input and converts into sign language, displaying the gif file as output. Sure! Here's a clearer and more polished version of that sentence. The Graphical User Interface is implemented to accept audio input through a microphone, which is then converted to text. Each obtained words are taken as text. The generated text is divided and saved in the dictionary file. Text Pre-processing is done by natural language processing and uses dictionary based Machine Translations of input text by using ISL grammar rules which helps in Sign language generation with GIF file.

In the second stage, the goal is to convert images and videos to text or speech. It entails gathering Indian Sign Language datasets specific to helping hearing-impaired individuals. The datasets are pre-processed through resizing, removal of noise, and grayscale conversion, after which they are divided into training and testing subsets. A Convolutional Neural Network model is used to train the model. The trained model handles video inputs, detects the user's hand gestures, and translates the resultant sign language.

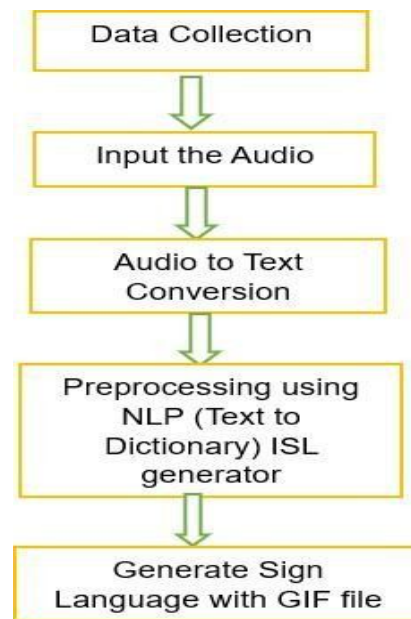


Fig 1: Conversion of Audio File to Sign Language

The algorithm used in the proposed model begins by capturing audio input, either through a microphone or by loading a pre-recorded audio file. The audio is then processed by converting it into text using a Speech-to-Text (STT) engine, such as the Google Speech API. This step involves normalizing the text by removing punctuation and converting it to lowercase to ensure consistency. The next step involves converting the text to sign language gestures, where the text is tokenized into individual words or phrases. Each token is then matched to its corresponding Indian Sign Language (ISL) gesture using a dataset that links words or phrases with ISL signs. In the subsequent stage, the system retrieves the corresponding ISL gesture for each matched word or phrase, facilitating the animation of the gestures. Lastly, the ISL gestures are shown sequentially on the screen as a visual output, giving a clear image of the words uttered in Indian Sign Language. In this process, Convolutional Neural Networks (CNNs) may be utilized to improve gesture recognition in instances where video- based sign language interpretation is needed, by processing image sequences of gestures and enhancing accuracy using deep learning models.

A Convolutional Neural Network (CNN) basically contains four layers: Convolutional, ReLU, Pooling, and Fully Connected layers. The Convolutional Layer applies filters on the input to detect features by replacing negative values with zeros in order to create a feature map. The ReLU (Rectified Linear Unit) Layer only activates nodes when the input is greater than a certain threshold, providing a linear relationship with the dependent variable for positive inputs and maintaining outputs as zero for negative inputs. The Pooling Layer decreases the image stack size through downsampling, which means picking a window size (usually 2x2 or 3x3), selecting a stride (ordinarily 2), moving the window over the feature map, and taking the maximum value from each window. This simplifies the data without compromising the important features. Lastly, the Fully Connected

Layer links all neurons from earlier layers to each neuron in later layers, emulating high-level thinking by analyzing all possible input-output routes. After going through the Convolutional, ReLU, and Pooling layers, the reduced and processed data is flattened into one list or vector, making classification or prediction operations in the final output possible. This design enables CNNs to efficiently extract and process features for intricate tasks.

IV. RESULT AND DISCUSSIONS

The Fig 2 represents the snapshot of translating the users input text through mic to convert into sign language.



Fig 2: Text to Sign Language Conversion

The Fig 3 snapshot represents the sign language representation for the given text “Hello” by the user’s input text through mic.

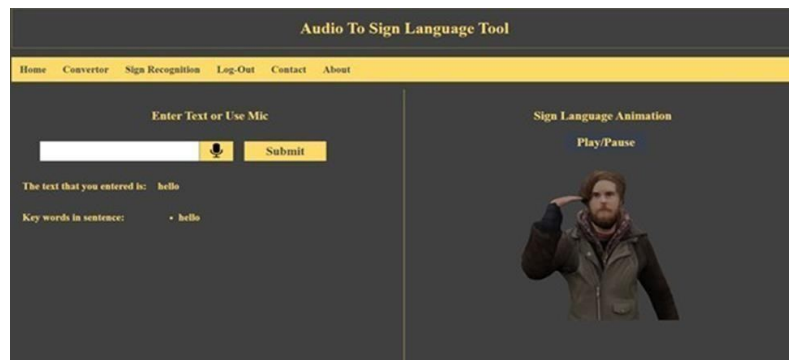


Fig 3: Output in Indian Sign Language

The Fig 4 snapshot represents the sign language representation for the given text ”Good” by the user’s input text through mic.

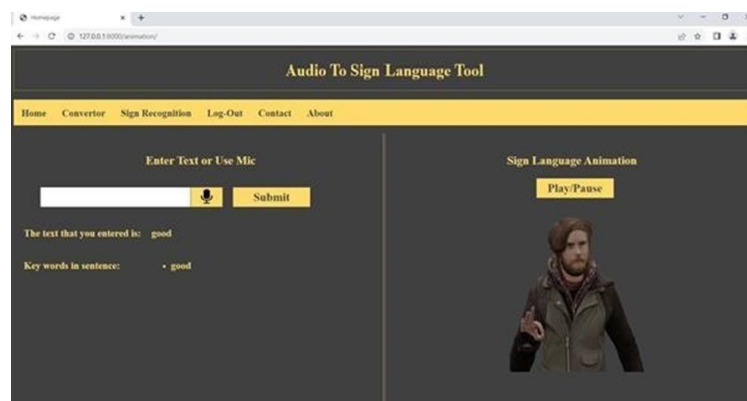


Fig 4: Output in Indian Sign Language

Fig 5 snapshot represents the conversion of sign language by reading the users input image to the text.

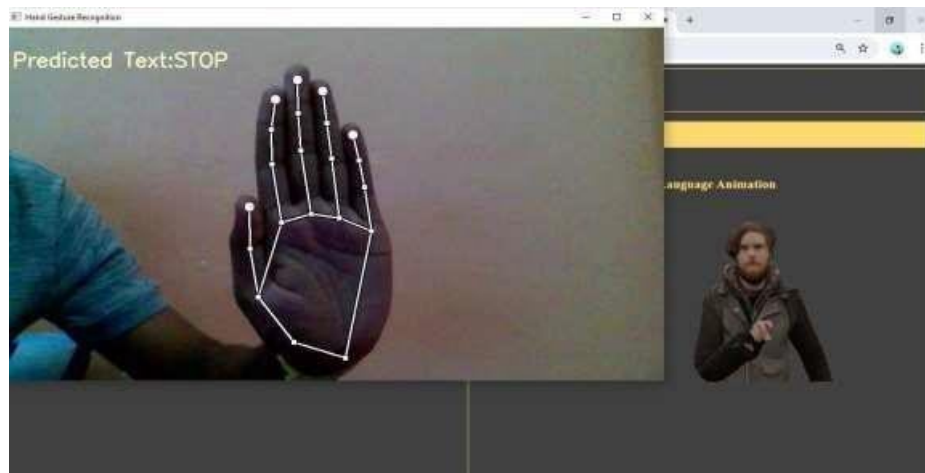


Fig 5: Predicted Text Results

V. CONCLUSION

The proposed web application is designed to automate the translation process that translates incoming speech or English text into phrase structure grammar form. This form is then translated to the grammatical structure of Indian Sign Language (ISL). To achieve accurate translation, the system removes stop-words from the input and uses stemming and lemmatization to convert words to their root forms, as ISL does not have word inflections. The system matches each word with its corresponding sign from a dictionary containing representations of words. If a word is not found in the database, the algorithm searches for synonyms and replaces the word with its equivalent, ensuring effective communication.

REFERENCES

- [1] Shetty, S., Hirani, E., Singh, A., & Koshy, R. (2024). Gesture-to-Text: A Real-Time Indian Sign Language Translator with Pose Estimation and LSTMs. *Procedia Computer Science*, 235, 2684-2692.
- [2] Thakar, S., Shah, S., Shah, B., & Nimkar, A. V. (2022, October). Sign Language to Text Conversion in Real Time using Transfer Learning. In *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)* (pp. 1-5). IEEE.
- [3] Papatsimouli, M., Kollias, K. F., Lazaridis, L., Maraslidis, G., Michailidis, H., Sarigiannidis, P., & Fragulis, G. F. (2022, June). Real Time Sign Language Translation Systems: A review study. In *2022 11th International Conference on Modern Circuits and Systems Technologies (MOCASST)* (pp. 1-4). IEEE.
- [4] Liang, Z., Li, H., & Chai, J. (2023). Sign language translation: A survey of approaches and techniques. *Electronics*, 12(12), 2678.
- [5] Garg, C., Attry, G. P., Mittal, D., & Bhat, A. (2023, May). Conversion of Native Speech into Indian Sign Language to Facilitate Hearing Impairment. In *2023 4th International Conference for Emerging Technology (INCET)* (pp. 1-9). IEEE.
- [6] Nayana, J., Bhat, S., Nair, R. R., & Babu, T. (2022, December). Audio to Sign Language conversion using Natural Language Processing. In *2022 3rd International Conference on Communication, Computing and Industry 4.0 (C2I4)* (pp. 1-6). IEEE.
- [7] Duarte, A. C., Roldan, F., Tubau, M., Ecur, J., Pascual, S., Salvador, A., ... & Giro-i-Nieto, X. (2019, May). WAV2PIX: Speech-conditioned Face Generation using Generative Adversarial Networks. In *ICASSP* (Vol. 2019, pp. 8633-8637).
- [8] Fernandes, L., Dalvi, P., Junnarkar, A., & Bansode, M. (2020, August). Convolutional neural network based bidirectional sign language translation system. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 769-775). IEEE.
- [9] Patel, B. D., Patel, H. B., Khanvilkar, M. A., Patel, N. R., & Akilan, T. (2020, August). ES2ISL: an advancement in speech to sign language translation using 3D avatar animator. In *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1-5). IEEE.
- [10] Baehaqi, M. N., Irzal, M., & Indiyah, F. H. (2019, October). Morphological analysis of speech translation into Indonesian sign language system (SIBI) on android platform. In *2019 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 205-210). IEEE.
- [11] Eshetu, Y., & Wolde, E. (2019). A Real-Time Ethiopian Sign Language to Audio Converter. *International Journal of Engineering Research and technology (IJERT)*, 8, 661-66.

- [12] Pu, J., Zhou, W., & Li, H. (2018, July). Dilated convolutional network with iterative optimization for continuous sign language recognition. In IJCAI(Vol. 3, p. 7).
- [13] Guo, D., Zhou, W., Li, H., & Wang, M. (2018, April). Hierarchical LSTM for sign language translation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).
- [14] Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017, July). Convolutional sequence to sequence learning. In International conference on machine learning (pp. 1243-1252). PMLR.
- [15] Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017, July). Convolutional sequence to sequence learning. In International conference on machine learning (pp. 1243-1252). PMLR.
- [16] Pu, J., Zhou, W., & Li, H. (2016). Sign language recognition with multi-modal features. In Advances in Multimedia Information Processing-PCM 2016: 17th Pacific-Rim Conference on Multimedia, Xi'an, China, September 15-16, 2016, Proceedings, Part II (pp. 252-261). Springer International Publishing.