Indian Sign Language Translation Using Deep Learning Techniques

Sunitha Sabbu, Sumaira Tanzeel A., Sravya K., Venkata Sai Kumar V. and Sai Charan Teja C.
Department of CSE(AI&ML), Srinivasa Ramanujan Institute of Technology, Rotarypuram Village, B K Samudram Mandal, Anantapur District-515701, Andhra Pradesh, India

Keywords: Indian Sign Language, Deep Learning, Text-to-Sign Conversion, Image-to-Sign Conversion, Speech-to-Sign Conversion.

Abstract: However, there exists a significant social barrier due to communication limitations even though Sign Language (SL) is the most viable form of expression for the deaf and mute community. We introduce a new technology that can yield communication with Text-to-Sign translation and Text, Speech, and images converted into Indian Sign Language outputs. This system caters explicitly to the Indian Sign Language community, a sector that has recently garnered comparatively less attention in terms of technical innovations, distinguishing it from competing scenarios. The proposed system is novel in that it will cater to the Indian Sign Language community, a widely under-served population group in terms of technological advancements. Our proposed approach has several features such as incorporating advanced Deep Learning (DL) techniques for the system to accurately identify hand gestures accurately for real-time recognition, using a dataset collected from Indian Sign Language Research and Training Centre, Kaggle in enhancing cultural specificity. It is a modular system that enables smooth integration into the different sectors of society, educational institutions, public service and healthcare centers. It forms a basic system that could lead on to more social inclusion throughout a variety of architecture, all of which could be extended in time. This is at the service of empowering the deaf-mute community and actively integrating this population into the social context.

1 INTRODUCTION

To about 70 million deaf people around the world, sign language is a crucial form of communication 5 million of whom are in India, where Indian Sign Language (ISL) is the predominant form in use. Though this is a significant step in narrowing that gap, there's still a communication divide between deaf people and hearing people much of that resulting from a lack of awareness and understanding of sign language. This communication gap affects several aspects of daily life including, accessing classrooms and health services as well as participating in social and professional activities. Technology based solutions exist for sign language interpretation, but primarily target the American sign language (ASL) or its international editions. This creates a huge gap in services tailored for Indian Sign Language users. The lack of ISL-based resources inhibits not just effective communication but also the deaf community's independence.

For Indian Sign Language, with its linguistic properties and cultural context, the availability and standardization of digital resources poses significant challenges. Compared to ASL and BSL, ISL has fewer datasets, documentation, and development tools available. This absence of resources is a major hurdle for academics and developers who are trying to build ISL-based solutions. Our project counteracts these limitations by utilizing datasets from the Indian Sign Language Research and Training Centre, while also supplementing the dataset with selective additional data found on Kaggle and paving the way for future ISL-based solutions by contributing to the growing number of ISL resources digitized.

Three independents but correlated categories are used in our approach to sign language conversion. The core part of our module processes text and maps it to the corresponding signs in text to text translation. Static images or animated GIFs of the signs are generated as the output of this process. This feature is enhanced in the speech-to-sign module, which employs speech recognition techniques to transcribe spoken words, passing them through the text-to-sign conversion process. Similarly, the image-to-sign capability utilizes optical character recognition (OCR) methods to recognize text from

Sabbu, S., A., S. T., K., S., V., V. S. K. and C., S. C. T.
Indian Sign Language Translation Using Deep Learning Techniques.
DOI: 10.5220/0013881500004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 2*, pages 279-286
ISBN: 978-989-758-777-1

input images, e.g., digital displays or pictures of written text, and passes the retrieved text to the same conversion mechanism used for text-to-signs. This approach of combining input modalities into an integrated yet modular design serves to optimize system usability and accessibility, while ensuring output consistency in the sign language output irrespective of the input registered type. By utilizing a single architecture for multiple insert formats, the system adapts to a variety of real-world communications scenarios by speeding the conversion process while maintaining correctness and reliability across several insert formats.

2 RELATED WORKS

Specifically, (Muhammad al-Qurishi, et.al, 2021) introduced a model that surveyed progress in sign language translation and recognition to cover comprehensive communication solutions. This review highlighted the deep learning methods for sign language identification while addressing issues like real-time processing, and contextual variability. The findings had an immediate influence on design considerations for sign recognition systems.

Transformers-based designs for handling sequential nature of sign language data were researched (Necati Cihan Camgoz, et.al) and proved their effectiveness, but the majority of the work was still centered on Western sign languages.

Real-time sign detection has been studied using advanced object detection frameworks (Shobhit Tyagi, et al, 2023). The study also focused on gesture-through-parts: the study of recognition of 55 different signs from Alphabets and Integers.

A detailed review on a variety of machine learning methods (A. Adeyanju, et al, 2021) assisted in narrowing down on deep learning techniques that led to robust recognition accuracy across diverse settings like varying ambient light and camera placements.

In the paper (Yogeshwar I. Rokade, Prashant M. Jadav, 2017) the problem of the recognition of Indian Sign Language (ISL) is discussed. The paper underscored the key takeaway of creating extensive datasets and ISL focused solutions — highlighting the necessity of individual-centric approaches for ISL recognition and translation.

Study investigated the differences in structure between ISL and spoken languages. In comparison, ISL does not have the equivalent of an auxiliary verb like "is" or "are" as in English. The English sentence, "The school opens in April," translates to ISL as "SCHOOL OPEN APR." ISL also employs fingerspelling, in which gestures mimic letters of the alphabet to spell names and technical terms. These structural differences require specialized models to be developed for ISL recognition.

Article (Sinha, et al, 2020) explored the broader implications of sign language recognition systems for accessibility, particularly how they can bridge communication gaps for the hearing-impaired community. The research highlighted the importance of user-centered customizable solutions to enhance inclusion in various contexts. Sign language translation efficiency was also suggested to be improved by combining deep learning and image processing methods according to the research. Note: Another similar work done by Sinha, Swapnil & Kataruka, Harsh & Kuppusamy, Vijayakumar (2020) entitled Image to Text Converter and Translator: Realization of Image to Text Converter and Translator using Deep Learning and Image Processing in the International Journal of Innovative Technology and Exploring Engineering. They explored how image-based text could be converted, forming an essential backdrop for implementation of sign language as well.

The work presented in paper focused on the impact of various machine learning models on improving the efficiency of sign language translation, specifically investigating strategies to optimize the translation model to enhance recognition accuracy while keeping the real-time user experience. The results underlined the importance of deep learning in speech to text applications and corresponds to artworks that subject a speech-driven sign language translation model. They correlated their findings with "Speech to Text using Deep Learning" (IJNRD, 2024): Speech is one of the methods of communication in which the person speaks, and the voice is converted into the text. One of the key highlights of this study was the acknowledgement of real-time processing and accuracy optimization in deep learning-enabled speech-to-text systems, which provides a significant step towards building an efficient sign language translation framework.

Perceived Usefulness (PU), Recognition Accuracy (RD), Perceived Ease of Use (PES), and Compatibility (CY) form perceived features of Indian Sign Language (ISL) translation systems. On the basis of these features the research suggests the following hypotheses:

The perceived features of Indian Sign Language (ISL) translation systems include Perceived Usefulness (PU), Recognition Accuracy (RD), Perceived Ease of Use (PES), and Compatibility

(CY). Based on these features, the research proposes the following hypotheses:

- Hypothesis 1 (H1**):** There is a positive relationship between the perceived usefulness (PU) of ISL recognition systems and the intention to use them.
- Hypothesis 2 (H2): There is a negative relationship between the perceived recognition difficulties (RD) of ISL recognition systems and the intention to use them.
- Hypothesis 3 (H3): The intention to use ISL recognition platforms will be positively connected to user attributes.
- Hypothesis 4 (H4**):** Interactive engagement with ISL content is positively related to the intention to employ ISL recognition technologies.
- Hypothesis 5 (H5**):** Interactive engagement with ISL content is negatively correlated with reliance on traditional communication methods.

3 METHODOLOGY

3.1 Theoretical Structure

The study uses an experimental design method, concentrating on the testing and deployment of a workable solution for translating Indian Sign Language into various formats. The system is developed incrementally, and its efficacy and dependability are guaranteed through iterative testing and validation stages. Figure 1 show the Schematic Flow of Theoretical Structure.

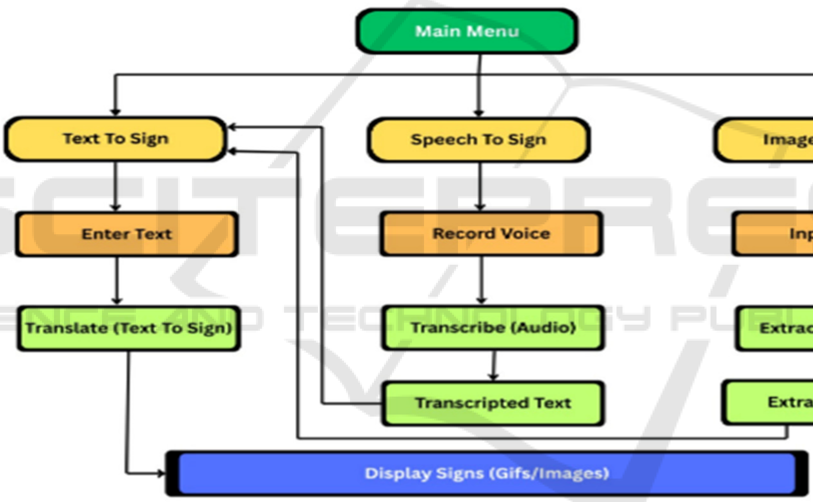

Figure 1: Schematic Flow of Theoretical Structure.

3.2 Objectives

3.2.1 Objective 1- Text-to-Sign Conversion

Implement a robust framework to convert textual input into corresponding ISL signs, enhancing accessibility for individuals with hearing impairments.

The Text-to-Sign Conversion where the related system is made to replace sign language gestures and translate text input into Indian Sign Language (ISL) using deep learning for hearing impaired people under this Section. The two main modules of the system are Phrase-to-GIF Translation that maps the text into a gif based on its semantic similarity (required a model trained on a sentence encoding), and Guidelines to Image Translation that would act as a more rudimentary fall back for those cases that do not match a phrase, which is translated based on characters. It's also robust to phrase-level translations, as well as individual character translations. Using the sentence Transformer package to create sentence embeddings and TensorFlow for tensor operations. It then uses the all-MiniLM-L6-v2 model from the Sentence Transformer library, which is a lightweight variant of the Mini LM architecture optimized for sentence embeddings. We take a 384-dimension vector space representation of the sentences. For a sentence SS the model produces an embedding vector es∈R384:es=Model(S) where: Model (·) denotes the pre-trained sentence embedding model. S is the sentence D is the dataset we are training it one is the embedding vector for sentence.

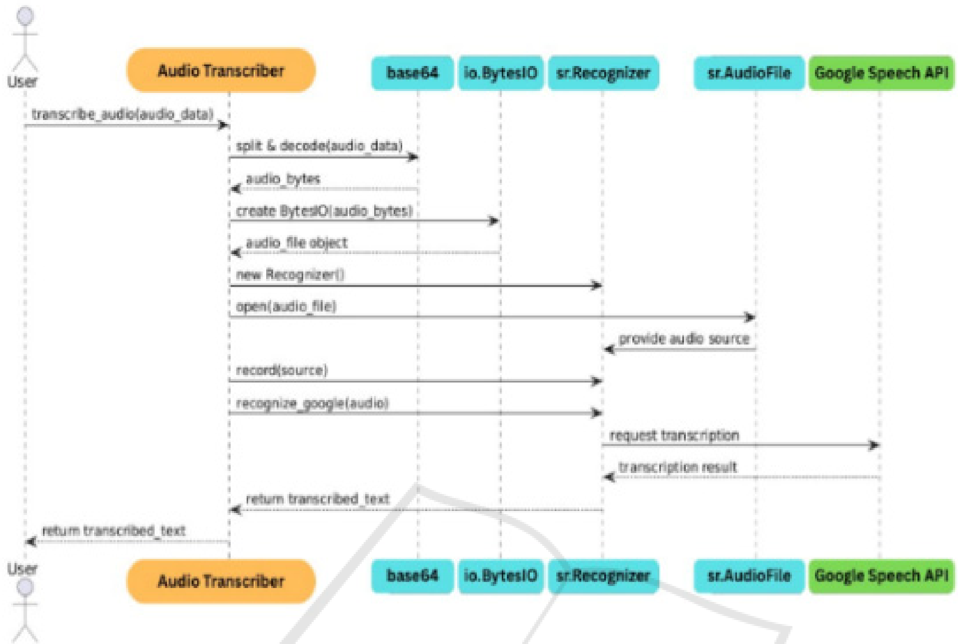

Figure 2: Flowchart of Text-To-Sign Conversion.

The system keeps a phrase-to-GIF dictionary to map well-known phrases to their respective GIF file paths. Phrases such as "hello" and "thank you" are mapped to the corresponding GIFs kept in the ISL translator/ ISL GIFs/ path. To increase efficiency, the embeddings for every known phrase are pre-computed and saved in phrase embeddings, which is represented as E= [e 1, e 2, e n], where E is the matrix of embeddings for every known phrase and e is the embedding vector for the i-th phrase. The algorithm breaks down the input phrase into individual letters and uses the letter to image dictionary to map each character to its appropriate image if no matching phrase is discovered. The algorithm processes every character c in the input phrase by determining whether it already exists in letter to image; if it does, the matching image path is added to the output. If not, an error on unsupported characters is recorded. The system is implemented using Python, TensorFlow, Keras, and OpenCV, leveraging a structured pipeline for processing text inputs. Figure 2 show the Flowchart of Text-to-Sign Conversion.

Resizable TF-IDF Vectorizer: train on raw text docSet, includes NLTK to handle utf-8 encoded text, decapitalize content, remove special character & split words, so for processing input O(n). This also allows for the proper transformation of textual information into movements, and ensures the correct meaning of the signs is preserved.

3.2.2 Objective 2 - Voice-to-Sign Conversion

Train on data until October 2023. Here, speech input get converts into text output, then this text has to further converted into Indian sign language (ISL) gestures.

This part explains how the system transforms spoken language into text through STT (Speech-to-text) conversion, using the Speech Recognition library which makes use of the Google Web Speech API to convert audio data into text. It utilizes common Python libraries, including base64 to encode audio bytes, io to create in-memory binary streams, and speech recognition to interface with voice recognition services. Google Web voice API: The cloud hosting service (Google cloud) and the voice recognition service (Google Web voice API). This strategy does have the advantage of using state-of-the-art speech recognition models trained on massive datasets, without needing to build and maintain sophisticated local speech processing capabilities. The transcription process begins with delivering the audio data, which can be provided as diverse methods, including base64, a common way to deliver binary data including audio over text-based paperwork. We are using the base64 library to decode this encoded audio data back to its original binary form. Upon receiving the encoded audio, the system uses the io library to convert the decoded audio bytes into an in-memory file-like object. However, this simplifies communication with the speech recognition engine because the speech recognition library can treat the audio data as if it were a normal file during the access and analysis stages. This pre-processing step is crucial for the subsequent transcription process to function correctly and efficiently. For hearing-impaired individuals, the transcribed text can then be translated into ISL gestures for seamless communication. Then the transcribed text is passed to Text-to-ISL Translation Module, which employs both phrase-level and character level translation techniques to convert the text to actual appropriate ISL signs. This integrated approach ensures a seamless process for bridging the communication gap, providing a reliable and accurate translation of spoken language into accessible ISL signs for individuals living with hearing impairments. Figure 3 show the Flowchart of Voice-to-Sign Conversion.

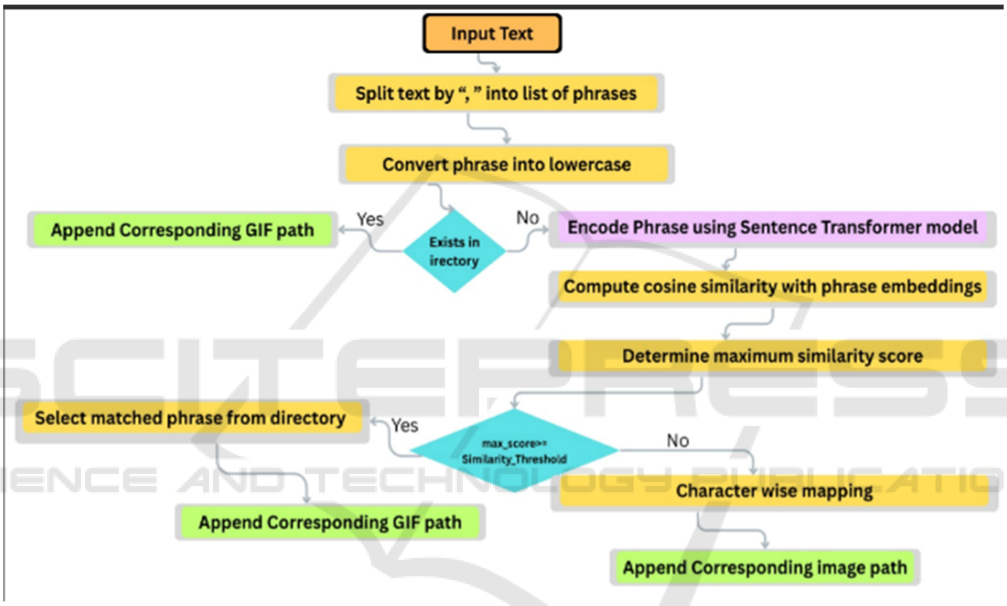

Figure 3: Flowchart of Voice-To-Sign Conversion.

Handling of user input as modular help modules are called successively makes the system easy to extend with new capabilities in the future, such as multilingual support or even a real-time processing module, ensuring that its practical application is flexible and future-proof. [8] The architecture of deep learning, such as Deep Belief Networks, sequence-to-sequence models, have also enhanced transcription accuracy significantly. However, cloud-based APIs (such as Google Speech Recognition) are here and, with them, you could have a good alternative instead of deep ASR systems, as they rely on efficient neural networks. API-based solutions improve performance and accessibility compared to their traditional counterparts and make development significantly easier.

3.2.3 Objective 3 Image-to-Sign Conversion

Structuring an Image-to-Sign Conversion, emphasizing the conversion of projected images input into their respective corresponding ISL signs, requires the integration of upper levels of image-to-text conversion fused with a text-to-ISL translator module. This will have the opportunity of people accessible for images4 to easily convert into ISL signs.

This section discusses the framework intended to transform image inputs into Indian Sign Language (ISL) signs. It has two main modules: image-to-text conversion and text-to-ISL translation. The Image-to-Text Conversion Module extracts text from images using an upgraded OCR (Optical Character Recognition) system, which employs

EasyOCR for text recognition and incorporates advanced preprocessing and postprocessing procedures to assure accuracy and reliability. The process starts with Image Input Handling, in which the system file uploads, and decodes them into binary data. OCR technology enhances its text recognition capabilities. Pattern matching is the purpose of the OCR model. However, accuracy was increased with the integration of deep learning and neural networks. Common image processing OCR techniques include denoising, segmentation, binarization, and skew correction. The primary component of OCR is features extraction. We employ a variety of feature extraction methods, such as stroke-based recognition, edge detection, and Gabor filters. By using preprocessing approaches and enhancing text detection with Pytesseract and EasyOCR, the suggested system image-to-text converter resolves the problems. Next, Image Preprocessing converts the image to RGB format for Easy OCR compatibility and optimizes OCR performance using techniques such as scaling, contrast correction, and noise reduction.

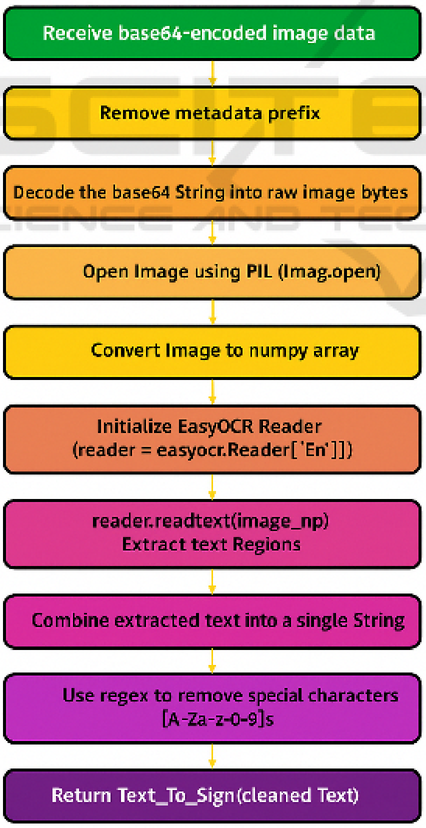

Figure 4: Flowchart of Image-To-Sign Conversion.

The Text Extraction phase detects and extracts text from the preprocessed image using EasyOCR, which combines fragmented text into a single string. Finally, Text Cleaning uses regular expressions to eliminate noise, unusual letters, and symbols while correcting frequent OCR problems with a dictionary-based spell-checker. The cleaned and corrected text is then routed to the Text-to-ISL Translation Module for additional processing. This program transforms captured text into ISL signs by mapping phrases to corresponding ISL GIFs based on semantic similarity, and for unmatched text, it falls back to character-level translation. Together, these components constitute a strong framework that bridges the communication gap, allowing for the seamless conversion of visual content into accessible ISL signs. Figure 4 show the Flowchart of Image-to-Sign Conversion.

4 RESULTS

4.1 Objective 1

The Text-to-Sign Conversion system effectively converts textual inputs into sign language. Because just like Indian Sign Language (ISL) gestures and movements assist people with hearing impairment to consume content. The model builds on the abilities of both deep learning models and linguistic algorithms, so that it can create correct and contextually accurate sign representation radically improving communication on a fundamental level. It offers a big edge in translating sign language with real-time voice conversion, allowing effective conversation in educational institutions, employments, and public service sectors.

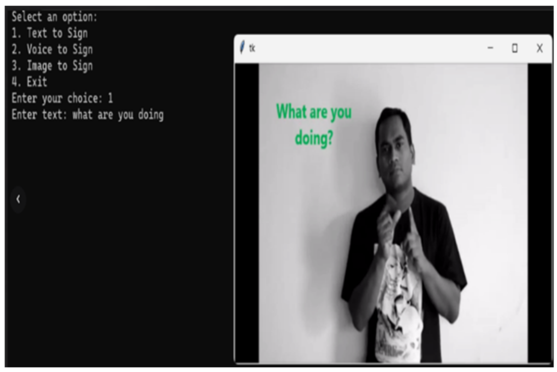

Figure 5: Outcome of Text-To-Sign Conversion.

Figure 5 show the Outcome of Text-to-Sign Conversion. The more people learn and get to know the language of the deaf and hard-of-hearing, the more inclusive society will be because that creates a communication link between the deaf and hearing communities, which enhances acceptance and social raw integration. It comes with a super versatile, diverse group of systems with multiple regional variations of ISL supported, capable of travelling successfully across human sign languages;

4.2 Objective 2

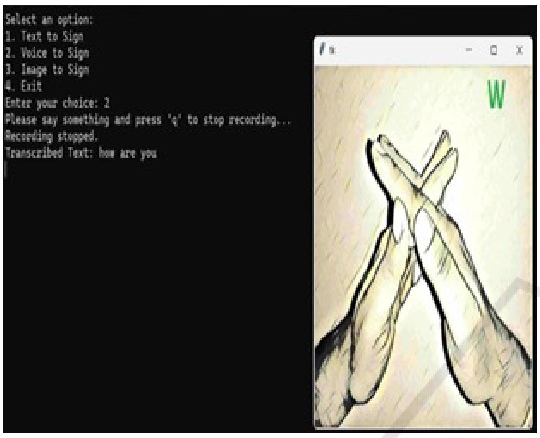

Figure 6: Outcome of Voice-To-Sign Conversion.

For that purpose, the technical breakthrough on Voice-to-Sign Conversion is important, especially when providing better communication for the hard of hearing. This is where the cloud-based speech recognition APIs (like the Google (Web Speech API) play an automatic role, as they have been designed specifically for the purpose of converting speech input to correctly transcribed text. In order to make the verbal communication easier for people who are deaf or hard of hearing, a new Text-to-ISL Translation Module translates the text to the ISL movements. Utilizing further advanced audio processing techniques as Python-based settings in pure audio format, it efficiently handles and processes Audio data, accurately detecting speech in real-time on high noise conditions, significantly improving performance. Being scalable and versatile, it can be applied into many environments like public places, offices, hospitals and schools. Its motion recognition based real-time speech to ISL gesture conversion system designed to help the deaf-mute people to have at most successful communication with vocally enabled person so that the people voice impaired person could also be included in the communication. Figure 6 show the Outcome of Voice-to-Sign Conversion.

4.3 Objective 3

The Image-to-Sign Language Transformation Framework is an innovative tool that utilizes uploaded visual content to create Indian Sign Language (ISL) movements effortlessly, catering to the needs of ISL users. Combining a sophisticated text-to-ISL translation module with advanced OCR technology to convert images to text facilitates reliable and accurate conversion of visual input into meaningful representations in ISL.

By conversion of text to visual information, quality of signs can be improved thus bridging the gap between knowledge and availability of the content in the format the deaf community can comprehend. It enables easier exchanges with digital text, printed materials, or signboards, promoting greater inclusion and independence, thanks to its real-time processing capacity. The dynamic of the system allows for really interesting areas of application, with some examples going as far as assistive technology, public services, and education where users are empowered by better access to information and seamless means of communication. This innovative technique really goes a long way in making ISL users more digitally and physically accessible to the world bridging the visual media with the language of signs. Figure 7 show the Outcome of Image-to-Sign Conversion.

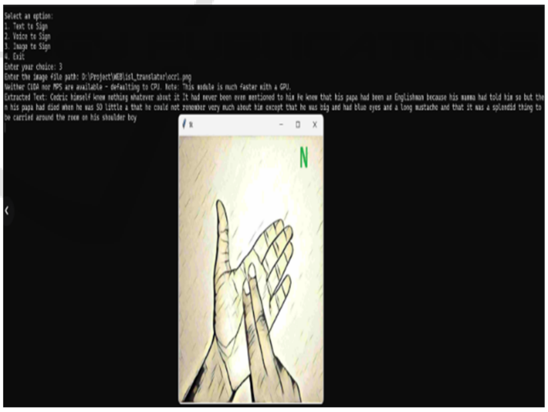

Figure 7: Outcome of Image-To-Sign Conversion.

5 DISCUSSION

It bridges this communication-gap smartly by an effective multi-modal sign language translation system by processing image, speech and text. Regardless of its capacity to deal with various input modalities, it is a key tool for ensuring inclusiveness and guarantees applicability and robustness in the real-world applications. But enhancing the accuracy and real-time performance is faced with stagnant constraints in processing speed and the size of sign databases.

This modular architecture of the system ensures we can define the size required for future improvements with better support for regional languages, enhanced machine learning integration, and high processing speed. Working with and working alongside the Deaf community and sign language experts will provide linguistic and cultural relevance, taking the system to the next level. These developments will result in transformative sign language translation technology, expanding the influence and accessibility of these technologies to a broader range of users. However, in order to be more universal and culturally relevant, the new translation system will also have to support regional dialects and languages. No only should they work together with linguists, but also with groups in the sign language and deaf umbrella to make sure that the linguistic correctness and actual needs are part of the system.

6 CONCLUSIONS

Utilizing advanced processing of speech, text, and image, the multi-modal sign language translation system they implemented, helps bridge communication between listeners and non-listeners alike. It serves as a flexible and consistent option for real-world scenarios. Its reliable functionality support for various input modalities reflects that it can be adopted in many practical scenarios, including public services, health care, health education, etc. This would be done by creating a larger sign database to optimize the time it takes to process and work in real-time while covering additional languages in context for international audiences. The system is modular and supports scalability, incorporation of latest techs like deep learning, and allows for creativity in the signs for Indian Sign Language interpretation. The system will develop to cover even more inclusive entities and it will help facilitate dialogue for the deaf and hard-of-hearing community in partnership with linguists, sign language experts and the deaf community."

REFERENCES

"Indian Sign Language", Published by National Institute of Open Schooling (An autonomous institution under Ministry of Education, Govt. of India), ISO 9001:2015 CERTIFIED.

"Speech to Text using Deep Learning", IJNRD - international journal of novel research and development (www.IJNRD.org), ISSN:2456-4184, Vol.9, Issue 4, page no.c86-c93, April-2024, Available: https://ijnrd.org/papers/IJNRD2404243.pdf

A. Adeyanju, O.O. Bello, M.A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis", ScienceDirect Intelligent Systems with Applications, 2021

Muhammad al-Qurishi, Thariq Khalid, Riad Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues", IEEE, s2021.

Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, Richard Bowden, "Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Shobhit Tyagi, Prashant Upadhyay, Hoor Fatima, Sachin Jain, Avinash Kumar Sharma, "American Sign Language Detection using YOLOv5 and YOLOv8", ResearchGate, 2023.

Sinha, Swapnil & Kataruka, Harsh & Kuppusamy, Vijayakumar. (2020). Image to Text Converter and Translator using Deep Learning and Image Processing. International Journal of Innovative Technology and Exploring Engineering. 9. 715-718. 10.35940/ijitee.H6695.069820.

Yogeshwar I. Rokade, Prashant M. Jadav, "Indian Sign Language Recognition System", ResearchGate, 2017.