International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

SIGN LANGUAGE TO TEXT CONVERTOR

Hrithik Suresh[1], Noufal Ismail[2], V Hari Krishna[3], Vysakh M Sudhakaran[4] , Santhi M[5]

[1]*Hrithik Suresh , Dept of CSE , Sree Narayana Gurukulam College of Engineering , Kochi , India*
[2]*Noufal Ismail , Dept of CSE , Sree Narayana Gurukulam College of Engineering , Kochi , India*
[3]*V Hari Krishna, Dept of CSE , Sree Narayana Gurukulam College of Engineering , Kochi , India*
[4]*Vysakh M Sudhakaran , Dept of CSE , Sree Narayana Gurukulam College of Engineering , Kochi , India*
[5]*Santhi M, Assistant Professor , Sree Narayana Gurukulam College of Engineering , Kochi , India*

Abstract – Sign language is a vital form of communication for the Deaf and Hard of Hearing community, yet there remains a significant barrier in real time interaction with non-sign language users. This project proposes the development of a Sign Language to Text Converter, an innovative system designed to translate sign language gestures into written text. The goal of the system is to produce an accessible real-time tool that transcends the communication gap between the Deaf people and those who are not familiar with sign language, thus enhancing interaction in diverse spheres such as education, healthcare, and social settings. This system will utilize techniques such as computer vision and machine learning in recognizing and interpreting hand gestures.

The system processes real-time video input using OpenCV for image processing and MediaPipe for hand and pose detection, extracting key features such as hand positions, movements, and orientations. These features are then fed into a Convolutional Neural Network (CNN) trained to classify individual sign language gestures. The output is translated to text, which can either be displayed on a screen or used with the text-to-speech systems to provide an auditory output.

Key Words: MediaPipe, Convolutional Neural Network (CNN), Text To Speech(TTS), American Sign Language (ASL)

1.INTRODUCTION

Communication is a fundamental human need; however, for individuals in the Deaf and Hard of Hearing community, accessing effective communication with the broader population can be challenging.

Sign language is the first and main way of communicating for millions of people across the globe. Still, it is often misunderstood by individuals who are not aware of the language. It will also be very easy to get excluded in education, health services, and social interactions among many others. Bridging this gap remains a significant challenge for fostering inclusivity and providing equal opportunities for individuals who rely on sign language. The Sign Language to Text Converter project aims to address this problem by developing an automated system that converts sign language gestures into written text. The primary goal of this project is to provide a real-time, accessible tool that can translate sign language (such as American Sign Language, ASL) into text, enabling communication between Deaf individuals and non-sign language speakers. By creating a system that can automatically interpret sign language gestures, we can enhance interaction and reduce the dependency on interpreters or specialized knowledge of sign language.

The system will utilize computer vision and machine learning methodologies to process visual inputs, particularly to recognize hand gestures and movements. To capture the real-time gesture, the system will be using a camera, applying OpenCV techniques for performing image processing operations like background subtraction, contour detection, and image normalization.

MediaPipe, which is a library developed by Google, will be used to track hands and detect keypoints. This will be an efficient way to recognize hand positions and orientations in real time.

These features will be fed into a deep learning model, probably CNN, trained to classify different signs using labeled data. Once the gestures have been identified, the system will translate them into text that is corresponding to that, and display it on a screen or utilize a TTS system to read the translation out loud. This real-time feedback will bridge the gap of communication and increase accessibility among Deaf individuals, in everyday life as well as professional and educational life. Ultimately, this project aims to promote social inclusion, increase access to resources, and reduce misunderstandings that often arise due to communication barriers. By making sign language more accessible to the wider public, the Sign

Language to Text Converter has the potential to improve the lives of Deaf individuals and foster a more inclusive society.

2. LITERATURE REVIEW

The purpose of a literature review is to "review" the literature surrounding a certain topic area. In this case, literature relates to various methods and implementations in sign language recognition systems. These sources provide insight into current approaches, techniques, advantages, and challenges associated with different sign language recognition technologies.

[1]"Sign Language Recognition Using Machine Learning" (Ms. Ruchika Gaidhani, Ms. Payal Pagariya, Ms. Aashlesha Patil, Ms. Tejaswini Phad, Mr. Dhiraj Birari), provides a gesture detection system based on Convolutional Neural Networks(CNNs). Developed in Python with Jupyter Lab, the system is based on background subtraction and skin detection on the Indian Sign Language (ISL) dataset. The method utilizes CNN with huge accuracy in image recognition, and it can be used for real-time hand gesture recognition with minimal equipment; all it needs is a webcam and a computer. The limitations include the proper interpretation of complex gestures, meeting the demands of large dataset, and handling regional language variations and privacy concerns.

[2]" Sign Language Recognition System"(Charu Srivastava, Deveshi Sagar), presents an alternate CNN-based system integrated with TensorFlow, involving stronger deep learning frameworks and transfer learning techniques. The communication breaks will be mitigated with this system by instant translation of sign language into text in real-time settings. This model has its advantages, which are cost-effectiveness and the potential of being an educational tool; however, it poses quite a challenge in terms of achieving a high level of accuracy, particularly with complex gestures and limited datasets.

[3]" Vision-Based Hand Gesture Recognition for Indian Sign Language" (Jayesh Gangrade & amp; Jyoti Bharti), describes an ISL recognition approach that combines CNNs with depth sensors like Kinect, enabling the system to manage complex backgrounds and lighting variations. The method has real-time gesture-to-text or audio capabilities help facilitate communication for the hearing-impaired. This method has high accuracy; however, reliance on Kinect sensors may be pricey and poses difficulties because Kinect sensors require special positioning and cannot handle complicated gestures.

[4]" Speech to Indian Sign Language Translator" (Alisha Kulkarni, Archith Vinod Kariyal, Dhanush V, Paras Nath Singh), aimed at filling the communication gap by translating spoken English into ISL via an online interface. This gadget would enable communication for the deaf This software does not only assist impaired persons but also is an education material for learning ISL. But the accuracy of the tool depends on the internet connection and the complexity of the inputs by voice, and it only performs with the system increases independence and inclusion and helps break away from the services of interpreters with advanced AI, involving real-time gesture analysis. Although useful, the accuracy of the translator may depend on regional differences in sign language while in low resource areas, access to technology may be limited.

[5]" Hand Gesture Detection and Conversion to Speech and Text" (K. Manikandan, Ayush Patidar, Pallav Walia, Aneek Barman Roy), presents a webcam-based gesture recognition system using contour analysis and feature extraction. This system captures hand gestures in real time, converting them to American Sign Language (ASL) letters. Advantages include accessibility for users who cannot use traditional input devices and the ability to offer a hygienic touchless interface. However, complex backgrounds, gesture variability, and lighting conditions present challenges to accurate recognition.

[6]"An Efficient Sign Language Translation Using Spatial Configuration and Motion Dynamics with LLMs" (Eui Jun Hwang, Sukmin Cho, Junmyeong Lee, Jong C. Park), proposes the SpaMo model which takes the spatial as well as motion features of sign language videos. By utilizing the fusion of visual encoders with an LLM, the proposed model, SpaMo achieves superior translation performance by getting rid of the necessity for gloss annotations. However, its performance can be improved using more extensive datasets, and the training is quite complex fine-tuning.

[7]" LLMs are Good Sign Language Translators" (Jia Gong, Lin Geng Foo, Yixuan He, Hossein Rahmani, Jun Liu), proposes a system, VQ-Sign, that converts sign videos into discrete tokens for hierarchical translation using pre-trained LLMs. This method improves the quality of translating accuracy because it is multi-lingual with the property of compatibility with available LLMs. Though, complex implementation, intensive resource and data dependency are the challenges one has in attaining

consistent accuracy dialectically which also raises privacy concern in usage of video data.

3. PROPOSED METHODOLOGY

The methodology for the Sign Language to Text Converter project will create a reliable, real-time system for translating sign language gestures into written text. This approach is divided into sequential phases, all of which are necessary for high accuracy and usability.

Phase 1: Data Collection and Preparation The first step is to collect a good set of sign language gestures, which should start with American Sign Language (ASL). This dataset will cover all the hand shapes, positions, movements, and orientations of commonly used signs. To ensure the model generalizes well across different users and environments augmentation techniques such as altering background settings, adjusting lighting conditions, and varying hand sizes will be employed. These variations aim to reduce overfitting and enhance the model's adaptability to real-world scenarios. Data collection will be employed such that obtaining a good quality, labeled video sequences or images from a wide variety of users is upheld in order to represent the natural gesture variations and prevent model bias toward some given demographic.

Phase 2: Image Processing and Preprocessing Once the data is collected, preprocessing is performed using OpenCV to prepare the images for accurate model input. This stage includes background subtraction in order to isolate the hand gesture from its environment thus enabling more focus on relevant visual information. Other step applied in this stage include contour detection to get the shape of the hand clearer. Image normalization techniques, including resizing and standardizing to color or grayscale formats, to normalize input dimensions and minimize noise. The goal of these preprocessing steps is to give the model a consistent and simple input, so that it may concentrate on identifying gestures rather than noise.

Phase 3: Hand Tracking and Keypoint Detection To enable real-time tracking, the system will be using MediaPipe, an open-source framework developed by Google for efficient hand tracking. MediaPipe extracts critical landmarks or keypoints from the hand; it identifies the position of joints and fingertips, which are critical to distinguish between gestures. The real-time capability of MediaPipe ensures that the system can continuously detect and track hand movements with minimal latency, making it suitable for dynamic, multi-step gestures. These detected landmarks are processed to form a simplified feature set that represents the gesture in a more structured format, such as coordinates or vectors, reducing the need for extensive raw image data.

Phase 4: Gesture Recognition Model The heart of the system is a Convolutional Neural Network (CNN) trained to recognize and classify gestures. CNNs are chosen for their ability to learn spatial hierarchies, making them effective for visual pattern recognition. Using the preprocessed images or extracted keypoints, the CNN model is trained to classify gestures into respective signs. Various CNN architectures will be used in this training phase with the proper hyperparameter tuning in order to have the right balance between the two accuracy and computational efficiency.
CNNs will be trained using the labeled dataset acquired before and some techniques that would avoid overfitting such as dropout and batch normalization. Furthermore, the model can include recurrent layers, such as LSTM, if necessary to handle temporal patterns in complex, data multi-step gestures.

Phase 5: Real-time Translation and Text Display After a gesture is detected by the CNN model, it is translated to its text representation. This translation is then displayed in real time on a user interface or device screen. The real-time text output ensures an interactive experience, allowing Deaf users to communicate easily with non-sign language users. The user interface for the system will be designed to provide clear, easily readable text and may permit users to scroll back through previous translations for reference in ongoing conversations.

Phase 6: Text-to-Speech (TTS) Conversion (Optional) In cases where there is a benefit in hearing the spoken words, an optional Text-to-Speech system would be implemented to take the translated text and generate actual speech. This makes the application a lot more applicable because the Deaf user can join verbal conversations without necessarily having a human interpreter on hand. The TTS system will be set up to sound as human-like as possible sounding speech and could be customized to use a specific voice or tone as preferred by the user.

Phase 7: System Testing and Evaluation One would translate the developed system with real users from the Deaf and Hard of Hearing community. For this, the

testing phase will be required to evaluate the usability, speed, and accuracy of the system, along with its adaptability across diverse users, backgrounds, and environmental conditions. The objective is to identify and rectify any constraints or frequent errors that might pop out in diversified real-world settings. Metrics such as recognition accuracy, response time, and user satisfaction will be tracked. Additionally, stress testing the system under different conditions (e.g., varying lighting or motion levels) will help pinpoint areas for further optimization.

Phase 8: Continuous Learning and Model Improvement To facilitate long-term usability, the system will include a feedback mechanism through which users can classify misclassifications in real-time. Such feedback will be utilized to incrementally update the model so that the system becomes capable of adapting to new gestures or idiosyncrasies specific to individual users. The system will improve its recognition accuracy over time by incorporating user input and may even learn new signs or modifications to existing signs. Continuous updates based on user feedback and new data collections will keep the model current and responsive to emerging needs.

Phase 9: Expansion and Scalability Finally, as a long-term objective, the system will target supporting multiple sign languages, expanding from ASL to include other languages like British Sign Language (BSL) or regional dialects. Scalability will also involve making the system available on mobile devices, enhancing accessibility for users in various settings. Privacy and data security will remain key considerations, particularly as the system expands to different platforms and geographical areas. Architecture Diagram By sequentially addressing each stage, this approach is intended to design a high-performance, user-centered Sign Language to Text Converter that fills the communication gap, supports social inclusion, and provides Deaf individuals with a reliable tool for real-time interaction.

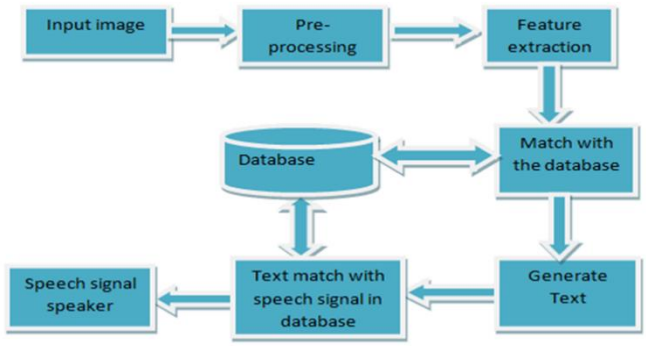

Figure 1 : Architecture Diagram

4.CONCLUSION

The purpose of Sign Language to Text Converter is to address issues faced by the Deaf and Hard of Hearing community as it aims at designing a real-time, accessible tool translating sign language gestures into text. The system, therefore is designed to allow the interaction of Deaf persons with others who may not understand sign language in learning and social lives as well as at the hospital. This system will use computer vision and machine learning by detecting hand gestures from a camera and then process them to produce a textual translation.

The project begins with collecting and preparing data, gathering a diverse set of sign language gestures, with special focus on American Sign Language (ASL). To improve robustness, data augmentation techniques are used to vary backgrounds, lighting, and hand sizes, ensuring that the model can generalize across different users and settings.

In the preprocessing stage, it uses OpenCV to separate hand gestures from the background and MediaPipe in determining the key hand landmarks, like joints and fingertips, for simplified gesture representations that increase the model's possibilities to trace hand movements. The core gesture recognition model is a Convolutional Neural Network (CNN) trained on the dataset to classify gestures with high accuracy. Different CNN architectures have been investigated to achieve a balance between speed and accuracy, both of which are very essential for real-time applications. Once a gesture is recognized, this is translated into text through a user interface and is potentially read out aloud by converting this text into speech with an integrated Text-to-Speech system. This feature allows the system to be used in verbal settings. This testing and evaluation will involve the collaboration with members of the Deaf and Hard of Hearing community for assessing how well the system performs in usability, accuracy, and adaptability. Users can make error corrections with mechanisms of feedback, making it more robust with time through continuous learning. In the future, expansion into more support for more sign languages will be added to the compatibility with mobile devices in enhancing the accessibility of the system.

In a nutshell, Sign Language to Text Converter is an important step in the direction of a more inclusive society. It gives Deaf people more autonomy and allows them to

communicate in situations where sign language interpreters or those who know sign language might not be available. As the project evolves, incorporating more languages and expanding platform availability, it has the potential to become a global tool for communication, breaking down barriers and fostering accessibility for all.

REFERENCES

[1] Gaidhani, R., Pagariya, P., Patil, A., Phad, T., & Birari, D. “Sign Language Recognition Using Machine Learning.” Department of Information Technology, MVP’s KBT College of Engineering, Maharashtra, India.

[2] Srivastava, C., & Sagar, D. “Sign Language Recognition System.” Inderprastha Engineering College, Sahibabad, Ghaziabad, India.

[3] Gangrade, J., & Bharti, J. (2020). “Vision Based Hand Gesture Recognition for Indian Sign Language Using Convolution Neural Network.”

[4] Kulkarni, A., Kariyal, A. V., Dhanush, V., & Singh, P. N. (2021). "Speech to Indian Sign Language Translator.”

[5] Strobel, G., Schoormann, T., Banh, L., & Möller, F. (2023). “Artificial Intelligence for Sign Language Translation.”

[6] Manikandan, K., Patidar, A., Walia, P., & Roy, A. B. “Hand Gesture Detection and Conversion to Speech and Text.”

[7] Hwang, E. J., Cho, S., Lee, J., & Park, J. C. “An Efficient Sign Language Translation Using Spatial Configuration and Motion Dynamics with LLMs.”

[8] Gong, J., Foo, L. G., He, Y., Rahmani, H., & Liu, J. LLMs are Good Sign Language Translators.”