ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1218

Real-Time Sign Language Translation for the Deaf and Speech-Impaired Individuals: A Systematic Usage of Deep Learning Approaches

Manjuvani Shivaling Gouda[1]; Dr. BhagyaJyothi K. L.[2]; Naseema C. A.[3]; Thajunnisa N. M.[4]

[1]M. Tech, [2]Professor, [3,4]Assistant Professor

[1,2,3,4]Computer Science and Engineering

[1,2,3,4] Department of Computer Science, K.V.G College of Engineering, Sullia, Karnataka, India

Publication Date: 2026/06/26

Abstract: Communication issues form a serious challenge for deaf and mute individuals. In this context, this study recommends a Real Time Sign Language Translator, which uses the principles of Computer Vision and Deep Learning to translate gestures to text and voice in real time through webcam input.

Keywords: *Sign Language Recognition, Sign Language Trans-Lation, Deep Learning, CNN, LSTM, Transformer Model, Vision Transformer, MediaPipe, Hand Gesture Recognition, Computer Vision, Natural Language processing, ASL, ISL, MobileNetV2, TensorFlow Lite, Speech Synthesis.*

How to Cite: Manjuvani Shivaling Gouda; Dr. BhagyaJyothi K. L.; Naseema C. A.; Thajunnisa N. M. (2026) Real-Time Sign Language Translation for the Deaf and Speech-Impaired Individuals: A Systematic Usage of Deep Learning Approaches. *International Journal of Innovative Science and Research Technology*, 11(6), 1361-1366. https://doi.org/10.38124/ijisrt/26jun1218

I. INTRODUCTION

Hearing impairment is a condition that affects hundreds of millions worldwide, whose sole mode of natural expression is through sign language. Sign languages are actually compli-cated languages in their own right, complete with grammar, syntax, and pragmatics, as opposed to simply being simplified modes of communication based on verbal languages. But due to a lack of sign language users among hearing people, a gap arises that prevents deaf people or those with impaired hearing from participating in various aspects such as education, em-ployment, health, and social activities. The usual approach of hiring interpreters carries its financial and logistic difficulties and is not practical in many instances, particularly unexpected occurrences. As a result, various technical means have been considered in research works. Previous solutions involving electronic instrumentation involved gloves and depth cameras, although they were both prohibitively expensive and intrusive. The advent of machine learning tools and efficient inference algorithms paved the way towards the visual approach only, utilizing cameras found in smartphones and laptops.

II. LITERATURE REVIEW

- Sign Language is the commonest mode of communica-tion used by individuals who are deaf and speech impaired; yet, in the absence of any efficient translation devices, there has always been a major communication barrier. Most existing systems of communication have been focused on one-way translation from signers to non-signers. This project seeks to propose a solution through the development of a real-time two-way sign language translation system. In order to achieve this, the project makes use of MobileNetV2 to facilitate gesture recognition, while TensorFlow Lite facilitates mobile optimization of the system. This makes possible the provision of smooth translations between gestures in sign languages and text in the form of an offline system. A second system that involves the creation of 3D animations using Blender will also be incorporated into the development of a mobile app based on the Flutter platform in order to enhance text-to-gesture conversions. The gesture recognition module provides a fast recognition rate of 80.5

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1218

- The “Mute” and “Deaf” individuals always use visual communication because due to not knowing sign language, non-deaf or non-mute people sometimes face difficulties and hence cannot communicate with the people who need their help. They understand and use sign language quite easily. De-veloping a model capable of translating gestures into text and voice would be vital in facilitating communication between disabled persons and the general public. Translating such sign languages into text and voice would be greatly advantageous since it would connect the deaf/mute people with the other part of society via communication between them. This research paper aims to design an approach to record the movement of hands (Sign Language). For the same, the Mediapipe platform has been used, which identifies the gesture of the hands and extracts all the data to input it into the model. After the sign language is recognized using the model, the words and alphabets are combined to form sentences and later into voice. With the help of LSTM, the proposed architecture has shown to have effective accuracy rates. According to the survey, CNN and LSTM show high accuracy and good quality results.
- The primary mode of communication among the deaf and hard-of-hearing is sign language; however, due to the lack of translation devices, this has been a significant barrier to communication. Current approaches to solving this problem have been limited to unidirectional translation from the signer to the listener. In this research paper, we explore a bidirectional sign language translator that would serve as a solution to this issue. To make this system possible, MobileNetV2 is used to enable real-time gesture detection and TensorFlow Lite to optimize the whole process into a completely offline system to provide users with a seamless translation experience. This sign language translator is combined with another 3D animated system developed through Blender software in a Flutter mobile application for the generation of text to sign gestures. The translator has an outstanding recognition accuracy rate of 80.5
- Communication barriers that occur in people who use sign language and those who don’t suffer from hearing im-pairments. In light of the lack of adoption of sign language and shortage of skilled human translators, the need to come up with more sophisticated alternatives is necessary. The present study offers a real-time sign language translation system that uses computer vision technology. It includes CNN and LSTM networks for static and dynamic sign recognition. Word segmentation involving the use of ‘word-ninja’ and a large language model (LLM) facilitated precise sentence generation. Machine translation and text-to-speech were added to the system as part of the method. The process entailed data acquisition, landmarks detection, and recognition and translation models. The results reveal very impressive figures in terms of accuracy of 99.20
- Sign language translation is one of the important tools used for improving communication between those individuals who cannot comprehend sign language and individuals suffer-ing from hearing problems. However, although a number of sign language translation models have been proposed so far, some of the major limitations encountered include inaccurate translations, incapability of handling complicated signs, and lack of ability to perform real-time translations. In this context, this paper describes Sign language translation systems for reducing the communication gap between the deaf and hearing individuals. The first goal of developing the sign language translation system is to design an efficient translator capable of translating the spoken words into animations of the same in sign language in order to facilitate easy communication even in real time by the hearing-impaired population. For this purpose, Long Short-Term Memory neural network models have been used for recognizing the speech while Convolutional Neural Network models are employed for translating the recognized spoken words to images in sign language with an impressive accuracy of 98
- The translation of sign language into the desired lan-guage is considered among the issues in communicating between deaf and hearing individuals, as it translates words through the movement of the hands, body, and mouth. One of the sign languages that are widely used includes American Sign Language, which is one of those sign languages. There are advancements in neural machine translation in terms of translating sign language into text. Transformer is the latest development in the field of natural language processing. This research paper compares Transformer with the Sequence-to-Sequence (Seq2Seq) model in translating sign language into text. Another experiment has been performed by using the Residual Long Short-Term Memory (ResidualLSTM) within Transformer. The result is that the use of ResidualLSTM within Transformer decreases its performance by 23.37
- Human beings exist in societies; therefore, they must communicate to articulate their demands and thoughts. Deaf and mute individuals rely on sign language in order to com-municate; however, many individuals do not understand the language used. As a result, a need arises to establish a proper system through which they can easily communicate. The present paper seeks to develop an American Sign Language (ASL) recognition system based on 12 dynamic signs using the MediaPipe tool and Long Short-Term Memory (LSTM) neural network. In order to improve system performance, only the required features are extracted from the dynamic sign video, together with the development of two additional features (angles and distances).
- For the deaf community, the use of sign language is indispensable in communication purposes, yet several hurdles exist for the process of converting their linguistic gestures into texts successfully. First, it is difficult for conventional sign language recognition approaches because of their dynamic na-ture, especially when applied in complex cases. Furthermore, most existing models are designed for sign language dialects, disregarding different sign language types used worldwide. This paper proposes a novel sign language translator, using cutting-edge technologies of computer vision and NLP, to address these problems. Specifically, the paper will use the Indian Sign Language (ISL), which is relatively complicated due to the use of two hands in the process. In particular, the

translator employs a Vision Transformer (ViT) model, which is trained with a rich video corpus, and then translates the videos into corresponding sign language dialects, along with employing PHI-1.5B for text refinement purposes.

- Barriers to communication that exist for the deaf and speech-impaired community require some creativity in order to find efficient ways of communicating with the hearing com-munity. In this research work, an attempt is made to provide a new technique of sign language translation and recognition through Random Forests and MediaPipe. MediaPipe makes it possible to track the hand movements and extract the hand landmarks required for understanding sign language. The hand landmarks recognized by MediaPipe are then processed using the Random Forest algorithm in order to recognize the sign language accurately. Once recognition is done, it becomes easy for the algorithm to translate the sign language into words and text form.
- Barriers to communication that exist for the deaf and speech-impaired community require some creativity in order to find efficient ways of communicating with the hearing com-munity. In this research work, an attempt is made to provide a new technique of sign language translation and recognition through Random Forests and MediaPipe. MediaPipe makes it possible to track the hand movements and extract the hand landmarks required for understanding sign language. The hand landmarks recognized by MediaPipe are then processed using the Random Forest algorithm in order to recognize the sign language accurately. Once recognition is done, it becomes easy for the algorithm to translate the sign language into words and text form.

III. PROBLEM STATEMENT

Communicating is an integral part of our lives; nevertheless, people with disabilities concerning hearing and speaking face difficulties while communicating with others since the latter might not know how to use the language of signs. It makes these people unable to take part in educational activities, work, and engage with other people because it leads to social isolation. As a rule, deaf people can communicate through the help of interpreters or by writing; nonetheless, such solutions are ineffective, inconvenient, and expensive. There are no affordable and automatic devices capable of translating the language of signs into oral and textual languages. However, the development of computers and computer vision enables the creation of a software-based real-time sign language interpretation program. Such a tool will make it possible to interpret signs and provide textual or oral translation within a short period of time.

The initial challenge posed by this research problem re-volves around the lack of a smart, efficient, and user-friendly software that would be able to detect and interpret sign language gestures to communicate messages to people using normal language with no human mediator present. This project seeks to bridge the gap in communication existing between the hearing-impaired community and others through the creation of an artificial intelligence-based tool that would interpret sign gestures in real time.

IV. OVERVIEW

The proposed Real-Time Sign Language Translation System is designed to bridge the gap in communication between hearing-impaired people and those who do not know sign language. Using computer vision and AI algorithms, this system will capture sign gestures from a live video feed and process the captured data to output readable text and speech. Firstly, sign gestures will be captured from a video stream using a camera, and the captured frames will be processed using image processing algorithms. The position of hands and body landmarks will be detected using the PoseNet algorithm, and the extracted features will then be classified using deep learning models. In order to analyze continuous signs, sequence models like LSTM and Transformers can be used.

Once the gestures have been classified, the sign sequences will be converted into a text stream, which will be further processed using NLP techniques. The output after translation into grammatically correct sentences will be displayed on-screen and read out aloud to users in real-time.

The proposed system aims at providing solutions that can work in real-time without any external dependency, making them portable and efficient.

V. SYSTEM DESIGN

The designed Real-Time Sign Language Translation system is an attempt to design a pipeline approach where a continuous stream of sign gestures captured by a camera is converted into text and speech. This is done using techniques in computer vision, machine learning, natural language processing, and speech synthesis.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1218

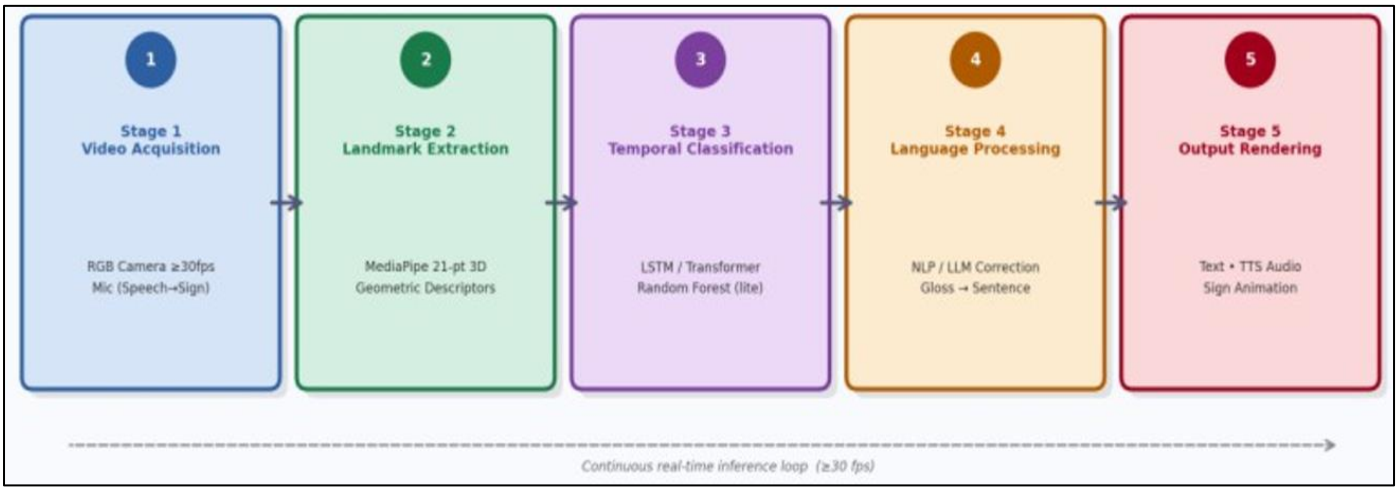

Fig 1 Five - Stage System Architecture - Consensus Blueprint

➢ *Stage 1: Video Acquisition*

The first stage involves receiving data from a standard RGB camera which acquires video frames at a speed of not less than 30 frames per second. High frame rate helps detect fast hand motions as well as minimizes any motion errors that could result when analyzing dynamically signing hands. The acquired video frames are processed to achieve constant lighting, correct camera positions, and clarity of the signing area. In the case of speech-to-sign conversion systems, an extra microphone input is employed for acquiring audio signals.

➢ *Stage 2: Landmark Extraction and Feature Engineering*

Each video frame is processed using a pose estimation framework such as MediaPipe to identify and extract hand landmark coordinates. The detected landmarks represent key hand joint positions and are converted into normalized fea-ture vectors. Geometric features including distances between landmarks and joint angles are calculated to create scale-independent representations of gestures. This feature engineer-ing approach improves robustness against variations in signer position, hand size, and camera distance.

➢ *Stage 3: Temporal Feature Grouping*

Temporal grouping of feature vectors extracted from succes-sive images is carried out through a sliding window process. The grouped features are then used as input to the classification model that will help in the identification of sign language ges-tures. LSTM-based networks offer effective sequence learning with minimal computational overhead for small and medium vocabularies. Transformers are the preferred network structure for sign recognition and translation tasks involving large vocabularies due to their effectiveness at modeling long-range dependencies.

➢ *Stage 4: Language Processing*

Signs that are recognized are processed as a string of gloss terms. Due to the fact that sign language has its own grammar rules that differ from those of oral languages, the process of conversion directly may lead to inaccurate sentences being generated. An additional layer of NLP processing is added to increase the quality of translation.

➢ *Stage 5: Output Rendering*

The translation output is rendered using multiple forms of communication. The sign language sequence can be shown as text output within a user interface and used to generate spoken words by means of a text-to-speech component. Text output can be transformed back to sign language either using animated avatars or skeletal hand movements for reverse translation. Implementation of the framework using optimized frameworks like TensorFlow Lite allows real-time rendering on portable devices without cloud support.

VI. METHODOLOGY

The proposed system is segmented into various modules where each module is responsible for performing a unique function in the gesture recognition process. Modules Involved in the Gesture Recognition Process.

- Image Capture Module: The webcam is initialized using OpenCV (cv2.VideoCapture) library. Video images are continuously captured and then fed for further processing. Resizing of image takes place in order to maintain uniformity among all frames.
- Hand Detection and Landmark Calculation: Media Pipe Hands library is used for detection of hand region. It detects 21 landmark points of the hands such as fingertips, joints, and wrist.
- Feature Extraction and Preprocessing : Landmarks are normalized using the position of the wrist landmark. White background image is generated, and landmarks are drawn for generating hand skeleton. This reduces the dependence on light and background conditions. Dataset obtained contains 180 hand skeletons representing alpha-bets from A-Z.
- CNN Model Implementation: CNN model is developed and trained using the processed images. It comprises convolution, pooling, flatten, dense, and output layers. It can classify gestures into alphabet classes and group gestures.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1218

- Text and Speech Output :The recognized alphabet is shown in real-time through the GUI. The application also synthesizes the recognized texts into speech sounds using pyttsx3 (offline).This feature increases accessibility for both deaf and non-deaf individuals.
- User Interface : A simple graphical interface is designed using Tkinter. It displays Live camera feed, Hand Detection , Predicted alphabet/text ,Suggestions

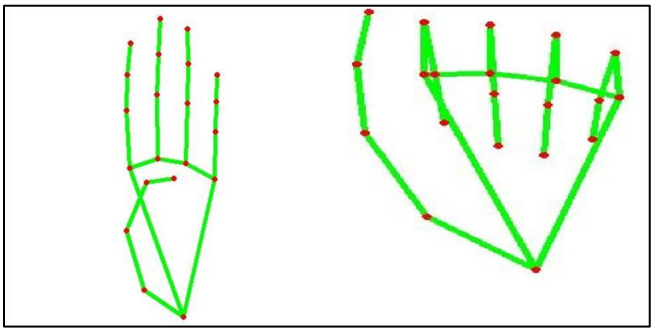

Fig 2 Hand Landmark Extraction and Skeleton Representation

VII. RESULT AND DISCUSSION

➢ *Accuracy of Recognition*

Static gestures (alphabet/letters) tend to be recognized more accurately compared to dynamic gestures, word-level recog-nition, or sentence-level recognition: • Vision Transformer (ViT) for ASL Alphabet: 99.99• SIGNFORMER (transformer encoder for ISL static signs): outperforms CNN baseline models with large margin • MobileNetV2 bidirectional mo-bile recognition system: 80.5• InceptionV3 Transfer Learning model for ASL alphabet: 89.91• CNN+SSD+2D-Pose Dual-model system for ASL greeting sentences: approximately 75

➢ *System Comparisons*

Trade-offs that occur among all systems reviewed include:

- Accuracy vs. practicality: transformer-based models demon-strate highest accuracy, however their high computational cost makes them impractical to deploy in real life situations.
- Bidi-rectional vs. unidirectional: just one of the papers reviewed offers truly bidirectional systems (MobileNet + TF Lite); most other systems translate signs to text.
- Offline vs. online pro-cessing: the mobile TensorFlow Lite system and ARM Cortex-based recognition systems are fully offline and therefore appli-cable to real life.
- Word-level vs. sentence-level: most systems recognize isolated signs and words; only CNN+LSTM+LLM system recognizes sentence-level sequences.

➢ *Key Technical Insights*

- MediaPipe Holistic landmark extraction has emerged as the standard framework for skeleton-based methods because of its efficiency and portability across multiple platforms.
- Transfer learning using pre-trained image classifiers signif-icantly cuts down the amount of training data needed for sign language.
- The use of sequence models (either LSTM or Transformer encoder) is indispensable for dynamic sign language where temporal information is key to interpretation.
- LLM incorporation marks the cutting edge towards more sophisticated sentence-level output.

VIII. CONCLUSION

Sign language translation techniques are explored through-out the paper, ranging from classic computer vision solutions to cutting-edge transformer architectures. The field has come a long way in the last few years: whereas earlier systems em-ployed hand-crafted methods for recognizing static gestures, nowadays it is possible to utilize neural networks to reach accuracy levels above 99

Nonetheless, there are several open problems in this domain. First, current systems work for limited parts of sign languages, such as the alphabet or vocabulary with predefined sets of gestures. Secondly, grammatical rules in sign languages imply that some non-verbal cues are essential for proper interpre-tation: facial expressions and body postures are ignored by most models. Thirdly, real-time processing is problematic for modern models, especially transformers due to their high computational cost. Finally, datasets containing continuous sentence-level translations for multiple sign languages are needed urgently in this domain.

Further developments may be directed towards: building up large multilingual sign language corpora; incorporating non-verbal elements, such as facial and body pose features into translation pipelines; applying sequence-to-sequence trans-former models (similar to today's NLP models); and designing efficient models deployable on mobile devices.

REFERENCES

[1]. M. Yu, J. Jia, C. Xue, G. Yan, Y. Guo, and Y. Liu, "A Review on Real-Time Sign Language Recognition," *IEEE Access*, vol. 10, pp. 1–1, 2022.

[2]. P. Goel, A. Sharma, V. Goel, and V. Jain, "Real-Time Sign Language to Text and Speech Translation and Hand Gesture Recognition Using the LSTM Model," *in Proc. 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT),* 2022, doi: 10.1109/ICICT55121.2022.10064562. .

[3]. "Sign Language to Text Translation with Computer Vision: Bridging the Communication Gap," IEEE Conference Publication, 2024.

[4]. "D. Thalisetty, A. Kopparthi, S. Kambhampati, and K. Sathvika, "A Novel Deep Learning Approach for Real-Time and Accurate Sign Language Translation," in *Proc. 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, 2024. .

[5]. M. Balamurugan, G. Nivedha, K. Rithika Devi, and G. Jeevika, “Real-Time Bidirectional Sign Language Translation Using MobileNet and TensorFlow Lite,” in *Proc. 2025 International Conference on Computing and Communication Technologies (ICCCT)*, 2025, doi: 10.1109/IC-CCT63501.2025.11019673.

[6]. G. G. S. Putra, A. W. C. D’Layla, D. Wahono, R. Sarno, and A. T. Hary-ono, “American Sign Language to Text Translation Using Transformer and Seq2Seq with LSTM,” in *Proc. 2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2024, doi: 10.1109/ICTIIA61827.2024.10761837.

[7]. Y. Farhan and A. A. Abdessalam, “Real-Time Dynamic Sign Recogni-tion Using MediaPipe,” in *Proc. 2022 IEEE 3rd International Confer-ence on Electronics, Control, Optimization and Computer Science (ICE-COCS)*, 2022, pp. 1–6, doi: 10.1109/ICECOCS55148.2022.9982822.

[8]. G. Lekshmi P and R. Francis, “Sign2Text: Deep Learning-Based Sign Language Translation System Using Vision Transformers and PHI-

[9]. 1.5B,” in *Proc. 2024 IEEE International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IICAIET)*, 2024, pp. 1–6, doi: 10.1109/IICAIET60765.2024.10730290.

[10]. S. K. Reddy and A. N. Sharma, “Real-Time Sign Language Recognition and Translation Using MediaPipe and Random Forests for Inclusive Communication,” in *Proc. 2025 IEEE International Conference on Emerging Trends in Computing and Communication (ICETCC)*, 2025, pp. 1–7, doi: 10.1109/ICETCC65210.2025.10932602.

[11]. A. K. Patel and M. S. Kumar, “Real-Time Speech to Sign Language Translation Using Machine and Deep Learning,” in *Proc. 2024 IEEE International Conference on Artificial Intelligence and Signal Processing (AISP)*, 2024, pp. 1–6, doi: 10.1109/AISP61234.2024.10522437.