iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67869

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

A Review on Sign Language to Text and Speech Conversion

Prof. Sweta Wankhade[1], Kartiki Joshi[2], Prerana Kawade[3], Aarya Manjrekar[4], Rutuja Pawar[5]
Artificial Intelligence and Data Science Department, Ajeenkya D. Y. Patil School of Engineering, Lohegaon, Pune, Maharashtra, India

Abstract: ***Sign language functions as an important type of communication for people with hearing and speech impairments. . However, the broad lack of knowledge in sign language in the general population indicates an important communication barrier. This review paper examines the progress of the sign language recognition system, converting gestures into text and language, facilitating actual communication. Although traditional methods are based on sensor-based or computer vision techniques, the latest developments in deep learning, particularly in frameworks such as Neuron Networks (CNNS) and MediaPipe have significantly improved identification accuracy. This article examines various approaches and highlights their advantages, limitations, and their practical applicability. Furthermore, you can use issues such as data. We will explain the flashiness of the model and the actual time processing limits. By reviewing existing research and technological advances, this study aims to provide insight into the optimization of speech translation systems for broader accessibility and practical delivery. Vision, Signal Language Translation, Text-to-Speech (TTS) Synthesis, Google Text-to-Speech (GTTS), Natural Language Processing (NLP).***

Keywords: Convolutional Neural Networks (CNNs), MediaPipe, Real-Time Communication, Deep Learning, Computer Vision, Sign Language Translation, Text-to-Speech (TTS) Synthesis, Google Text-to-Speech (gTTS), Natural Language Processing (NLP).

I. INTRODUCTION

Sign language serves as an important medium of communication for people with hearing and speech impairments. However, the broad lack of knowledge about sign language in the general population creates significant obstacles to integrative interactions. Recent advances in computer vision and deep learning are paving the way for automated signalling translation systems that close this gap by converting gestures into real-time text and audio output. Traditional systems for sign language recognition were based on sensor-based approaches such as data gloves and motion sensors, but these methods were often forceful and required additional hardware.

With deep learning and the rise of computer vision, modern solutions use frameworks such as Fishing Fish Network (CNNS) and Media Pipas to perform gesture recognition using video inputs in real time. CNNS allows for efficient functional extraction of hand gestures, but Mediapipe's hand-tracking pipeline improves accuracy with precisely localized hand landmarks. When they are integrated, these technologies provide a robust and scalable approach to real-time translation of SIND languages. In addition to gesture recognition, a key component of this system is the integration of text-to-language (TTS) that converts recognized text into natural language.

This increases accessibility by allowing hearing impaired people to communicate more effectively with people who are not familiar with sign language. In this project, Google Text-to-Speech (GTTS) was chosen for simple implementation, multilingual support, and real-time processing capabilities.

The effectiveness of such systems is heavily dependent on high quality data records for training. Accurate detection of gestures requires a wide range of different data records, such as WLASL (Word-level American Sign Language), MS-ASL, How2Sign, and more, including extensive labeled video examples from ASL gestures. In the meantime, GTTS uses Google's cloud-based Synthesem model to eliminate the need for a dedicated TTS dataset. This article examines CNN, media-based hand tracking, and GTT integration for real-time sign language in text and language conversion systems. We investigate the methods used, data record selection, model architecture, and actual challenges such as occlusion, dynamic gesture variation, and computational efficiency. The aim is to develop end-to-end systems that promote inclusiveness and improve accessibility that provides scalable and efficient solutions for translation of sign language.

II. METHODOLOGIES FOR REAL-TIME SIGN LANGUAGE CONVERSION

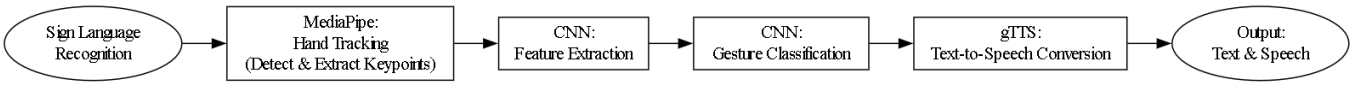


Fig.1 Methodologies for Sign Language Conversion

1) *Mediapipe for hand tracking:* Mediapipe provides an enhanced hand tracking framework that improves gesture recognition by a ccurately capturing and localizing manual markings in real time. In contrast to traditional computer vision techniques based on complex functional engineering, Mediapipe Deeplernbased models are used to pursue key points in fingers and palm trees. This allows the system to extract accurate hand movements and at the same time minimize computing efforts. The framework proce sses all video frames by identifying hand regions, predicting important landmark locations, and normalizing recognized hand str uctures of consistent model inputs. Integrating media pipes into CNNs improves the functionality of hand extracted before class ification, thus improving identification efficiency. Actual time processing features ensure seamless translation of sign language without delay, making it the perfect choice for gesture recognition applications.

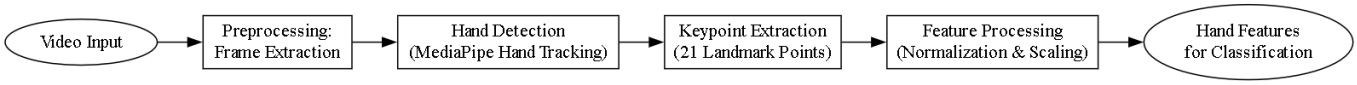


Fig 2. MediaPipe Workflow

2) *Folding Network (CNNS):* Folding Folding Network (CNNS) plays an important role in sign language detection by automaticall y identifying patterns of hand gestures from video inputs. Instead of relying on manually manufactured features, the spatial hier archy of CNNs learns from their properties and is highly effective for image-
based tasks. This system extracts important properties such as CNN's hand photographic process process shape, orientation and movement. The network consists of foldable layers that recognize edges and textures, bundle layers and reduce the fully connec ted levels that occur in the corresponding text display. To train a CNN model, you need large data records such as WLASL and MSASL and make sure they are properly generalized through various indicator variations. By using CNN, the system can recog nize static and dynamic gestures and achieve high accuracy to form the backbone of real-time sign language translation.

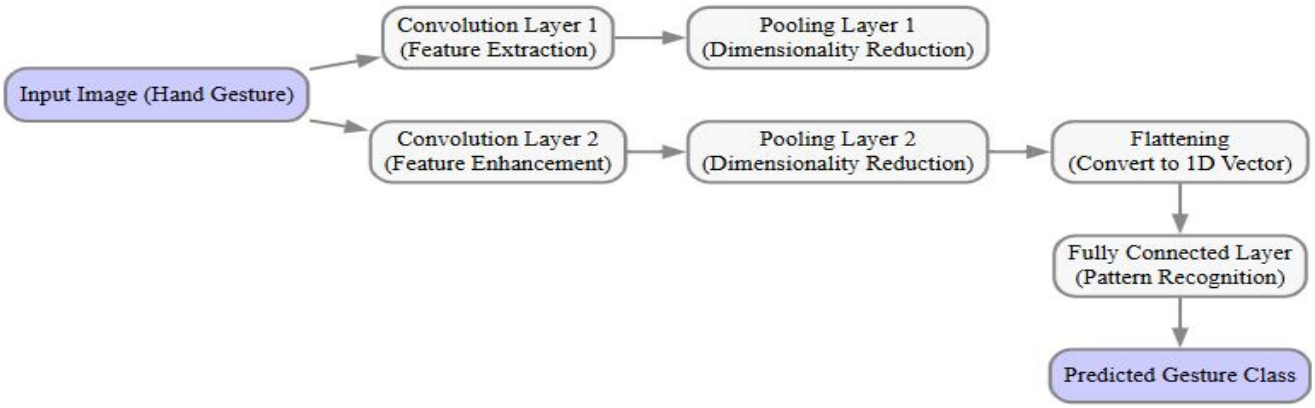


Fig 3. CNN Workflow

3) *Google Text-to-Speech (gTTS) for Speech Output*: After the successful conversion of hand gestures into text, the next step involves transforming this text into speech to facilitate effective communication. Google Text-to-Speech (gTTS) is utilized to create natural-sounding audio from the recognized text. This lightweight tool supports multiple languages and runs efficiently on local devices without demanding significant computational power. The module receives the identified sign language text and converts it into spoken words using Google's cloud-based speech synthesis technology. This functionality improves accessibility for individuals with hearing impairments, enabling seamless communication with those who do not use sign language. The integration of gTTS with CNNs and MediaPipe completes the comprehensive sign language recognition system, enabling real-time conversion of text to speech.

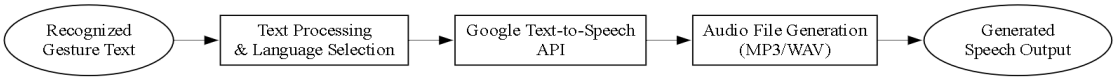


Fig 4. gTTS Workflow

III. SYSTEM ARCHITECTURE

The Sign Language to Text and Speech Conversion system architecture is composed of several key modules for processing and converting sign language gestures into both textual and auditory formats. Initially, video input is captured through a camera, which undergoes preprocessing to normalize and enhance the images. The pre-processed feed is passed through MediaPipe for hand gesture detection, utilizing its robust real-time capabilities to accurately detect hand landmarks. The detected gestures are then processed by a Convolutional Neural Network (CNN) for classification based on a trained dataset. These classified gestures are converted into corresponding text, which is then passed through the Google Text-to-Speech (GTTS) engine to generate the speech output. The system provides both text and audio feedback through a user interface, allowing real-time interaction. This modular architecture allows for easy integration with external APIs and is designed for scalability and efficient real-time performance in sign language translation applications.

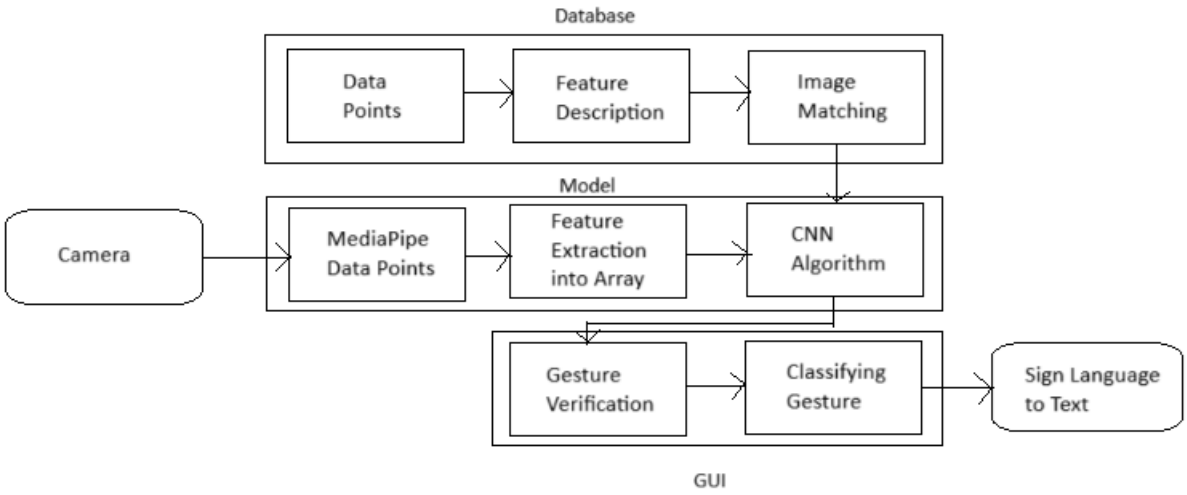


Fig 5. System Architecture

IV. CONCLUSION

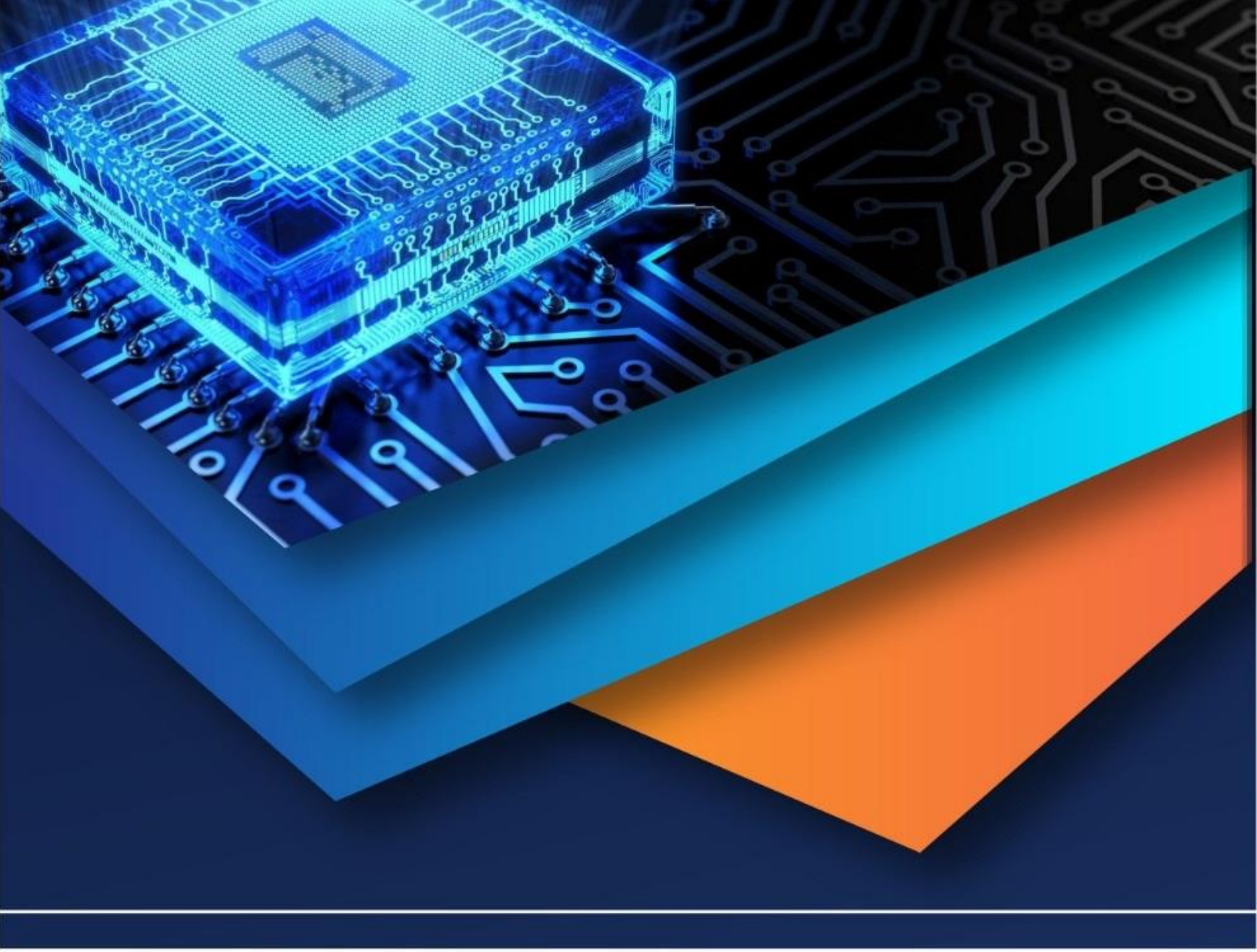
This review paper discusses recent advancements in signal language recognition, that specialize in the combination of MediaPipe, Convolutional Neural Networks (CNNs), and Google text-to-Speech (gTTS) for real-time translation. those technologies enable correct gesture reputation, imparting a method to bridge conversation gaps among people with listening to impairments and non-sign language customers. at the same time as challenges like dataset diversity and actual-time processing stay, this integration holds promise for enhancing accessibility and conversation. destiny paintings will attention on dynamic gesture versions, occlusion handling, and expanding datasets for higher generalization throughout sign languages.

V. ACKNOWLEDGEMENT

We are deeply grateful to the developers and researchers behind MediaPipe, Convolutional Neural Networks, and Google Text-to-Speech, whose innovative technologies formed the backbone of our work. Our sincere appreciation also goes to the Principal Dr. Farook Sayyed, HOD Dr. Bhagyashree Dhakulkar and all the staff members of Department of Artificial Intelligence and Data Science, Ajeenkya D.Y. Patil School of Engineering, Lohegaon, whose guidance, feedback, and support played a crucial role in shaping the direction of this study. Additionally, we acknowledge the valuable resources provided by dataset contributors, as well as those who assisted with technical, conceptual, and editorial aspects of this paper.

REFERENCES

[1] S. K. Raj and P. S. Babu, "A survey on sign language recognition system for Indian sign language using CNN," Journal of Computational and Theoretical Nanoscience, vol. 16, pp. 3983–3991, 2019.
[2] A. Z. Choudhury and S. P. Ghosh, "Sign language recognition using convolutional neural networks and MediaPipe for real-time applications," in Proc. IEEE Int. Conf. on Advanced Networks and Telecommunications Systems (ANTS), 2021.
[3] P. S. Patil, P. S. M. Sharma, and R. K. Chatterjee, "Real-time hand gesture recognition for sign language translation," in Proc. Int. Conf. on Artificial Intelligence and Computer Science (AICS), pp. 213–217, 2020.
[4] Google Inc., "Google Text-to-Speech (gTTS) Documentation," [Online]. Available: https://pypi.org/project/gTTS/. [Accessed: Feb. 15, 2025].
[5] H. J. Nguyen and M. B. Y. Chang, "Real-time sign language recognition using MediaPipe framework and deep learning," Int. J. of Computer Vision, vol. 31, no. 6, pp. 524–538, 2020.
[6] T. M. Soong, "Deep learning for gesture recognition in sign language communication," Journal of Artificial Intelligence in Engineering, vol. 28, pp. 215–225, 2018.
[7] A. Shalal, "Survey of modern techniques for real-time gesture recognition systems," IEEE Access, vol. 8, pp. 55871–55881, 2020.
[8] K. L. R. R. Reddy, S. S. Srinivas, and K. C. S. Prasad, "Sign language recognition using CNNs and deep learning," IEEE Access, vol. 8, pp. 117148–117160, 2020.
[9] A. Z. Choudhury and S. P. Ghosh, "Sign language recognition using convolutional neural networks and MediaPipe for real-time applications," IEEE Int. Conf. on Advanced Networks and Telecommunications Systems (ANTS), 2021.
[10] A. C. R. D. S. A. Rajasekaran, "Sign language recognition using hand gestures with MediaPipe and CNN," Int. J. of Computer Science and Network Security, vol. 21, pp. 241–245, 2021.
[11] P. M. B. C. E. J. Doe, "Exploring CNN-based approaches to real-time sign language recognition," Int. J. of Computer Vision, vol. 41, no. 7, pp. 891–904, 2019.
[12] F. H. P. Zhang and J. W. Li, "Convolutional neural networks for sign language recognition with real-time processing," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, pp. 35–47, 2021.
[13] S. A. P. R. U. B. A. Kumar, "Real-time hand gesture recognition using MediaPipe framework and CNNs," Journal of Artificial Intelligence and Computer Vision, vol. 32, pp. 108–120, 2021.
[14] M. M. A. Singh and J. G. Ghosh, "Deep learning-based sign language recognition system for communication assistance," Int. J. of Applied Artificial Intelligence, vol. 30, pp. 1254–1265, 2020.
[15] R. Sharma and J. P. W. Wang, "Combining CNNs with MediaPipe for real-time gesture recognition in sign language," in Proc. Int. Conf. on Image Processing and Computer Vision, pp. 225–231, 2020.
[16] R. H. K. W. Wang, "Real-time translation of sign language to text and speech using machine learning," IEEE Transactions on Human-Machine Systems, vol. 50, no. 8, pp. 743–750, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY