iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: VI Month of publication: June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83782

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Sign Language Gesture-to-Speech Model Using ML

Gulshan[1], Prof. Sunila[2]
Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar (Haryana)

Abstract: The fact that there is still no universally accepted solution to the accessibility issues that affect people with hearing loss is a huge challenge in modern society. One of the most common ways that the deaf and hard of hearing communicate is through sign language; however, when speaking with someone who does not know sign language, it becomes more difficult to convey meaning. Based on the findings of this study, SL Gesture-to-Speech Model powered by ML can facilitate communication between sign language users and those who do not know sign language. The suggested system includes a real-time sign language gesture detection system that uses DL and computer vision techniques to convert the gestures into understandable text and speech. The architecture integrates Media Pipe for gesture and hand tracking detection, CNNs for spatial feature extraction, and LSTM networks for temporal sequence analysis. To improve the recognition accuracy and train the software with more examples, transfer learning techniques are employed. After NLP identifies the movements, TTS are generated. The study's overarching goal is to develop a robust, accurate, and efficient real-time sign language recognition system that can handle challenges such as diverse singers, varying lighting sources, and unpredictable motions. The fields of education, medicine, customer service, and assistive technology can all benefit from this approach. A more accessible communication environment and better living conditions for those with speech and hearing impairments are both aided by the integration of computer vision and ML.
Keywords: Sign Language Recognition, ML, DL, CNN, LSTM, Media Pipe, Gesture Recognition, Computer Vision, Text to Speech, Human Computer Interaction, Assistive Technology, AI.

I. INTRODUCTION

Communication is an essential tool in human relations which allows people to convey information, thoughts and feelings. People who are deaf or do not speak English primarily communicate through sign language. Problems arise in the areas of education, employment, healthcare, and social interactions as a result of the severe lack of communication between sign language users and non-sign language users.

The fast pace at which the AI, ML and Computer Vision technologies are developing allows to develop intelligent systems that can recognize human gestures and convert them to meaningful outputs. SLR systems are an important part in minimizing communication barriers as they are able to recognize hand gestures and then translate the gestures into text or speech.

The proposed Sign Language Gesture-to-Speech Model is based on ML and DL [1] method to recognize the sign language gesture in real time. The goal of system is to capture hand motions by using a camera, extract the relevant features, classify the hand motion accurately and produce accurate speech output from the system. These systems can make a huge difference to the accessibility, independence and social inclusion of those who are hearing or speech impaired.

The objective of this research is to build an efficient and correct gesture-to-speech translation system that combines computer vision, DL algorithms with speech synthesis technologies. The study is also a step towards developing inclusive communication solutions to be used in education, health care, customer service and assistive technology applications.

A. Background of the Study

One of the most crucial communication methods that is utilized by those who have speech and hearing problems is sign language. It allows to communicate concepts, feelings and information with hand gestures, facial gestures and body movements. Although important, there is communication barrier due to the lack of knowledge of most of the population regarding sign language. New advancements in AI, ML, and CV have changed the game in this area, allowing for the creation of intelligent systems capable of recognizing and translating sign language motions [2].

CNNs and LSTMs have recently been used to successfully accomplish sequence analysis and gesture recognition. One potential use of recent tech developments is the creation of real-time sign language translation systems, which would greatly benefit communication and promote social inclusion [3].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

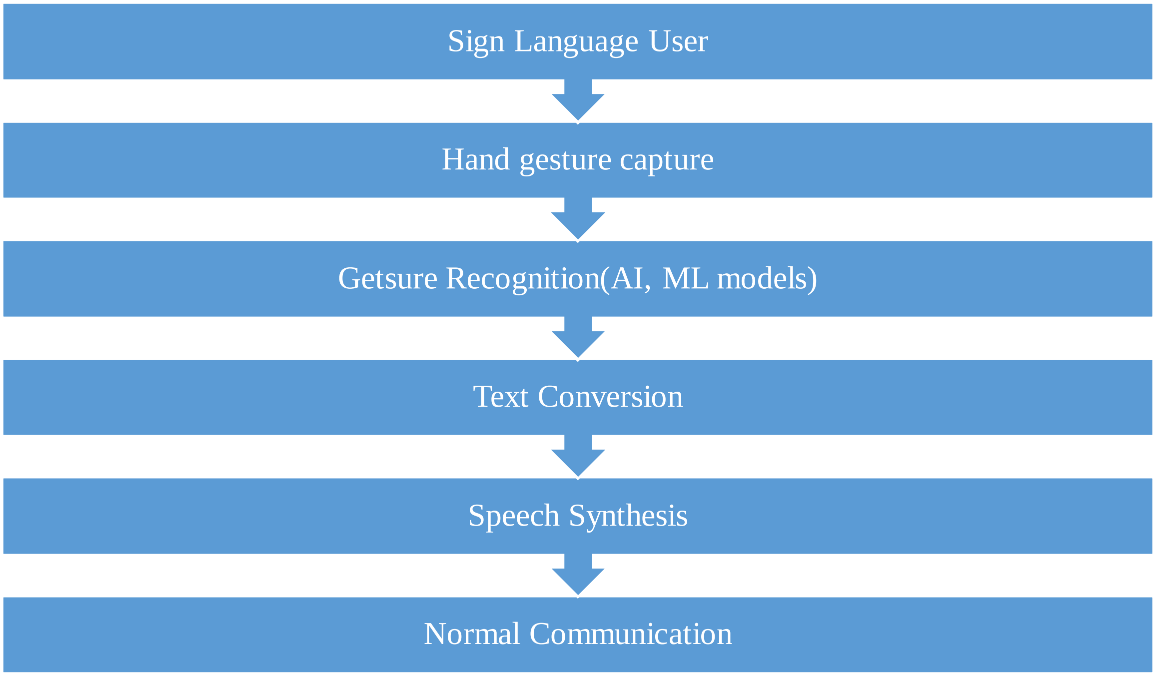


Figure 1. shows Flowchart of Conceptual Framework

B. Need of the Study

As more people are using sign language communication systems, there is a need to create effective translation solutions for sign language [4]. Current systems do not support limited vocabulary, do not achieve high accuracy, require high computation and do not perform well in real time. People who have hearing loss often face difficulties in educational, health and care, workplace and public services settings [5]. For the deaf and hard of hearing, this might mean more freedom, independence, and effective communication.

C. Challenges of the Study

The development of the effective recognition system for signed language has several challenges:

- Different hand shapes, gestures and signing styles of users.
- Various light conditions and background effects that impact on the quality of the image.

- Challenges with recognizing gestures that have sequential movements.
- There are few large sign language annotated and standardized datasets.
- Requires real-time processing, and low latency.
- Variations in skin colour and hand colour.
- Gesture recognition can be complicated in combination with speech synthesis systems [6].

- High accuracy and high efficiency of computation.

D. Motivation of the Study

Improving the accessibility of communication for persons with speech. Because of difficulties in communicating, people have less options when it comes to school, work, and social contacts. Intelligent assistive devices capable of accurate sign language interpretation have recently become feasible as a result of developments in computer vision, AI and ML [7][8][9]. The motivation for this study is the need for a practical and real time communication tool which will contribute to the promotion of inclusiveness and independence. The research aims to use cutting edge technologies to create an efficient system which allow people in the sign language community to interact with the wider community easily [10].

E. Sign Language Gesture-to-Speech Model Using ML

The suggested system has the following process:

1) Data Collection

- Collect sign language gesture images and sign language video data.
- Label and explain gestures and give meaning to them.

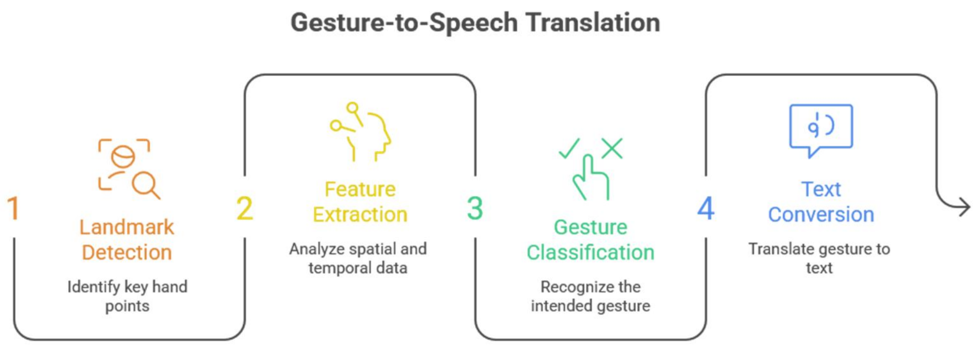


Figure 2. Gesture to speech translation

2) Data Preprocessing

- Normalize, clean and augment collected data.
- Using Media Pipe, extract hand landmarks.

3) Feature Extraction [11] [12]

- Use CNN models to Extract Spatial gesture features.

4) Sequence Analysis

- Use LSTM networks to model temporal aspects of gesture sequences.

5) Gesture Recognition [13-17]

- Recognise gestures and translate into sign language words.

6) Text Generation

- Transform known gestures to meaningful text.

7) Speech Synthesis [18][19]

- Convert text to speech using Text to Speech technology.

8) Performance Evaluation

- Calculate measure of accuracy, precision, recall, f1-score and inference time.

II. RECENT RESEARCH

Recent research in sign language translation and assistive communication has increasingly focused on developing intelligent, real-time, and bidirectional systems that enhance accessibility for individuals with speech and hearing impairments. Gogoi et al. (2025) proposed an integrated graphical user interface that connects both ISL gesture-to-voice and voice-to-sign translation in a single framework to enhance the seamless interaction between hearing and non-hearing users [20]. Bhosale et al. (2025) also modified the traditional sign language translation to add emotion recognition in the generation of the spoken output, enabling ML models to produce voice outputs that express the signer's emotion when considering the context [21]. The glove system designed by Zhang et al. (2025) is an intelligent bidirectional sign language recognition glove based on multi-sensor fusion and AI analysis, which has potential to enhance the accuracy of gesture acquisition and enable two-way communication through the glove system [22]. In the same way, Arora et al. (2025) have designed a vision-based sign-to-speech system which uses computer vision techniques to improve accessibility by accurately recognizing the gestures without any requirement of special hardware [23].

Additional progress has been focused on incorporating deep learning architectures for strong real-time capabilities. To capture both spatial and temporal information, Nyunt et al. (2025) used CNNs along with LSTM networks to translate gestures and speech in real time, enabling better machine interaction [24].

To keep up digital inclusion for speech-impaired societies, Betri (2026) proposed a real-time gesture-to-speech transformation system with alphabet recognition to achieve it by an efficient recognition of the gestures and the generation of speech [25]. In field of SL recognition, Mandu et al. (2026) designed AI based DL framework to directly map the identified sign language gestures to the voice output which is also a practical assistive technology [26]. Similarly, GESTURA, a computer vision and deep learning-based sign language translation system, was proposed by Anjana et al., 2026, which focused on the high recognition accuracy and the easy implementation in real world setting [27].

Personalization, multilingualism, and hybrid pedagogies have been other recent topics investigated to improve the effectiveness of communication. Yanushkevich et al., (2026) explored personalized augmentative and alternative communication systems that incorporate biometric authentication with adaptive communication technologies in order to deliver personalized user experience [28]. Rajasekar (2026) presented a new model named BiSign which is a bidirectional and multilingual model of sign language interpretation that supports multiple spoken and sign languages, tackling linguistic diversity and enhancing accessibility for different populations [29]. In this paper, Sivan and Eswaran (2026) proposed a hybrid deep learning model for Indian Sign Language alphabet recognition and speech synthesis, which integrated multiple learning paradigms to enhance the recognition accuracy and natural voice output of proposed model [30]. The overall trend of these studies is towards the realization of sign language translation systems that are computer vision-dependent, wearable-sensor-based, real-time, and multi-lingual, and that are powered by AI.

III. LITERATURE REVIEW

Binwant Kaur et al. (2023) proposed real-time SL to speech conversion system to enhance communication opportunities for deaf and hard of hearing people. Communication barriers hinder the ability of hearing-impaired people to engage in common communication, which is essential for intelligent assistive technologies. ML and CV-based framework capable of real-time recognition of SL gestures and translation to spoken language was introduced in the study. In contrast with the conventional approach which relies on large datasets and complex recognition techniques, the proposed method demonstrated improved recognition accuracy without compromising on real-time performance. The fact that there is still no universally accepted solution to the accessibility issues that affect people with hearing loss is a huge challenge in modern society. One of most common ways that deaf and hard of hearing communicate is through SL; however, when speaking with someone who does not know sign language, it becomes more difficult to convey meaning. The suggested system would use DL for sign language identification in real time, and then convert that data into useful text and voice. The architecture integrates Media Pipe for gesture and hand tracking detection, CNNs for spatial feature extraction, and LSTM networks for temporal sequence analysis. To improve recognition accuracy and train the software with more examples, transfer learning techniques are employed. AI detects the motions and converts them TTS. The results highlight the potential of AI to create intelligent and accessible human-computer interaction systems. [3]

V. Guruviah et al. (2023) proposed ISL recognition system using the real-time framework of body and gesture tracking, Media pipe Holistic. Current ISL recognition methods are usually hardware-intensive or assume a very large dataset in order to be effective. The proposed solution will use the landmark detection features of media pipe to detect all the hand gestures, facial expressions and body posture at the same time. Accurate recognition of ISL gestures while providing real-time response was achieved using DL models. This method contrasts from the conventional vision-based approaches, which rely on handcrafted features and are costly, impractical, and hard to scale.[4]

A. S. Nandhini et al. (2021) presented a sign language recognition framework using CNNs Traditional recognition techniques have had difficulties with sign variation, varying illumination, and inter-signer variation. By using CNN-based approach, the spatial features are automatically extracted from the gesture images which further boosts the accuracy of gesture recognition and robustness of the system. Unlike feature-engineered/rule-based approaches, CNNs are capable of learning meaningful representations directly from the raw image data which greatly diminishes the number of pre-processing steps needed [5].

The comparative study of various sign language recognition techniques and algorithms was done by R. Kumar et al. (2023) where the different techniques and algorithms were analyzed based on their recognition accuracy, computational complexity and their suitability for the real-time application. Traditional recognition systems have relied on handcrafted features and rule-based approaches; however they are not effective for different environments or handcrafted gestures. The current study was dedicated to the latest architectures (CNN, RNN and hybrid) for SL recognition tasks and performance of these architectures were evaluated. Results revealed the advantages and disadvantages of each strategy as well as the advantages of being accurate, but not too slow. The study found the need to develop more efficient tools that can be based on artificial intelligence (AI) technology to recognize communication and access for people with hearing loss to be more accessible. [6]

S. Thakar et al. (2022) presented a system that converts sign language to text in real-time using the principle of transfer learning. The goal is to enhance the efficiency and accuracy of recognition. Modern sign language recognition systems rely on antiquated methods that are difficult to apply in real-time settings due to the large amounts of data and computing power needed for training. In this research, the pre-trained DL model is employed to get a high accuracy for the gesture and meaningful gesture representation is obtained in a short training time. The system could finely tune a translated transfer learning model that achieved good generalization on the translation of sign language gestures to text on different datasets different users. This method has a few advantages over the normal methods of recognition particularly in the aspect of scalability and adaptability of an application to a real scenario. The research project highlights the significance of the transfer learning approaches, especially in the context of making the HCI systems more accessible for the persons with hearing and speech loss in assistive communication, with the aid of AI tools.

M. Al-Qurishi et al. (2021) provided an extensive overview of DL methods for sign language identification, reviewed all previous benchmark studies, and brought attention to the difficulties in DL-based sign language recognition. Due to variety of signers, backgrounds, and lighting conditions, most sign language identification systems cannot be applied in real-time, or in a generalizable or scalable manner. Exploring and classifying DL approaches, such as CNN, RNN, and hybrid ones, and evaluating their performance to enhance accuracy and robustness is goal of this work. Also discussed are some of the difficulties that may arise from putting the findings into practice, such as the scarcity of benchmark datasets and the difficulty in making models understandable.

Z. Liang et al. (2023) provide a comprehensive survey of sign language translation methodologies, investigating the trends, problems and future insights of sign language translation studies. The existing systems that process sign language have some drawbacks including the complexities of sign language, differences in sign gestures and absence of rich annotated databases. The authors classify various methods presented in the literature such as: Statistical machine translation, Deep-learning based machine translation and transformer models, and test them to support sign language to spoken language translation. Furthermore, the research explores the possibilities for incorporating multimodal data (hand movements, facial expressions, body movements and more).

To increase the accuracy of spam detection, I. Basyar et al. (2020) [10] suggest a classification system for email spams using a GRU and a LSTM networks. Conventional spam filtering approaches like rules-based and keyword-based filtering methods can suffer from high false positives and the ability to adapt to changing spam techniques is difficult. This study employs DL models that are able to capture the sequential dependencies between the words in the text to enhance classification accuracy over traditional ML algorithms. The combination of GRU and LSTM eliminates the vanishing gradient problem frequently encountered in recurrent networks and efficiently processes long sequences. This reveals a newfound significance of AI powered NLP. [10]

To improve gesture classification accuracy L. Pigou et al. (2015) propose SL recognition system using CNN. The traditional approaches to SL recognition are based on hand crafted features, but they are often affected by the variability of the lighting conditions, signer's movements and background noise. In this work, it investigates the possibility of automatically extracting the spatial features from the video frames of sign language using DL approaches with the focus on CNNs, for making the recognition more robust. The proposed model is an effective gesture classification model that greatly alleviates manual feature engineering and attains high accuracy in the classification of dynamic gestures. CNN based methods are more scalable and adaptable in various sign languages than the traditional vision based methods. The findings in this research have highlighted the possibility of DL in sign language processing, which will be beneficial to the advancement of HCI and accessibility solutions for the people with a hearing and speech impairment.[11]

S. B. Goyal et al. (2021) propose a framework for an IoT information analysis using fusion-based learning with DNNs to improve information processing and decision making. However, traditional data analysis methods are not efficient in dealing with the large-scale and heterogeneous data streams efficiently for IoT. This research combines the various learning methods with DNNs to enhance feature extraction, pattern recognition and prediction accuracy in IoT environments. The proposed fusion-based learning method optimally combines different data sources, leading to enhanced insights and system performance. Unlike the traditional ML models, DL models are more scalable and flexible in real-time application of IoT. The study underscores growing significance of AI-driven IoT analytics, paving the way for advancements in smart environments, industrial automation, and intelligent decision-making systems. In the recently published paper S. S. Gill et al. (2019) introduce the recent development and future trends of IoT, Blockchain and AI and discuss how these technologies are influencing the domain of cloud computing, while also addressing some of the issues that are yet to be resolved. When integrating IoT-based applications, traditional cloud computing systems have challenges to address in terms of security, scalability and real-time data processing. This research explores potential of Blockchain to boost the security and trustworthiness of cloud-based IoT systems, and examines the possibilities of AI solution to optimize use of resources, data analysis, and automation. Some of challenges are mentioned to be interoperability, energy efficiency and data privacy which need new frameworks to be addressed.

The findings underscore the evolving convergence of IoT, Blockchain, and AI in driving the future of cloud computing, further solidifying their importance in creating more intelligent, secure, and decentralized computing ecosystems.

A. M. Buttar et al. (2023) suggest a hybrid DL system for SL recognition to tackle problems of static and dynamic sign recognition. Recognition systems for traditional sign language are in general capable of recognizing only static or static and spatial signs, and they fail to recognize dynamic gestures, which necessitate a temporal analysis along with spatial feature extraction. The current study combines the use of CNNs for static sign recognition and RNNs for dynamic sign sequences, which increases the recognition accuracy and robustness. The proposed model is able to effectively capture spatial-temporal dependencies, and outperforms traditional ML-based classifiers.

This study emphasizes the importance of hybrid models based on DL for enhancing real-time SLR, which will help develop more accessible and effective HCI systems for deaf and hard of hearing. [14]

E. H. Tusher et al. (2024) summarize the current state of knowledge and issues in email Spam Detection and summarize the open research problems. The traditional methods of spam detection (rule-based filtering, blacklisting, etc.) are unable to stay one step ahead of the spammers and therefore result in high false-positive and false-negative rates.

Spam Pattern Detection using RNNs. Also, it points various issues like anti attack, data imbalance and privacy issues in email filtering. The findings underscore the need for creating more effective, adaptable, and explainable AI spam filtering solutions that can tackle real-world spamming challenges.

The work by W. Zhang et al. (2022) is about SL recognition system using depth image processing to enhance recognition accuracy of the gesture interpretation. Conventional vision-based recognition approaches, however, have difficulties in recognizing the variations in the environment, occlusion and background noise, making them difficult to use in reality. In this study, depth sensors are used to see the three-dimensional hand and body movements to get a more powerful and invariant description of the sign language gestures. The method is based on image processing and ML modeling to improve the accuracy of image recognition and minimize error and false recognition.

Minimizing dependence on RGB based datasets. The depth image processing is more flexible than the 2D image based recognition system in various lighting environments and signer's variations. The results highlight the need of depth-based vision techniques for the creation of more reliable and scalable sign language recognition systems which will be helpful to progress in HCI and assistive communication technologies.[16]

M. Alaghband et al. (2023) provide a comprehensive analysis of subjects on sign language, covering methods of study, evolution and challenges. The traditional sign language recognition systems have limitations and restrictions because of the availability of sign language datasets, variability of signs and complexity of sign language. The aim of this research is to summarize and analyze the works done in the field of vision-based recognition, sensor-based recognition, and DL-based models. Hand tracking, facial expressions, and body posture are all examples of multimodal data that are integrated in the study.

In area of movement, the ability to recognize has been improved. Additionally, the study points out several challenges, including the lack of data, signer independence, and constraints on real-time processing, which further emphasize the need for more powerful and flexible sign language translation models.

From the systematic literature review by N. Fatima et al. (2022), the developments, challenges and future trends of NLP for NLG using DNNs are covered. The classic text generation techniques, including rule-based and statistical techniques, are not well suited for the generation of contextually coherent, fluent and semantically rich text. The methods used in this study are classified as state-of-the-art DL-based methods like RNNs, LSTM, and CNN. In this study, the modern DL techniques like as RNNs, LSTM and CNN are classified. It highlights three important models that have pushed forward the field of text generation: LSTM, GANs and Transformer models, with a particular focus on how these models enhance the quality of text generation. Further, the review highlights some of the key challenges such as bias in data, computational complexity, and evaluating text coherence and diversity metrics. Overall, the study underscores the capacity of DL in NLP and its potential applications in automated content generation, chatbots, and communication support systems, among others, to greatly improve the performance of automated systems in NLP-related tasks. In the paper M. Papatsimouli et al. 2023, the authors give a comprehensive survey of developments of real-time sign language translators, focusing on them integrating with IoT technology. The traditional sign language recognition system either vision-based or sensor-based, both have hardware and computational complexity limitations, and vision-based has environmental limitations, making it difficult to get real-time performance. This study aims to investigate application of IoT based systems and their advantages to make sign language interpretation more efficient and accessible through use of cloud computing, edge devices and AI. The trends the research points out are wearable sensor integration, DL-based gesture recognition, and real-time. Moreover, it also deals with issues of data transmission delay, privacy and scalability. [19]

Table: Literature Review

Ref. No.	Author/Year	Objective	Methodology	Conclusion	Findings	Research Gap
1	Kaur et al. (2023)	Develop real-time sign-to-speech conversion for deaf and hard-of-hearing individuals	Computer vision and speech synthesis techniques	Effective communication bridge between signers and non-signers	Real-time translation improves accessibility	Limited vocabulary and scalability for continuous sign recognition
2	Renz et al. (2021)	Improve sign language segmentation	Temporal Convolutional Networks (TCN)	Accurate temporal segmentation achieved	TCN effectively identifies sign boundaries	Does not address complete sign-to-text/speech translation
3	Papastratis et al. (2021)	Review AI technologies for sign language processing	Survey of AI, ML, and DL approaches	AI significantly improves recognition performance	Deep learning dominates modern SLR systems	Lack of unified frameworks and real-world deployment studies
4	Guruviah & Goyal (2023)	Recognize Indian Sign Language using Mediapipe	Mediapipe Holistic landmark extraction with classification	Efficient ISL recognition achieved	Skeleton-based features reduce computational complexity	Limited handling of dynamic gestures and sentence formation
5	Nandhini et al. (2021)	Recognize sign language gestures	Convolutional Neural Network (CNN)	CNN provides high recognition accuracy	Deep learning outperforms traditional methods	Dataset diversity and real-time implementation remain challenging
6	Kumar et al. (2023)	Compare SLR techniques and algorithms	Comparative analysis of ML and DL methods	DL methods generally outperform conventional approaches	CNN and RNN show superior performance	Need for hybrid models and multilingual support
7	Thakar et al. (2022)	Convert sign language to text in real time	Transfer Learning with pretrained CNN models	Transfer learning improves accuracy with limited data	Reduced training time and improved recognition	Limited support for continuous sentence translation
8	Al-Qurishi et al. (2021)	Review DL techniques for sign recognition	Comprehensive survey of DL architectures	DL provides robust recognition capabilities	CNN, RNN, and LSTM dominate SLR research	Need for lightweight and interpretable models
9	Liang et al. (2023)	Survey sign language translation approaches	Analysis of recognition and translation techniques	Translation systems are rapidly evolving	Transformer-based methods show promise	Real-time multilingual translation remains unresolved
10	Basyar et al. (2020)	Classify email spam	GRU and LSTM networks	Sequential models achieve high classification accuracy	Deep recurrent networks handle text effectively	Not focused on sign language applications
11	Pigou et al. (2015)	Recognize sign language using CNN	CNN-based gesture classification	CNN can effectively classify hand gestures	Established DL foundation for SLR	Limited dynamic gesture recognition capability

12	Goyal et al. (2021)	Analyze IoT information using DNN	Fusion-based learning and Deep Neural Networks	Improved IoT data analysis performance	Fusion learning enhances accuracy	Not directly applicable to sign language translation
13	Gill et al. (2019)	Study IoT, Blockchain, and AI in cloud systems	Survey-based analysis	Emerging technologies improve intelligent services	AI and IoT integration enhances automation	Limited discussion on accessibility applications
14	Buttar et al. (2023)	Recognize static and dynamic signs	Hybrid deep learning framework	Hybrid model improves recognition accuracy	Better performance for mixed gesture datasets	Computational complexity remains high
15	Tusher et al. (2024)	Review spam detection methods	Comprehensive literature review	Advanced AI improves spam detection	Identified open research challenges	Not related to sign language communication
16	Zhang (2022)	Recognize signs using depth images	Depth image processing techniques	Depth information improves gesture recognition	Robustness against lighting variations	Limited integration with speech generation
17	Alaghband et al. (2023)	Survey sign language literature	Systematic review of SLR research	Significant progress in AI-driven recognition	Highlighted trends and challenges	Need for real-time multilingual systems
18	Fatima et al. (2022)	Review text generation using DNNs	Systematic literature review	DNNs effectively generate text	Transformers outperform traditional models	Limited focus on sign-language-generated text
19	Papatsimouli et al. (2023)	Survey real-time sign language translators with IoT	Review of IoT-integrated translators	IoT improves accessibility and deployment	Real-time systems becoming feasible	Security, latency, and scalability challenges remain
20	Gogoi et al. (2025)	Develop GUI for ISL gesture-to-voice and voice-to-sign translation	Integrated GUI with bidirectional translation	Improved communication accessibility	Supports both sign-to-voice and voice-to-sign conversion	Limited personalization and multilingual capability
21	Bhosale et al. (2025)	Transform sign language into emotion-enhanced speech	Machine learning-based emotion recognition and speech synthesis	Emotion-aware communication improves user experience	Generated more natural speech output	Limited real-time performance evaluation
22	Zhang et al. (2025)	Design bidirectional intelligent recognition gloves	Multi-sensor fusion and AI techniques	Wearable devices improve recognition accuracy	Sensor fusion enhances gesture detection	High hardware cost and user dependency
23	Arora et al. (2025)	Enhance accessibility through vision-based SLR	Computer vision and deep learning	Vision-based systems provide effective translation	Improved recognition without wearable devices	Challenges with environmental variations
24	Nyunt et al. (2025)	Enable real-time gesture and speech translation	CNN-LSTM hybrid architecture	Combined spatial-temporal learning improves translation	Real-time machine interaction achieved	Limited multilingual and contextual understanding

25	Betri (2026)	Develop real-time alphabet recognition system	Computer vision-based alphabet recognition	Supports digital inclusion of speech-impaired users	Accurate alphabet-level recognition	Does not support sentence-level interpretation
26	Mandu et al. (2026)	AI-powered sign detection with voice output	Deep learning framework	Effective sign-to-voice conversion achieved	Improved accessibility and automation	Limited contextual and semantic understanding
27	Anjana et al. (2026)	Build sign language translator using computer vision	Deep learning and computer vision techniques	Automated sign translation achieved	High recognition performance reported	Needs continuous sign recognition capability
28	Yanushkevich et al. (2026)	Develop biometric-enabled AAC systems	Biometrics integrated with AI communication systems	Personalized communication support enhanced	User-specific adaptation improves accessibility	Not specifically optimized for sign language translation
29	Rajasekar (2026)	Create bidirectional multilingual sign interpretation system	AI-based multilingual translation framework	Supports communication across languages	Bidirectional interpretation improves usability	Requires large multilingual datasets
30	Sivan & Eswaran (2026)	Recognize ISL alphabets and synthesize speech	Hybrid deep learning approach	Improved ISL recognition and speech output	Higher accuracy than conventional methods	Limited continuous gesture and sentence recognition

IV. PROBLEM STATEMENT

Even though the number of people that use ASL has increased, human to sign language communication remains challenging around the world. The communication barrier has been noted to impact on a number of areas of life such as education, employment, access to health services and social interactions.

While some sign language recognition systems have been implemented, most of the current systems have been restricted in vocabulary, recognition accuracy, requiring specific hardware, and not performing in real-time. The absence of any large-scale standardized sign language datasets and evaluation frameworks adds to the difficulties of developing reliable and universal sign language recognition systems.

The challenges could be met with the latest development of AI, ML and Computer Vision. But, still, it is required to have an efficient, scalable and accurate system for real-time recognition of the hand gestures in sign language and converting it to understandable speech. This system should be computationally efficient and not require large amounts of memory. A system of this type should have a small computational complexity, high recognition accuracy and adaptability to various users and environments.

Hence, this research focuses on the creation of a ML based model for SLG2S which can accurately detect real-time sign gestures, convert them into textual format and then translate them into the speech output. The system presented here aims to enhance the accessibility of communication, minimize reliance on an interpreter, and ensure inclusiveness for people who are deaf and/or have a speech disorder in different practical contexts.

V. FUTURE SCOPE

Currently, the system can only recognize a small subset of sign language movements and convert them to spoken text. However, as time goes on, it has the potential to be enhanced to accommodate a broader range of communication demands and a larger sign language vocabulary.

Future research should focus on addressing the issue of multilingual sign language recognition. Worldwide, sign languages such as ASL, BSL, and ISL are spoken by people. Better cross-cultural communication and access can be achieved through the development of more sign language translation and recognition systems.

A promising avenue to explore is the incorporation of cutting-edge Transformer-based architectures and LLMs to enhance the ability to understand context and generate sentences. These can be used to improve the understanding of fluent sequences of signs and provide more natural and meaningful speech outputs.

Wearable devices, smart gloves, and IoT technologies can further enhance the accuracy of gesture detection and its responsiveness in real-time. Also, it would make the technology more accessible and useful in the mobile and edge computing environments.

One possible future research is to combine the interpretation of the facial expression, eye gaze and body posture with the gesture recognition to identify the entire situation of sign language communication. Such a multimodal method can prove to be beneficial to recognition and translation accuracy.

The application of AR and VR technologies can create a communication space with a virtual or enhanced experience for educational, training and rehabilitation purposes. In addition, systems can find ways to adapt to the specific users and enhance their performance over time through the use of continuous learning mechanisms.

With the continuous development of technology, the sign language translation systems of the future will be more advanced and intelligent, and with greater accuracy and ease. The future of sign language translation systems is bright, and with advancements in DL, computer vision, and speech technologies, these systems will become even more sophisticated and intelligent, with improved accuracy and accessibility, creating a more inclusive and barrier-free society.

VI. CONCLUSION

Making a model that uses ML approach to convert sign language motions into spoken words is the goal of this project, which aims to help individuals who are deaf or hard of hearing with their communication issues. The deaf and hard-of-hearing rely heavily on sign language, but communicating with non-sign language users can be challenging. In this regard, the suggested system excels; it is able to recognize sign language gestures and immediately render them in understandable text and speech.

This research improves gesture recognition using state-of-the-art technologies like Computer Vision, ML, and DL. The gesture detection and gesture analysis modules are built using Media Pipe, CNNs to extract the features of the gestures, and the LSTM networks to analyze these gestures, which is a system capable of statically and dynamically interpreting gestures accurately. Moreover, NLP and TTS technologies can be utilized and enable the fluent translation of recognized gestures to speech.

The challenges faced with sign language recognition are discussed, such as variation in signer's hand movements, lighting, background and signer diversity. For large-scale effectiveness verification of the proposed model, real-world deployment inference times with comprehensive metrics like recall, precision, and F1-score can be utilized. The results of this research have implications for educational, health care, assistive communication and customer service technologies. The communication system proposed can improve the social inclusion, independence and quality of life of people with a hearing loss by providing an efficient and easily-accessed communication tool.

REFERENCES

[1] B. Kaur, A. Chaudhary, S. Bano, Yashmita, S. R. N. Reddy, and R. Anand, “Fostering inclusivity through effective communication: Real-time sign language to speech conversion system for the deaf and hard-of-hearing community,” Springer Science+Business Media, Oct.18, 2023.

[2] K. Renz, N. C. Stache, S. Albanie, and G. Varol, “Sign Language Segmentation with Temporal Convolutional Networks,” Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), May 13, 2021.

[3] Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, “Artificial Intelligence Technologies for Sign Language,” Sensors, vol. 21, no. 17, p. 5843, Aug. 30, 2021.

[4] V. Guruviah and K. Goyal, “Indian Sign Language Recognition Using Mediapipe Holistic,” Cornell University, Jan. 01, 2023

[5] S. Nandhini, D. S. Roopan, S. Shiyaam, and S. Yogesh, “Sign Language Recognition Using Convolutional Neural Network,” IOP Publishing, May 01, 2021.

[6] R. Kumar, A. Sinha, A. Bajpai, and S. Singh, “A Comparative Analysis of Techniques and Algorithms for Recognising Sign Language,” Cornell University, Jan. 01, 2023.

[7] S. Thakar, S. Shah, B. Shah, and A. V. Nimkar, “Sign Language to Text Conversion in Real Time using Transfer Learning,” Oct. 07, 2022.

[8] M. Al-Qurishi, T. Khalid, and R. Souissi, “DL for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues,” IEEE Access, Jan. 01, 2021.

[9] Z. Liang, H. Li, and J. Chai, “Sign Language Translation: A Survey of Approaches and Techniques,” Electronics, vol. 12, no. 12, p. 2678, Jun. 15, 2023.

[10] Basyar, A. Adiwijaya, and D. T. Murdiansyah, “Email Spam Classification Using Gated Recurrent Unit and Long Short-Term Memory,” J. Comput. Sci., vol. 16, no. 4, pp. 559–567, Apr. 01, 2020.

[11] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, “Sign Language Recognition Using Convolutional Neural Networks,” in Lecture Notes in Computer Science, Springer Science+Business Media, 2015, p. 572.

[12] S. B. Goyal, P. Bedi, D. K. Yadav, and N. A. Vakil, “Internet of Things Information Analysis using Fusion based Learning with Deep Neural Network,” J. Phys.: Conf. Ser., vol. 1714, no. 1, p. 012022, Jan. 01, 2021.

[13] S. S. Gill et al., "Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges," Internet of Things, vol. 8, p. 100118, Sep. 19, 2019.

[14] M. Buttar, U. Ahmad, A. Gumaei, A. Assiri, M. A. Akbar, and B. F. Alkhamees, "DL in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs," Mathematics, vol. 11, no. 17, p. 3729, Aug. 30, 2023.

[15] E. H. Tusher, M. A. Ismail, Md. A. Rahman, A. H. Alenezi, and M. Uddin, "Email Spam: A Comprehensive Review of Optimize Detection Methods, Challenges, and Open Research Problems," IEEE Access, vol. 12, pp. 1–15, Jan. 01, 2024.

[16] W. Zhang, "Sign language recognition based on depth image processing," Proceedings of HSET, vol. 23, Dec. 03, 2022.

[17] M. Alaghband, H. R. Maghroor, and I. Garibay, "A survey on sign language literature," ML with Applications, vol. 13, p. 100504, Oct. 24, 2023.

[18] N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota, and A. Abdullah, "A Systematic Literature Review on Text Generation Using Deep Neural Network Models," IEEE Access, vol. 10, pp. 56331–56350, Jan. 01, 2022.

[19] M. Papatsimouli, P. Sarigiannidis, and G. F. Fragulis, "A Survey of Advancements in Real-Time Sign Language Translators: Integration with IoT Technology," Technologies, vol. 11, no. 4, p. 83, Jun. 22, 2023.

[20] Gogoi, P., Karsh, B., & Karsh, R. K. (2025, August). Bridging Communication Gaps: An Integrated GUI for ISL Gesture-to-Voice and Voice-to-Sign Translation. In International Conference on Signal Processing and Computer Vision (pp. 404-418). Cham: Springer Nature Switzerland.

[21] Bhosale, A., Patil, L., Netake, G., Surwase, S., Gade, R., & Sahane, P. (2025, January). Transforming Sign Language into Emotion-Enhanced Speech with Machine Learning. In International Conference on Smart Trends for Information Technology and Computer Communications (pp. 209-223). Singapore: Springer Nature Singapore.

[22] Zhang, G., Wen, R., Zhou, Q., Cen, G., Luo, J., & Jin, Z. (2025, August). Design and Implementation of Bidirectional Intelligent Sign Language Recognition Gloves Based on Multi-sensor Fusion and Artificial Intelligence. In International Conference on Computer Science and Education (pp. 210-224). Singapore: Springer Nature Singapore.

[23] Arora, Y., Bajaj, S. B., Kaur, A., Garg, A., Narang, A., & Diwan, P. (2025, September). Sign to Speech: Enhancing Accessibility Through Vision-Based Sign Language Recognition. In Congress on Intelligent Systems (pp. 91-103). Cham: Springer Nature Switzerland.

[24] Nyunt, A. T., Vora, N., Vaghela, D., Kotak, B., Chauhan, R., & Zala, K. (2025, January). Real-Time Interaction with Machines Through Gesture and Speech Translation: A CNN and LSTM-Based Approach. In International Conference on Smart Trends for Information Technology and Computer Communications (pp. 173-182). Singapore: Springer Nature Singapore.

[25] Betri, T. J. (2026). Gesture to Speech Transformation: A Real-Time Sign Language Alphabet Recognition System for Digital Inclusion of the Speech Impaired Community. Citizen: Jurnal Ilmiah Multidisiplin Indonesia, 6, 1-10.

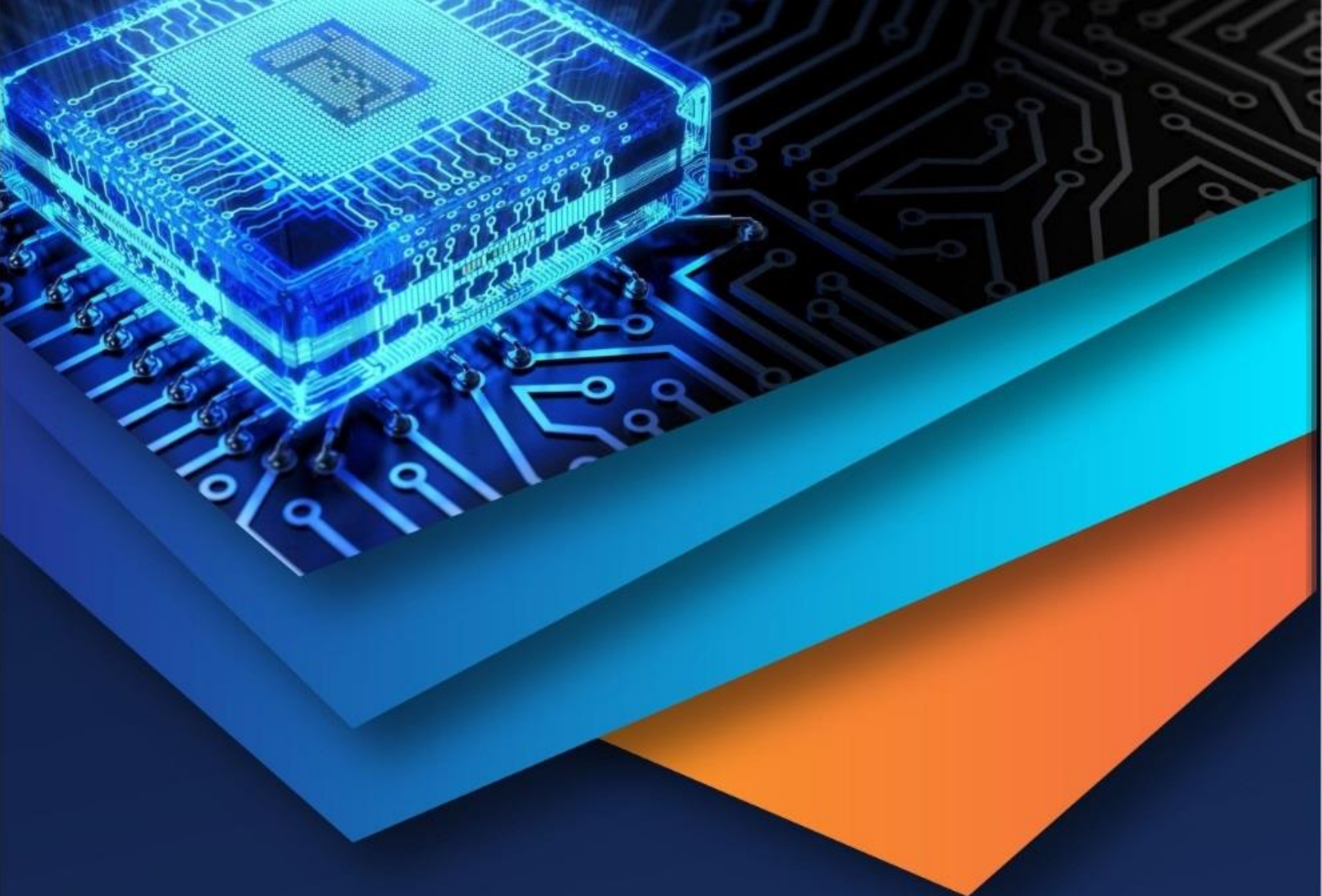
[26] Mandu, S., Nandhakumar, P., Akshitha, E., Anjana, D. S., Kumar, V., & Kumar, G. P. (2026). AI powered sign language detection with voice output using deep learning. In Computational Techniques and Smart Manufacturing (pp. 650-658). CRC Press.

[27] Anjana, K. S., Dhanvarsha, G., & Kanish, R. (2026, March). GESTURA: Sign Language Translation using Computer Vision & Deep Learning. In 2026 Second International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI) (pp. 1-6). IEEE.

[28] Yanushkevich, S., Berepiki, E., Ciunkiewicz, P., Shmerko, V., Wolbring, G., & Guest, R. (2026). Biometric-enabled Personalized Augmentative and Alternative Communications. arXiv preprint arXiv:2603.05512.

[29] Rajasekar, P. (2026, January). BiSign: Bidirectional and Multilingual Sign Language Interpretation. In 2026 5th International Conference on Communication, Computing and Electronics Systems (ICCCES) (pp. 2036-2043). IEEE.

[30] Sivan, A., & Eswaran, C. (2026). Indian Sign Language Alphabet Recognition and Speech Synthesis Using a Hybrid Deep Learning Approach. Archives of Acoustics.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY