IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⵋ Peer-reviewed & Refereed journal ⵋ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13602

Smart Translation For Physically Challenged People

Malashree M S[1], Sujal N Murthy[2], Thejaswini J[3], Yashapradha M[4], Yashwanth K[5]

Assistant Professor, Department of CSE, Maharaja Institution of Technology Mysore, Mandya, India[1]

Student, Department of CSE, Maharaja Institution of Technology Mysore, Mandya, India[2]

Student, Department of CSE, Maharaja Institution of Technology Mysore, Mandya, India[3]

Student, Department of CSE, Maharaja Institution of Technology Mysore, Mandya, India[4]

Student, Department of CSE, Maharaja Institution of Technology Mysore, Mandya, India[5]

Abstract: This paper introduces an AI-driven translation platform developed to enhance communication for individuals with speech and hearing disabilities. The solution integrates advanced technologies such as Natural Language Processing (NLP), gesture recognition using MediaPipe, and Artificial Neural Networks (ANNs) to enable seamless two-way interaction. It supports real-time conversion of spoken language into Indian Sign Language (ISL) through animated GIFs, while also interpreting gestures captured via webcam into coherent text. The system is deployed using a Django-based web application, ensuring both usability and scalability. Unlike conventional tools, this integrated solution enables dynamic, real-time, bidirectional communication, making it highly applicable in fields like education, healthcare, and public services. Experimental evaluations indicate the system achieves an 85% accuracy rate in recognizing gestures and a 92% accuracy rate for speech-to-sign translation, marking a notable advancement over existing approaches

Keywords: Sign Language Translation, Assistive Technology, Artificial Neural Networks, Natural Language Processing, Accessibility, Gesture Recognition, Indian Sign Language

I. INTRODUCTION

In today's increasingly digital world, communication remains a significant hurdle for individuals who are deaf or mute. With around 70 million people globally relying on sign language as their main form of communication [1], there exists a major disconnect, as most hearing individuals are not proficient in understanding or interpreting sign language. This communication gap presents ongoing challenges in key areas such as education, healthcare, and everyday social interactions.

Although advancements in artificial intelligence and computer vision have opened doors to innovative assistive technologies, many existing tools are limited in scope—focusing solely on either converting speech to sign language or translating sign language into speech [2]. This one-way functionality restricts the fluid, two-way communication essential for natural conversations. Additionally, regional variants like Indian Sign Language (ISL) are often overlooked in favor of more widely supported systems such as American or British Sign Language.

To overcome these challenges, our research introduces a comprehensive system that offers:
1. Real-time conversion of spoken language into ISL using NLP methods
2. Vision-based gesture recognition with an accuracy rate of 85%
3. A web-based interface designed for broad accessibility
4. Context-sensitive translation to enhance communication accuracy

Built on advanced deep learning frameworks, the system is optimized for performance while ensuring it can run efficiently on standard consumer devices. This approach ensures the solution is not only powerful but also practical for use in schools, public services, and other resource-constrained environments, particularly within developing regions.

II. PROBLEM STATEMENT

Existing solutions are either one-way translators or limited in real-time interaction. Most depend on pre-recorded datasets or require human interpreters. Audio-to-sign and sign-to-text translations rarely coexist within a single platform, highlighting the need for an integrated, AI-enabled solution.

DOI: 10.17148/IJIREEICE.2025.13602

III. SYSTEM DESIGN

Architecture Overview

The system architecture (Fig. 1) comprises three main components:

1. **Frontend Interface**: Django-based web application with responsive design
2. **Processing Modules**:
   - Audio processing pipeline (speech-to-text)
   - Visual processing pipeline (gesture recognition)
3. **Backend Services**:
   - NLP engine for text processing
   - ANN model for gesture classification
   - Database of ISL signs and synonyms

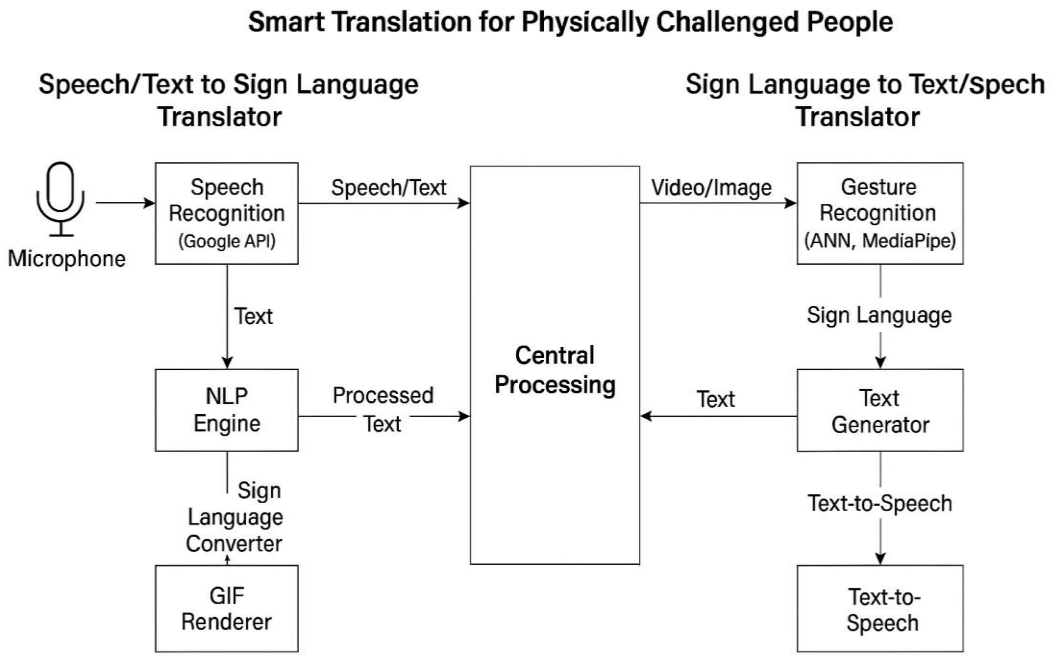

FIG. 1 SYSTEM ARCHITECTURE

IV. METHODOLOGY

A. Audio-to-Sign Module

The audio processing pipeline follows these steps:

1. **Audio Capture**: PyAudio module records speech input
2. **Speech Recognition**: Google Speech API converts audio to text
3. **Text Processing**:
   - Dependency parsing analyzes sentence structure
   - Stopword removal and lemmatization
   - Synonym replacement for missing vocabulary
4. **Sign Generation**:
   - Dictionary lookup for ISL GIFs
   - Grammar-aware sign sequencing

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025
DOI: 10.17148/IJIREEICE.2025.13602

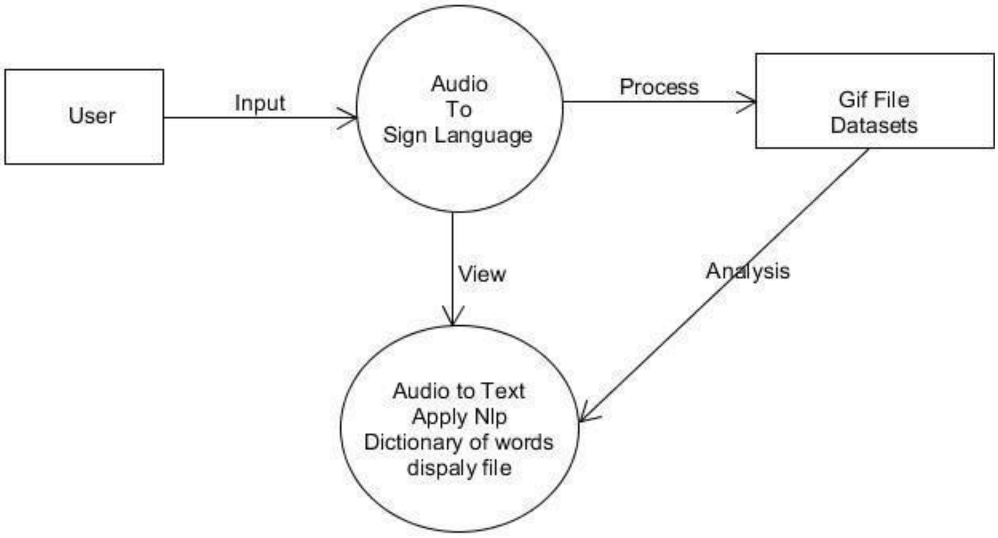

FIG. 2: TEXT TO SIGN LANGUAGE

B. Gesture-to-Text Module

The visual processing pipeline implements:

1. **Hand Tracking**: MediaPipe extracts 21 landmark points
2. **Feature Extraction**: Normalized spatial coordinates
3. **Classification**:
   - ANN model with 3 convolutional layers
   - Softmax output for 26 ASL letters + 5 common words
4. **Sentence Formation**:
   - Space/del gestures for editing
   - TextBlob for grammatical correction

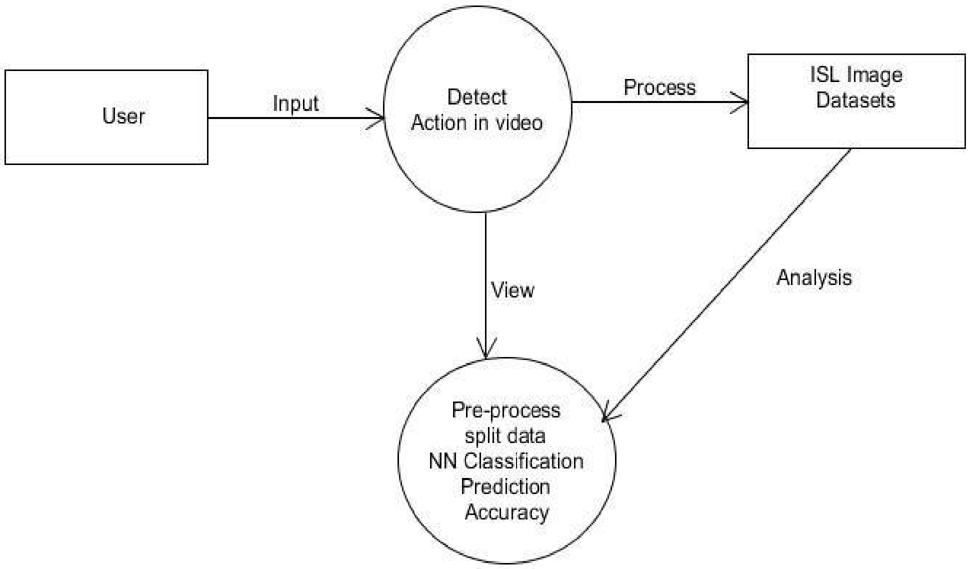

FIG.3: SIGN TO TEXT LANGUAGE

DOI: 10.17148/IJIREEICE.2025.13602

V. IMPLEMENTATION

A. Dataset and Training

We collected two primary datasets:

1. **ISL GIF Database**: 1,200 words with multiple GIF representations
2. **Gesture Dataset**: 5,000 images of hand signs (20 samples per class)

The ANN model was trained using:

- 80/20 train-test split
- Adam optimizer (lr=0.001)
- Categorical cross-entropy loss
- Early stopping (patience=5)

B. Technical Stack

Hardware Requirements:

- Intel i3 3.30 GHz processor
- 4GB RAM
- Webcam and microphone

Software Components:

- Backend: Python 3.8, Django 3.2
- ML: TensorFlow 2.4, MediaPipe 0.8.9, Python
- Frontend: HTML5, CSS3, JavaScript
- Framework: Django
- Database: MySQL 8.0

C. Optimization Techniques

Key optimizations included:

- Landmark normalization for viewpoint invariance
- Sliding window averaging for gesture stability
- Caching frequent sign sequences
- Asynchronous processing for responsive UI

VI. LITERATURE SURVEY

- One proposed system utilizes computer vision, image processing, and neural networks to recognize Indian Sign Language without the use of markers. It captures hand movements via webcam and translates commonly used sentence gestures into textual content, which is then converted to audio for seamless communication.
- Another study applied AdaBoost and Haar-like feature classifiers to interpret American Sign Language (ASL) from live video streams containing complex backgrounds. The use of a large, labeled dataset allowed for improved model accuracy. Each letter was captured over 20 video frames to evaluate performance.
- A speech interface system reported an accuracy of over 94% using the Android Speech API. Morphological analysis and stemming algorithms further enhanced performance, resulting in an overall effectiveness of approximately 81%, validating the system's translation reliability.
- Recent developments in Sign Language Translation (SLT) have explored neural-based methods. While SLT has shown promise, current implementations face challenges in managing long sign sentences and ensuring contextual accuracy. Research continues into tokenization and modeling strategies to address these concerns.
- Researchers initially used flex sensor-based gloves to map gestures to text, but accuracy limitations prompted the adoption of Convolutional Neural Networks (CNNs). To boost model reliability, a new dataset was curated, and training was enhanced to improve prediction capabilities across the full alphabet range.
- In another work, a mobile application was designed to convert Iraqi Sign Language gestures into Arabic speech using Android Studio. This platform-specific app aimed to promote communication among deaf individuals in Arabic-speaking regions.
- A wearable device incorporating flex sensors and accelerometers, named the E-Glove, was introduced for gesture recognition. This device wirelessly transmitted sensor data to a computer to display the corresponding text. It was noted for its affordability, simplicity, and usefulness in sign language education.
- Assistive technologies have the potential to revolutionize the learning environment for students with disabilities. These systems help integrate students with physical challenges into mainstream classrooms, enhancing educational access through tools that support their specific needs, rather than trying to correct the disabilities themselves.

DOI: 10.17148/IJIREEICE.2025.13602

- **A Two-Way Communication System** was developed by Areesha Gul et al., allowing both hearing-impaired and regular individuals to communicate without requiring knowledge of sign language. The system includes a wearable module for the deaf and a mobile app for the hearing individual, achieving an accuracy rate of 92.5%.
- **A Duplex Communication Framework** by Surbhi Rathi and Ujwalla Gawande implements gesture recognition for Indian Sign Language using techniques like skin color segmentation, Eigen features, and Euclidean classifiers. It supports bidirectional communication through text and speech.
- **An Inclusive Communication Aid** by Rajapandian et al. supports interaction among individuals with various disabilities such as blindness, deafness, or muteness. It bridges communication gaps through output formats like Braille, speech, or visual text, depending on the user's needs.
- **A Gesture-to-Voice Translator** by L. Anusha and Y. Usha Devi employs a trajectory recognition algorithm to convert gestures into English alphabets. The system features training, testing, and translation modes, utilizing Voice RSS and Microsoft Translate APIs for multilingual audio output.

VII. ALGORITHM

Natural Language Processing (NLP) is a branch of artificial intelligence (AI) that focuses on enabling computers to comprehend and interpret human language, whether it's written or spoken. NLP allows machines to process and analyze language data in a way that's similar to how humans understand language.

When humans interact with the world, they use senses like hearing and sight to perceive information, and their brains to process it. Similarly, computers use microphones to capture audio and programs to read text. These inputs are then processed by specific algorithms that convert them into a format a computer can work with—typically machine-readable code.

Artificial Neural Network (ANN) Algorithms:
One commonly used ANN architecture for image processing is the **Convolutional Neural Network (CNN)**. A CNN is composed of several key layers that work together to extract features and classify data. These layers include:

- **Convolutional Layer**
- **ReLU Activation Layer**
- **Pooling Layer**
- **Fully Connected Layer**

Step 1: Convolutional Layer
The convolutional layer is fundamental to CNNs. It applies a set of filters to the input image, producing feature maps that highlight specific patterns such as edges, corners, or textures. This layer helps the network understand spatial relationships in the data.

Step 2: ReLU Activation Layer
ReLU, or Rectified Linear Unit, is an activation function applied after convolution. It replaces all negative pixel values in the feature maps with zero, ensuring that the model remains computationally efficient and avoids vanishing gradients. ReLU activates only when the input is positive, allowing the network to introduce non-linearity.

Step 3: Pooling Layer
Pooling reduces the dimensions of the feature maps while retaining important information. It is applied after the activation layer. The pooling process typically involves:
1. Choosing a window size (e.g., 2x2 or 3x3)
2. Selecting a stride (commonly 2)
3. Sliding the window across the feature map
4. Extracting the maximum (Max Pooling) or average value within each window

This step helps to minimize computational complexity and prevent overfitting.

Step 4: Fully Connected Layer
The final layer in a CNN is the fully connected (FC) layer. Here, the output from the previous layers is flattened into a one-dimensional vector and passed through one or more dense layers. Each neuron in an FC layer is connected to every neuron in the previous layer, enabling the network to perform high-level reasoning and classification based on the features extracted earlier.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13602

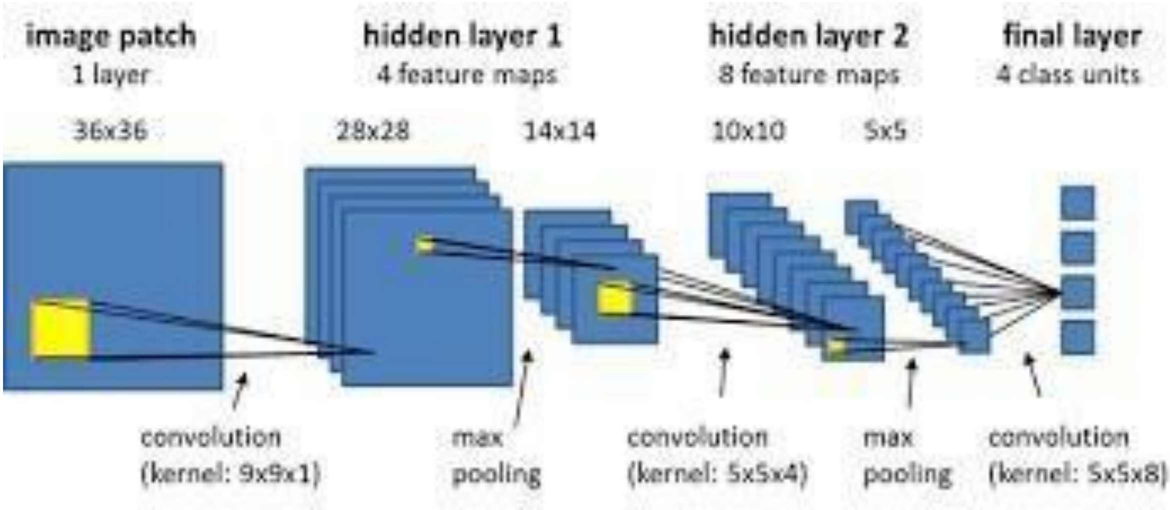

FIG.4 ANN

VIII. RESULTS DISCUSSION

The proposed system is designed to interpret hand gestures made by individuals with speech impairments. Unlike other methods that require users to wear specialized gloves or electromechanical devices, this framework operates without any physical attachments. It listens to spoken input and translates it into individual alphabets. For instance, if the word "riya" is spoken, the system processes it into letters – r, i, y, a – then retrieves the corresponding hand signs from a gesture database, making them visible for individuals who are deaf or hard of hearing.

This web-based application aims to build an automated translation mechanism that features a parser capable of converting speech or English text into a phrase structure grammar. This structure is then interpreted by another module that aligns with the grammatical rules of Indian Sign Language (ISL).

As part of this process, stop words are removed from the input. Since ISL lacks word inflections, the system uses techniques like stemming and lemmatization to reduce words to their root forms. After filtering the sentence, each word is checked against a database that maps words to their respective sign language video representations. If a particular word is not found, the system attempts to identify a synonym and use that instead.

This approach improves upon existing systems, which typically perform only direct word-to-sign translation without considering the grammatical structure of ISL. In contrast, this system restructures phrases according to ISL grammar rules, offering a more accurate and natural representation.

Being a web-based solution, the application is easily accessible and user-friendly. It is platform-independent, offering greater flexibility and enabling real-time translation of spoken or written phrases into Indian Sign Language.

XI. CONCLUSION AND FUTURE ENHANCEMENT

In conclusion, the Smart Translation System for the physically challenged, particularly the deaf and mute community, marks a significant step toward inclusive communication. By translating spoken audio into Indian Sign Language and vice versa, the system facilitates seamless interaction between hearing and non-hearing individuals. It leverages technologies such as Google Speech API, Natural Language Processing (NLP), and Artificial Neural Networks (ANN) to efficiently convert speech to text and text to sign language visuals, using predefined dictionaries and GIFs. The system's intuitive interface, real-time processing, and platform independence make it accessible and practical for educational, social, and daily use.

Moreover, the image/video to text/speech module enhances the learning and expressive abilities of users by recognizing Indian Sign Language gestures and converting them into understandable formats for the general population. The model, trained using deep learning techniques and robust datasets, ensures high accuracy and adaptability.

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 6, June 2025
DOI: 10.17148/IJIREEICE.2025.13602

With the increasing demand for assistive technologies, this dual-function platform not only bridges the communication gap but also empowers the deaf and mute community by promoting digital inclusivity, learning, and independence.

Furthermore, the system's modular architecture and use of modern machine learning frameworks ensure scalability and future enhancements. As more diverse datasets are incorporated and the model continues to learn, the accuracy and contextual understanding of sign translations are expected to improve significantly.

The project not only serves as a technological aid but also sets a foundation for further research and development in human-computer interaction for the differently-abled.

With continued refinement and integration into everyday platforms like mobile applications or educational tools, this smart translation system holds the potential to transform how society engages with the deaf and mute community.

While the current implementation of the Smart Translation System addresses core communication needs for the deaf and mute community, several enhancements can be introduced to improve functionality, scalability, and user experience:

- **Real-Time Video Translation**: Future versions can incorporate real-time video stream processing to recognize continuous sign language gestures, enabling more fluid and natural conversations without needing segmented inputs.
- **Expanded Language Support**: Currently focused on Indian Sign Language (ISL), the system can be expanded to support other regional and international sign languages, making it more inclusive for global users.
- **Emotion and Facial Expression Detection**: Integrating emotion recognition will help capture the tone and intent of sign language communication more accurately, as facial expressions are a crucial component of ISL.
- **Offline Functionality**: Developing offline capabilities using on-device models would allow users in remote or low-internet areas to benefit from the system without requiring a constant online connection.
- **Mobile App Integration**: A lightweight mobile application can be developed to offer on-thego accessibility, particularly beneficial for use in schools, workplaces, and public interactions.
- **Voice Customization and Multilingual Speech Output**: Users could select different voices and regional languages for the speech output to enhance personalization and accessibility across diverse linguistic backgrounds.
- **User Training Module**: An embedded learning module could be included to teach users basic sign language, promoting awareness and facilitating smoother communication between hearing and non-hearing individuals.
- **AI-Powered Grammar Correction**: Enhancing NLP components to not only translate but also correct grammar and syntax would improve the quality and clarity of communication output.
- **Gesture Speed Adaptation**: Implement adaptive algorithms that can recognize sign language performed at different speeds—whether slow or fast—making the system more robust and accommodating to individual signing styles and proficiency levels.
- **Context-Aware Translation**: Incorporating contextual understanding using AI can help the system interpret ambiguous signs based on surrounding words or phrases, resulting in more accurate and meaningful translations in real-world conversations.

REFERENCES

[1] A translator for American sign language to text and speech, 2016 IEEE 5th Global Conference on Consumer Electronics, 2016, pp. 1-2, doi: 10.1109/GCCE.2016.7800427.
[2] Morphological Analysis of Speech Translation into Indonesian Sign Language System (SIBI) on Android Platform, 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), 2019, pp. 205-210, doi: 10.1109/ICACSIS47736.2019.8980000.
[3] An Improved Sign Language Translation Model with Explainable Adaptations for Processing Long Sign Sentences Computational Intelligence and Neuroscience Volume 2020, Article ID 8816125.
[4] Convolutional Neural Network based Bidirectional Sign Language Translation System, 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 769-775, doi: 10.1109/ICSSIT48917.2020.9214272.
[5] N. El-Bendary, H. Zawbaa, M. Daoud, A. ella Hassanien, and K. Naka matsu, "Arslat: Arabic sign language alphabets translator," in Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on, Oct 2010, pp. 590–595.
[6] M. Melnyk, V. Shadrova, and B. Karwatsky, "Article: Towards computer assisted international sign language recognition system: A systematic survey," International Journal of Computer Applications, vol. 89, no. 17, pp. 44–51, March 2014, published by Foundation of Computer Science, New York, USA.
[7] K. Ellis and J. C. Barca, "Exploring Sensor Gloves for Teaching Children Sign Language," Advances in Human-Computer Interaction, p. 8, 2012.

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⧖ Peer-reviewed & Refereed journal ⧖ Vol. 13, Issue 6, June 2025
DOI: 10.17148/IJIREEICE.2025.13602

[8] K. Murakami and H. Taguchi, “Gesture recognition using recurrent neural networks,” in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., New York, NY, USA, 1991, pp. 237–242.
[9] P. R. V. Chowdary, M. N. Babu, T. V. Subbareddy, B. M. Reddy, and V. Elamaran, “Image processing algorithms for gesture recognition using MATLAB,” presented at the Int. Conf. Adv. Commun. Control Comput. Technol., Ramanathapuram, India, Jan. 2015.
[10] T. Khan and A. H. Pathan, “Hand gesture recognition based on digital image processing using MATLAB,” Int. J. Sci. Eng. Res., vol. 6, no. 9, pp. 338–346, Sep. 2015.

