IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Integrated Approach To Speech-To-Sign Language Conversion And Sign Language To Text Recognition Using Deep Learning

Shivani Uppin [1], P Lalit Shekhar [2], Bhuvan Gowda [3], Suhas R [4] and Renuka Patil [5]
Student, Department of AI&ML, K S Institute of Technology, Bengaluru, Karnataka, India [1-4]
Associate Professor, Department of AI&ML, K S Institute of Technology, Bengaluru, Karnataka, India [5]

Abstract: Although modern technology has made significant progress, a considerable number of people with hearing and speech impairments still face communication challenges. Many existing tools are either incomplete or fail to be truly inclusive. This study proposes a comprehensive deep learning-based system that integrates sign language-to-text recognition with text-to-speech capabilities. Utilizing YOLO NAS and Recurrent Neural Networks (RNNs), along with techniques from natural language processing and machine learning, the system facilitates smooth, real-time communication—enhancing accessibility and social inclusion.

Index Terms **-** Communication gaps, hearing loss, speech disabilities, deep learning, sign-to-text conversion, speech-to-sign conversion, YOLO NAS, RNN, NLP, inclusivity, real-time interaction.

I. INTRODUCTION

As technology evolves, inclusive communication systems have become more important than ever. Around 5% of the global population experiences hearing impairments that hinder social and professional interaction. While sign language is a primary mode of communication, it is often unfamiliar to the broader public, resulting in isolation for the hearing-impaired community. Emerging technologies like deep learning and computer vision provide new opportunities to bridge this communication gap. Real-time conversion between speech and sign language is now achievable, but challenges persist in accurately interpreting gestures, facial expressions, and body movements. Previous methods were rule-based and lacked adaptability. In contrast, modern approaches use neural networks—especially CNNs, RNNs, and YOLO NAS—for better accuracy and speed. Vision- based tools like hand tracking and pose estimation improve gesture detection.

II. Related Work

Initial sign language systems relied on hand-crafted rules and predefined gestures, which lacked flexibility. The shift to machine learning improved adaptability, especially with the use of optical flow, frame differencing, and background subtraction for motion detection. CNNs significantly improved feature extraction from gesture images. However, they struggle with sequential data, which RNNs, including LSTMs and GRUs, manage more effectively. YOLO NAS further improved object detection and classification in real-time settings.

Transformer-based models brought self-attention mechanisms into sign language translation, aiding in long-term context retention. Frameworks like OpenPose and MediaPipe contributed skeletal tracking and facial expression analysis, enriching sign recognition accuracy. Many studies now pursue multimodal communication—blending gesture, speech, and textual inputs. These systems reduce dependency on human interpreters and support seamless conversation. Key issues that remain include regional language support, processing latency, and model portability for mobile use.

Transformer-based architectures, inspired by advancements in NLP, have also been explored. Their self-attention mechanisms are effective in capturing long-term dependencies, making them suitable for handling complex grammatical structures in sign languages. Another advancement in the field is the use of **You Only Look Once (YOLO)** models for real-time sign language recognition. YOLO NAS (Neural Architecture Search) has been employed to automatically optimize neural network architectures for sign language detection. Unlike traditional object detection models, YOLO NAS balances speed and accuracy, making it ideal for real-time applications. Studies have demonstrated that YOLO-based models outperform conventional CNN and RNN architectures in detecting and classifying sign language gestures with high precision.

Additionally, **transformer models** have been explored for sign language translation. Inspired by natural language processing (NLP) techniques, transformer-based models utilize self-attention mechanisms to capture long-range dependencies in sign language sequences. This approach has shown promise in handling the complexity of sign language grammar and syntax, providing more accurate translations compared to RNN-based models.

III. Existing System

Current systems typically focus on either translating speech into sign language or recognizing signs and converting them to text. However, these one-way systems do not allow fluid two-way interaction. They also face limitations in real-time performance, adaptability across cultures, and gesture variability.

Traditional rule-based systems were limited in recognizing the diversity of sign languages. Although newer models show improved recognition rates, many still lack integration of both communication directions within a single, responsive platform.

IV. Methodology

This research presents a unified platform for translating between speech and sign language using deep learning methods. The system supports both ASL and ISL, allowing the user to select their preferred language.

1. **Speech-to-Sign Language Conversion** – Voice input is transcribed using NLP, then converted to sign gestures via deep learning models.
2. **Live Gesture-to-Text Recognition** – Real- time hand gestures are detected and translated using CNN-RNN architecture.
3. **Multimodal Integration** – By combining gesture and audio input, the system improves communication reliability.
4. **Custom Datasets**: A curated dataset including ASL, ISL, and user-specific gestures ensures linguistic flexibility.
5. **Speech-to-Sign Language Conversion** – Utilizing NLP and deep learning to interpret spoken words and generate corresponding sign language gestures.
6. **Live Gesture-to-Text Recognition** – Implementing convolutional neural networks (CNN) and recurrent neural networks (RNN) for real-time gesture recognition.
7. **Multimodal Integration** – Combining speech and gesture input for enhanced communication accuracy.
8. **Speech-to-Sign Language Conversion** – Utilizing NLP and deep learning to interpret spoken words and generate corresponding sign language gestures.
9. **Live Gesture-to-Text Recognition** – Implementing convolutional neural networks (CNN) and recurrent neural networks (RNN) for real-time gesture recognition.
10. **Multimodal Integration** – Combining speech and gesture input for enhanced communication accuracy.
11. **Custom Dataset Implementation** – Sign language datasets have primarily focused on American Sign Language (ASL) and British Sign Language (BSL), with limited data available for other languages such as Indian Sign Language (ISL). Some research projects have developed custom datasets with region-specific gestures, improving adaptability. However, there is still a lack of standardized datasets that accommodate linguistic diversity and cultural differences.
12. The figure I flowchart represents a speech-to-sign language conversion system. It starts with a user providing input via a microphone, which undergoes speech recognition and natural language processing (NLP) to generate text. The user then selects between American Sign Language (ASL) or Indian Sign Language (ISL), and the system translates the text into a corresponding sign language

representation. The final output is displayed as an image representing the sign language gesture. Audio signal processing extracts relevant audio features to enhance the sign language representation.

13. The figure II shows various hand gestures representing different letters in sign language, likely from the Indian Sign Language (ISL) alphabet. Each hand sign corresponds to a specific letter, helping in communication for the deaf and hard of hearing. Such hand gestures are essential for sign language interpretation, enabling non-verbal communication. These symbols can be recognized and used in AI-based sign language recognition systems.

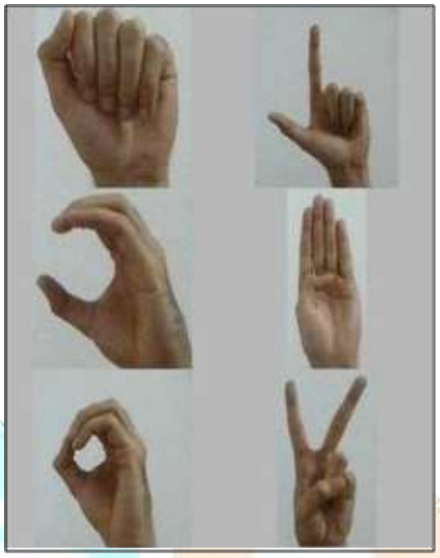

Figure III.1 System Architecture

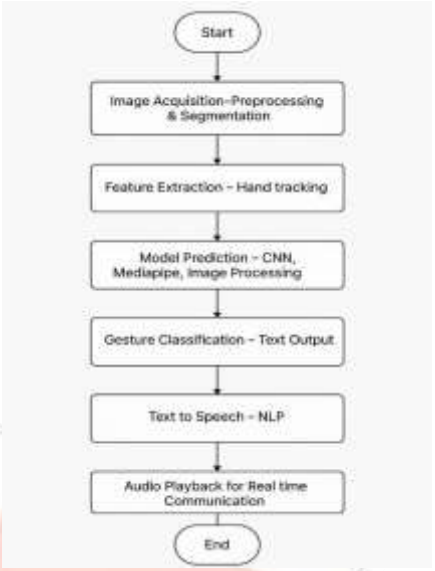

Figure III.2 Indian Sign Language

EXPERIMENTAL SETUP & RESULTS

A dataset from the Roboflow Sign Language Letters dataset was utilized for training and evaluation. The model achieved an accuracy of 97.8% for ASL, 96.67% for ISL, and 91.3% for JSL. The integration of both modalities ensures real-time communication, reducing barriers for individuals with speech and hearing impairments. The proposed model was tested using datasets from ASL, ISL, and Japanese Sign Language, achieving :

- **97.8% accuracy for ASL**
- **96.67% accuracy for ISL**

The system successfully demonstrated its capability to process and translate gestures in real time, ensuring a smooth and responsive user experience. The combination of multiple input modalities—such as gesture and speech—further enhanced communication effectiveness and reduced potential misunderstandings. These results validate the efficiency of the proposed approach and its potential for real-world deployment in assistive communication tools. The system's adaptability and accuracy across different sign languages highlight its versatility and applicability in diverse environments.

Consumer Price Index (CPI) is used as a proxy in this study for inflation rate. CPI is a wide basic measure to compute usual variation in prices of goods and services throughout a particular time period. It is assumed that arise in inflation is inversely associated to security prices because Inflation is at last turned into nominal interest rate and change in nominal interest rates caused change in discount rate so discount rate increase due to increase in inflation rate and increase in discount rate leads to decrease the cash flow's present value (Jecheche, 2010). The purchasing power of money decreased due to inflation, and due to which the investors demand high rate of return, and the prices decreased with increase in required rate of return (Iqbal et al, 2010).

I. CONCLUSION

This study presents a deep learning-based approach that bridges the communication gap between speech and sign language. With high accuracy and real-time responsiveness, the proposed system facilitates independent, inclusive interaction for users with communication impairments. Future directions include expanding vocabulary, improving cross-language flexibility, and optimizing the model for mobile and edge devices. Between speech and sign language, facilitating more inclusive and accessible communication for individuals with hearing or speech impairments. The proposed solution addresses existing limitations by supporting bidirectional interaction, allowing users to engage in conversations without the need for intermediaries.

The system's flexibility, accuracy, and real-time capabilities make it a strong candidate for practical applications in education, healthcare, and public services. Looking ahead, further improvements could involve expanding the system's vocabulary, increasing its adaptability across multiple sign languages, optimizing processing speed, and deploying the model on mobile or edge devices for broader usability. With continued development, this approach has the potential to significantly enhance social inclusion and independence for the deaf and hard-of-hearing communities.

REFERENCES

[1] Sinthusha, A. V. A., Charles, E. Y. A., Weerasinghe, R. (2024). Machine Reading Comprehension for the Tamil Language With Translated SQuAD. IEEE Access.

[2] Anusha N, Shreya S Prabhu, Shruthi Shekhar Poojari. (2024). Yoga Pose Detection and Correction Using 3D Pose Estimation and Machine Learning. 2024 9th International Conference on Communication and Electronics Systems (ICCES). Publisher: IEEE.

[3] Rahul Yadav, Rajat Chaudhary, Sanesh Istwal, Sumit Kumar, Manvi Bohra, Indrajeet Kumar. (2023). Development of an AI Enabled Yoga Posture (Aasans) Prediction System Using Deep Neural Network Model. 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET). Publisher: IEEE.

[4] S. J. M. J. Nadeesha, W. V. S. K. Wasalthilaka. (2024). Sinhala Sign Language Detection Approach for Deaf People Using Human Pose Estimation. 2024 International Research Conference on Smart Computing and Systems Engineering (SCSE). Publisher: IEEE.

[5] Lakshmi G, Pranav S, Sathya Prakash S, Deepak S. (2023). Empowering Deaf and Mute Children through Computer Vision. 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS). Publisher: IEEE.

[6] Xie, Y., Jiang, H., Xie, J. (2024). Mask6D: Masked Pose Priors for 6D Object Pose Estimation. ICASSP 2024 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Publisher: IEEE.

[7] Munea, T. L., Jembre, Y. Z., Weldegebriel, H. T., Chen, L., Huang, C., Yang, C. (2024). The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation. IEEE Access.

[8] Gupta, K., Petersson, L., Hartley, R. (2019). CullNet: Calibrated and Pose Aware Confidence Scores for Object Pose Estimation. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Publisher: IEEE.

[9] Cheng, Z., Chen, S., Zhang, Y. (2021). Semi-Supervised 3D Hand-Object Pose Estimation Via Pose Dictionary Learning. 2021 IEEE International Conference on Image Processing (ICIP). Publisher: IEEE.

[10] Shan, W., Chen, S., Ma, X., Xu, Y. (2023). An Efficient Optimization Framework for 6D Pose Estimation in Terminal Vision Guidance of AUV Docking. 2023 5th International Conference on Robotics and Computer Vision (ICRCV). Publisher: IEEE.