International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)

AI-ASSISTED VISUAL COMMUNICATION FOR INCLUSIVE HUMAN INTERACTION

[1] Dr. M. Navaneetha Krishnan, Professor & Head of Department

Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu

Email Id – mnksjce@gmail.com

[2] Ms. Shiny Devadoss, Student

Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu

Email Id – leninn065@gmail.com

Abstract – **This project presents an AI-based system designed to support real-time sign language interpretation and learning. The system enables seamless communication by translating sign language into understandable spoken output. It supports multilingual speech generation, allowing the interpreted signs to be conveyed in multiple languages. In addition to real-time interpretation, the project includes an interactive learning module that helps users understand and practice sign language visually. The solution combines intelligent gesture recognition with speech and visual assistance to improve accessibility. It focuses on ease of use, real-time performance, and inclusivity. By integrating interpretation and learning in a single platform, the system benefits both learners and users with hearing or speech impairments. The proposed approach enhances human–computer interaction through natural communication methods. The project aims to reduce communication barriers and promote wider adoption of sign language.**

I. Introduction

Effective communication is a fundamental requirement for social interaction, education, and professional growth. However, individuals with hearing or speech impairments often face significant challenges in communicating with the hearing population due to the limited understanding of sign language. Although sign language is a well-established mode of communication, its adoption remains restricted, creating a communication gap in daily life situations such as classrooms, workplaces, and public services.

With recent advancements in artificial intelligence, machine learning, and computer vision, it has become possible to develop systems capable of understanding human gestures in real time. These technologies enable the interpretation of sign language through visual inputs, reducing the dependency on human interpreters.

The proposed project leverages these advancements to build an AI-based sign language interpretation system that can recognize hand gestures and convert them into meaningful speech output.

In addition to real-time interpretation, there is a growing need for user-friendly platforms that support sign language learning. To address this, the project incorporates a learning module where users can input text and view corresponding sign language demonstrations through videos. This dual functionality of interpretation and learning makes the system useful not only for people with disabilities but also for students, educators, and individuals interested in learning sign language.

II. Scope of the Project

This project focuses on building an intelligent application that understands hand gestures and presents the interpreted meaning in an accessible format. It emphasizes real-time operation, user-friendly interaction, and language flexibility. The system also provides a visual learning environment to support sign language education. The solution is intended for assistive communication and educational use cases. Its modular design allows future enhancement without major structural changes.

III. Objective

The project aims to:

- Improve prediction reliability and consistency using regularization-based classification techniques.
- Design a learning module that processes user-provided text using NLP techniques such as tokenization and normalization, mapped to appropriate sign language visual representations.
- Optimize feature handling and model performance for real-time execution.
- Integrate robust classification and NLP-based learning techniques into a unified intelligent system.
- Improve communication and accessibility for people with hearing and speech impairments.

IV. Background and Motivation

A. Overview

Communication is one of the most essential aspects of human interaction, enabling people to express ideas, emotions, needs, and thoughts effectively. For individuals with hearing and speech impairments, sign language serves as the primary medium of communication. Although sign language is widely used within the deaf and mute community, many people outside this community are unfamiliar with it, creating a major communication barrier in everyday life. These developments inspired the design of an intelligent sign language recognition system that can automatically interpret gestures from live video feeds and convert them into meaningful outputs. Such systems have the potential to revolutionize assistive communication technologies.

B. Importance of AI – Based Sign Language

The motivation behind this project arises from the need to create a more inclusive society where communication is accessible to everyone regardless of physical limitations. Many hearing-impaired individuals face challenges while interacting with teachers, healthcare professionals, public service providers, and ordinary people who do not understand sign language. A technological solution capable of translating sign gestures into understandable text or speech can significantly reduce these challenges and improve social integration.

Recent advancements in artificial intelligence, computer vision, and deep learning have opened new possibilities for gesture recognition systems. Cameras and machine learning models can now analyze human hand movements and facial expressions with high accuracy in real time. These developments inspired the design of an intelligent sign language recognition system that can automatically interpret gestures from live video feeds and convert them into meaningful outputs. Such systems have the potential to revolutionize assistive communication technologies.

Another important motivation for this project is the growing demand for real-time and contactless communication systems. Traditional communication assistance methods often require human interpreters, which may not always be available or affordable. By automating sign language recognition using image processing and machine learning algorithms, the proposed system can provide instant communication support without human intervention. This makes the solution more scalable, cost-effective, and accessible for daily use.

The project is also motivated by the need to improve accessibility in educational institutions and workplaces. Students with hearing impairments may struggle to communicate effectively with teachers and classmates, while employees may face barriers during professional interactions. An automated sign language interpretation system can support better inclusion by enabling smooth communication between differently abled individuals and the wider community. This can enhance confidence, participation, and equal opportunities for users.

Furthermore, the increasing availability of affordable cameras, mobile devices, and embedded computing platforms has made it practical to develop portable and real-time recognition systems. Integrating machine learning models with modern hardware allows the system to operate efficiently with minimal latency. This technological feasibility motivates the implementation of a user-friendly application that can be deployed in homes, schools, hospitals, customer service centers, and public spaces.

The project also seeks to contribute to the research and development of assistive technologies. Many existing systems are limited by low accuracy, restricted gesture datasets, or dependence on controlled environments. By applying advanced feature extraction techniques and robust classification algorithms, the proposed project aims to improve recognition performance under varying lighting conditions and hand orientations. This contributes to the advancement of intelligent human-computer interaction systems.

V. Problem Statement

In today's digital era, effective communication stands as a fundamental human right, yet individuals within the deaf and hard-of-hearing community encounter notable barriers. Among these challenges, the limited accessibility of sign language interpretation services remains prominent, often due to factors such as high costs, limited availability, or lengthy delays. Additionally, the diverse range of global sign languages further complicates the provision of accurate interpretation services.

Leveraging cutting-edge technologies like computer vision and natural language processing, there is an opportunity to develop real-time and cost-effective sign language interpretation systems. These systems must possess the capability to recognize and interpret diverse sign languages accurately, ultimately translating them into spoken or written language, empowering individuals within the deaf and hard-of-hearing community and promoting equal participation in society.

VI. Literature Survey

Paper I

Title: An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion

Author: Hamzah Luqman

Year: 2022

Abstract: This paper proposes a trainable deep learning network for isolated sign language recognition that captures spatiotemporal information using a small number of sign frames. The system comprises a hierarchical sign learning module with Dynamic Motion Network (DMN), Accumulative Motion Network (AMN), and Sign Recognition Network (SRN). It outperformed other techniques by 15% on the KArSL-190 dataset in signer-independent mode.

Disadvantages: Resource-intensive training of the hierarchical network; key posture ambiguity causes errors in fast or unclear signing.

Paper II

Title: Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation

Author: B. Natarajan, E. Rajalakshmi, R. Elakkiya, Ketan Kotecha.

Year: 2025

Abstract: This paper introduces novel approaches for a complete framework handling sign language recognition, translation, and production tasks in real time. It uses MediaPipe and a hybrid CNN + Bi-LSTM model achieving above 95% classification accuracy. The model secured a 38.06 average BLEU score with remarkable visual quality metrics.

Disadvantages: High computational requirements; dependent on quality of MediaPipe landmark extraction.

VII. Proposed System

The proposed system is an AI-based platform designed to interpret sign language in real time and provide a learning environment for users. It uses a webcam to capture hand gestures and applies the Media Pipe framework to extract precise hand landmark features. These features are processed using a Ridge Classifier, enabling accurate recognition of a wide range of gestures. Once a gesture is predicted, the system generates multilingual speech output, allowing communication across different languages. The system also incorporates a learning module where textual input is processed through NLP techniques such as tokenization and normalization, and mapped to corresponding sign language videos for interactive learning.

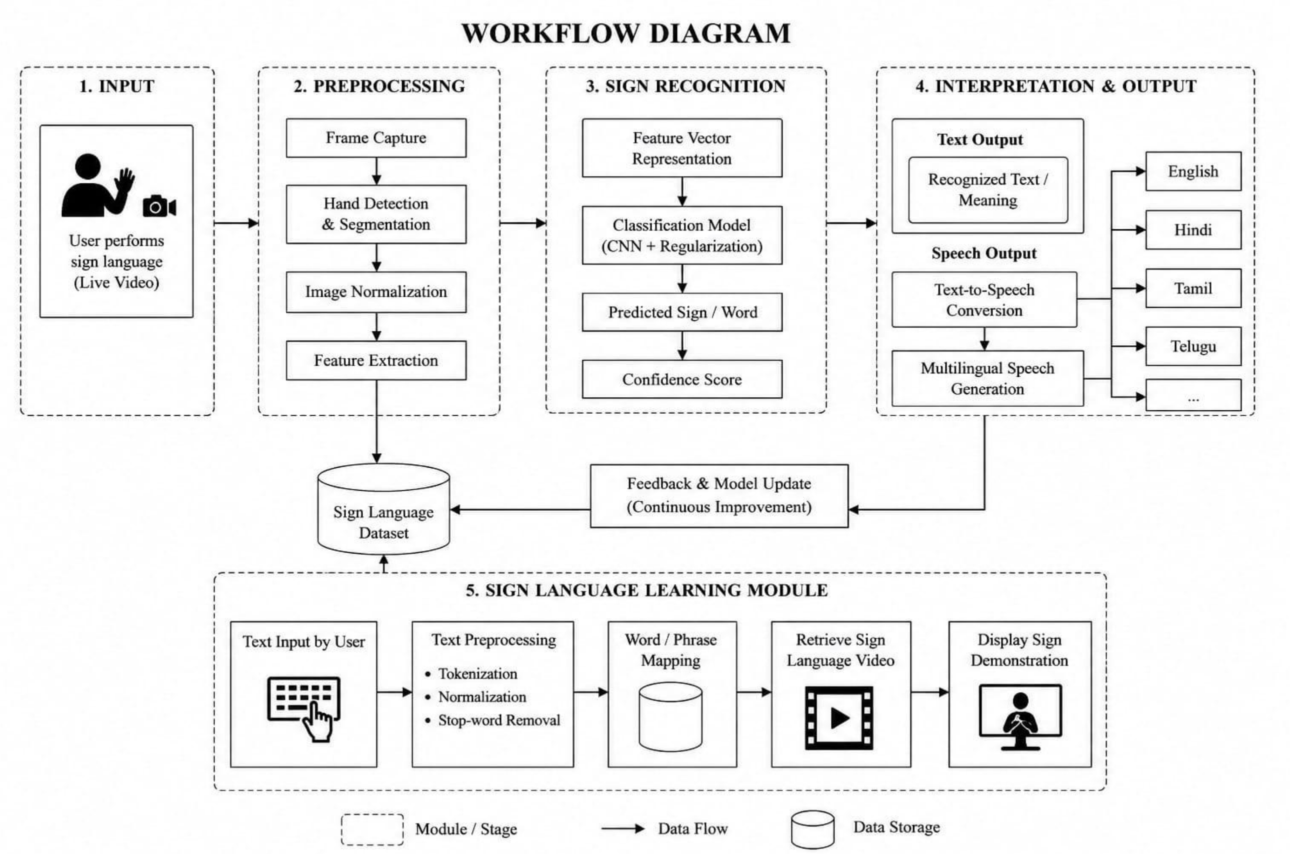

Figure: Workflow of the System

The proposed system is an AI-based real-time sign language recognition and learning platform designed to improve communication between hearing- and speech-impaired individuals and the general public. The system combines computer vision, machine learning, and natural language processing techniques to recognize hand gestures, convert them into meaningful speech output, and provide an interactive sign language learning environment. The main objective of the system is to eliminate communication barriers and promote inclusivity through intelligent automation. The system uses a webcam to capture live video streams containing hand gestures performed by the user. Unlike traditional systems that depend on sensor gloves or specialized hardware devices, the proposed solution works using standard camera equipment, making it affordable, portable, and easy to deploy. The captured frames are continuously analyzed in real time to detect hand movements and gesture patterns accurately.

To achieve precise hand detection and tracking, the system utilizes the MediaPipe framework. MediaPipe is an advanced computer vision library capable of identifying multiple hand landmarks with high accuracy. It extracts important key points from the hand, such as finger joints, fingertip positions, and palm coordinates. These landmarks act as feature vectors that help the system understand gesture orientation, movement, and finger positioning effectively.The extracted hand landmark features are then processed using the Ridge Classifier algorithm, which serves as the core machine learning model of the system. Ridge Classification is an extension of Ridge Regression adapted for classification tasks. It transforms the classification problem into a regression-based approach and predicts gesture classes based on continuous output values. This algorithm is particularly suitable for high-dimensional feature datasets generated from hand landmark coordinates.

The Ridge Classifier applies L2 regularization to improve prediction stability and reduce overfitting. The cost function used by the algorithm is represented as:

$$\text{Cost} = \| y - Xw \|^2 + \alpha \| w \|^2$$

where $y$represents the class label vector, Xdenotes the feature matrix, $w$represents the weight vector, and $\alpha$is the regularization parameter. The regularization component helps the model manage correlated features efficiently and improves the overall generalization capability during real-time prediction.

One of the major advantages of using the Ridge Classifier is its ability to handle multicollinearity among features extracted from hand landmarks. Since many gesture features are interrelated, the regularization mechanism stabilizes the learning process and ensures reliable predictions. Additionally, the algorithm is computationally efficient, making it suitable for real-time applications where low latency and fast response times are essential.

After gesture recognition, the predicted output is converted into text and multilingual speech using a text-to-speech module. This feature enables users to communicate across different languages and improves accessibility in multilingual environments. The generated speech output allows non-sign-language users to understand the intended message immediately, thereby enhancing communication effectiveness.

The proposed system also includes a dedicated learning module for users interested in learning sign language. In this module, textual input provided by the user undergoes Natural Language Processing (NLP) operations such as tokenization, normalization, and text mapping. The processed text is then linked to corresponding sign language videos or gesture demonstrations, creating an interactive and educational learning experience. The inclusion of the learning module makes the system beneficial not only for hearing-impaired individuals but also for students, teachers, caregivers, and the general public. Users can practice sign language through visual demonstrations and improve their understanding of gestures gradually. This educational capability increases awareness of sign language and supports inclusive communication practices in society.

The system is designed to support efficient processing of high-dimensional data while maintaining recognition accuracy and performance. The combination of MediaPipe-based feature extraction and Ridge Classification enables stable operation even under varying lighting conditions and hand orientations. The lightweight computational requirements also make the system suitable for deployment on laptops, embedded devices, and mobile platforms. Although the Ridge Classifier offers several advantages such as simplicity, scalability, and fast execution, it also has certain limitations. The algorithm assumes a linear relationship between features and output classes, which may reduce flexibility compared to nonlinear models such as Support Vector Machines or Decision Trees. Additionally, it does not directly provide probability estimates for predictions. However, despite these limitations, the algorithm remains highly effective for real-time sign language recognition because of its reliability, efficiency, and ability to handle large feature sets effectively.

VIII. Conclusion

The AI-based Sign Language Interpretation and Learning System developed in this project demonstrates the effective integration of computer vision, machine learning, and natural language processing techniques. By leveraging the Ridge Classifier, the system achieves accurate real-time recognition of hand gestures, ensuring reliable interpretation of sign language. The inclusion of multilingual text-to-speech conversion enhances accessibility, allowing users from diverse linguistic backgrounds to understand the gestures.

Furthermore, the learning module provides an interactive platform where textual input can be mapped to sign language videos, supporting structured and effective learning. Overall, this project not only bridges communication gaps for individuals with hearing or speech impairments but also promotes sign language education. The proposed system is scalable, user-friendly, and can serve as a foundation for future enhancements in assistive and educational technologies.

IX. Future Enhancement

The rapid advancement of artificial intelligence, computer vision, and deep learning technologies, the system can be further enhanced to achieve greater accuracy, scalability, and usability. Future improvements can focus on expanding gesture recognition capabilities, supporting multiple sign languages, integrating advanced AI models, and developing mobile and cloud-based solutions. These enhancements will help transform the system into a more intelligent, accessible, and user-friendly communication platform suitable for real-world applications across education, healthcare, public services, and social interaction environments.

- Support for dynamic gesture recognition (time-based gestures).
- Recognition using hand movement, pose, and facial expressions.
- Integration of Text-to-Speech features for improved accessibility.
- Support for multiple languages using Google API.
- Cross-platform development using MediaPipe (including iOS).
- Mobile application development.
- Improve accuracy using deep learning architectures.

REFERENCES:

[1] H. Luqman, "An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion," IEEE Access, vol. 10, pp. 93785–93798, 2022.

[2] B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," IEEE Access, vol. 10, pp. 104358–104374, 2022.

[3] A. S. M. Miah, M. A. M. Hasan, Y. Tomioka and J. Shin, "Hand Gesture Recognition for Multi-Culture Sign Language Using Graph and General Deep Learning Network," IEEE Open Journal of the Computer Society, vol. 5, pp. 144–155, 2024.

[4] T. Shanableh, "Two-Stage Deep Learning Solution for Continuous Arabic Sign Language Recognition Using Word Count Prediction and Motion Images," IEEE Access, vol. 11, pp. 126823–126833, 2023.

[5] O. El Ghoul, M. Aziz and A. Othman, "JUMLA-QSL-22: A Novel Qatari Sign Language Continuous Dataset," IEEE Access, vol. 11, pp. 112639–112649, 2023.

[6] S. Sharma and S. Singh, "Vision-based hand gesture recognition using deep learning for the interpretation of sign language," Expert Syst. Appl., vol. 182, Nov. 2021.

[7] T. Kapuscinski and K. Inglot, "Vision-based gesture modeling for signed expressions recognition," Proc. Comput. Sci., vol. 207, pp. 1007–1016, 2022.

[8] D. Tasmere, B. Ahmed, and S. R. Das, "Real time hand gesture recognition in depth image using CNN," Int. J. Comput. Appl., vol. 174, no. 16, pp. 28–32, 2021.

[9] S. B. Abdullahi and K. Chamnongthai, "American sign language words recognition of skeletal videos using processed video driven multi-stacked deep LSTM," Sensors, vol. 22, no. 4, p. 1406, 2022.

[10] S. B. Abdullahi and K. Chamnongthai, "American sign language words recognition using spatio-temporal prosodic and angle features," IEEE Access, vol. 10, pp. 15911–15923, 2022.