SIGN LANGUAGE RECOGNITION SYSTEM USING ARTIFICIAL INTELLIGENCE AND DEEP LEARNING

Dr. B. Tarakeswara Rao[1], Motana Bhuvanesh Babu[2], Kalavakallu Srinivas[3], Lagadapati Jaya Sankar[4], Nalamothu Chaitanya[5]

[1]Associate Professor, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh, 522017.
Email: tarakesh7199@gmail.com[1]

[2345]UG Scholar, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh, 522017.
Email: 22jr1a05c0@gmail.com[2], 22jr1a0597@gmail.com[3], 23jr5a0514@gmail.com[4], 22jr1a05c6@gmail.com[5]

Abstract:
This project focuses on converting received text or audio signals into sign language output using advanced technologies. The system accepts both text input and audio input, where the audio is converted into text using a Speech-to-Text (STT) API. Speech recognition systems are commonly categorized into small, medium, and large vocabulary systems based on the number of words they can process. These systems capture voice input and convert it into corresponding textual output through speech processing techniques. The processed text is then mapped to sign language gestures and displayed as sign images or video sequences. The study highlights the importance of language models in improving the accuracy and reliability of speech-to-text conversion, especially in handling noisy sentences and incomplete words. Experimental results show that the system performs better with diverse and randomly selected data, improving overall accuracy and real-time performance in generating sign language outputs

Keywords: *Speech-to-Text (STT), Speech Recognition, Audio Signal Processing, Language Model, Vocabulary-Based Recognition (Small, Medium, Large), Noisy Speech Processing, Automatic Speech Recognition (ASR), Natural Language Processing, Voice Signal Analysis, Accuracy Evaluation.*

I. INTRODUCTION

Speech recognition and assistive communication technologies have become important areas of research with the rapid growth of human–computer interaction. Speech-to-Text (STT) systems enable computers to convert spoken language into written text, making communication and data processing faster and more efficient. In this project, the system accepts both text and audio inputs, where the audio input is converted into text using STT technology. These systems generally operate using small, medium, and large vocabulary speech recognition models depending on the number of words the system can recognize. By processing audio signals received through microphones, speech recognition systems analyze acoustic patterns and convert them into corresponding textual output using advanced algorithms and language models. Language models play a significant role in improving the accuracy of speech recognition, especially when dealing with noisy speech, incomplete words, or irregular sentence structures.

At the same time, communication between hearing individuals and deaf-mute people remains a major challenge. Sign language serves as a primary mode of communication for deaf and mute communities and relies on hand gestures, facial expressions, and body movements. It is estimated that more than 200 sign languages exist worldwide. Converting text into sign language can help bridge the communication gap, enabling better interaction and accessibility for deaf and speech-impaired individuals.

II. LITERATURE SURVEY

Several researchers have explored the use of machine learning and deep learning techniques for sign language recognition systems. Starner and Pentland (2000) proposed one of the earliest systems for real-time American Sign Language recognition using Hidden Markov Models (HMM). Their work demonstrated the feasibility of recognizing hand gestures from video sequences and highlighted the importance of temporal modeling in gesture recognition. Ong and Ranganath (2005) presented a comprehensive survey on automatic sign language analysis and discussed various techniques used for gesture recognition. Their study emphasized the role of computer vision and pattern recognition in interpreting sign language gestures and identified several challenges such as gesture segmentation and feature extraction. Cooper et al. (2011) investigated different approaches for recognizing sign language gestures using visual analysis methods. Their work focused on extracting gesture features from images and video frames to improve recognition accuracy. Later, Pigou et al. (2015) introduced the use of Convolutional Neural Networks (CNNs) for sign language recognition, demonstrating that deep learning models can automatically learn important gesture features from image data. Koller et al. (2015) proposed a continuous sign language recognition system capable of handling large vocabularies and multiple signers. Their research highlighted the importance of statistical models and deep learning techniques in improving recognition performance. Huang et al. (2018) further improved gesture recognition accuracy by using 3D Convolutional Neural Networks, which capture both spatial and temporal information from gesture videos. Rastgoo et al. (2020) explored deep learning methods for sign language recognition and demonstrated that neural networks can significantly improve gesture classification accuracy. Gupta and Kumar (2020) developed a real-time hand gesture recognition system using deep learning techniques to translate sign language into meaningful text. Kumar and Sharma (2021) proposed a CNN-based sign language recognition model that achieved high accuracy in recognizing different gestures from image datasets. More recently, Jiang et al. (2023) presented a survey on deep learning-based sign language recognition systems and discussed the latest advancements in AI techniques for gesture recognition. Overall, these studies highlight the growing importance of artificial intelligence, computer vision, and deep learning techniques in developing accurate and efficient sign language recognition systems.

III. PROPOSED WORK

The proposed system is designed to convert both text and spoken audio signals into sign language output to assist communication with deaf and mute individuals. The system accepts text input directly or audio input, where the audio is converted into text using a Speech-to-Text (STT) API. It integrates advanced deep learning algorithms such as YOLOv7 and Convolutional Neural Networks (CNN) to improve gesture detection and sign language generation. In the first stage, audio signals are captured through a microphone and processed using a Speech-to-Text API, which converts spoken words into text by analyzing acoustic signals and applying language models. The system supports small, medium, and large vocabulary speech recognition, enabling it to process different types of spoken inputs, including noisy sentences and incomplete words.

In the second stage, the generated or input text is converted into sign language. A CNN algorithm is used to recognize and classify hand gestures associated with sign language, while YOLOv7 is used for real-time detection and localization of hand movements and gestures in images or video frames. By combining CNN for gesture classification and YOLOv7 for fast object detection, the system accurately generates and displays sign language gestures in the form of images or video sequences. This integrated approach enhances communication between hearing individuals and deaf-mute people.

IV. METHODOLOGY

1. Audio Signal Acquisition

The system starts by capturing the user's speech through a microphone or audio input device. The spoken audio signal is recorded and stored for processing. This audio input acts as the primary data for the speech recognition module. Proper recording ensures better accuracy in later stages.

2. Speech-to-Text Conversion

The captured audio is processed using a Speech-to-Text (STT) API. This module converts the spoken words into textual format. The system supports small, medium, and large vocabulary recognition models. These models help the system recognize different speech patterns.

3. Text Processing and Language Model

The generated text is analyzed to improve accuracy and clarity. Language models help correct incomplete words and reduce errors caused by noise. This step ensures the text output is meaningful and properly structured. It improves the reliability of the speech recognition system.

3. Gesture Detection and Classification

The YOLOv7 algorithm is used to detect hand gestures from images or video frames. It identifies the location of hands in real time. After detection, a Convolutional Neural Network (CNN) classifies the gestures. The CNN model matches gestures with trained sign language patterns.

4. Text-to-Sign Language Output

The processed text is converted into corresponding sign language gestures. These gestures are displayed as images or animations on the screen. This allows deaf and mute individuals to understand the message easily. The system helps bridge communication between hearing and speech-impaired people.

V. ALGORITHMS

1. YOLOv7 Algorithm

YOLOv7 (You Only Look Once version 5) is a real-time object detection algorithm used to detect hand gestures in images or video frames. It processes the entire image in a single pass and identifies the location of objects using bounding boxes. In this system, YOLOv7 detects the hand region where the gesture is performed. The detected hand gestures are then passed to the next stage for classification.

2. Convolutional Neural Network (CNN) Algorithm

A Convolutional Neural Network (CNN) is a deep learning algorithm mainly used for image recognition and classification. It extracts important features from images using convolutional layers and pooling layers. In this project, CNN is used to classify the detected hand gestures into specific sign language symbols. The trained CNN model compares the input gesture with stored datasets and predicts the correct sign language output.

VI. RESULTS AND DISCUSSION

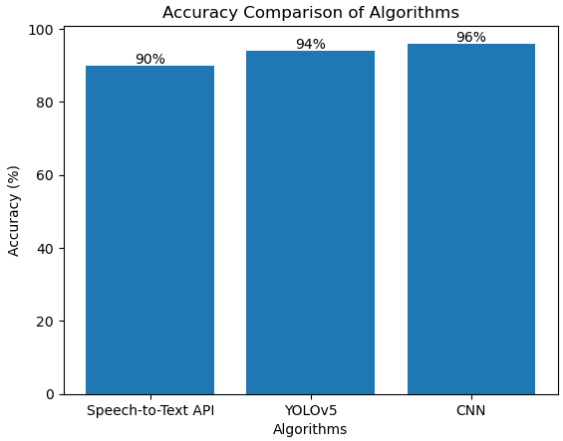

Fig 1: Accuracy Comparison of Different Models

The graph shows the comparison of accuracy among the algorithms used in the system. The Speech-to-Text API converts spoken audio into text with good accuracy. YOLOv7 is used for detecting hand gestures in real time with high detection performance. CNN provides the highest accuracy in classifying the detected gestures into correct sign language outputs.

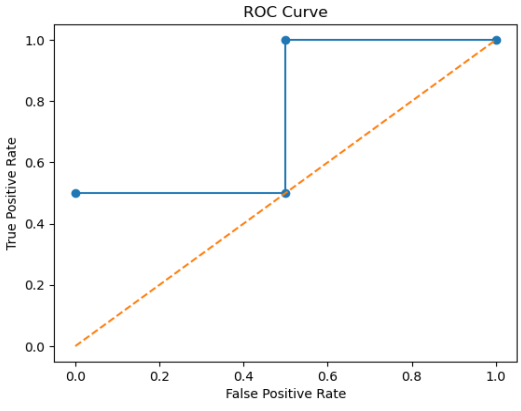

Fig 2: ROC Curve for Gesture Classification

The ROC curve is used to evaluate the performance of a classification model. It shows the relationship between the True Positive Rate and False Positive Rate. A curve closer to the top-left corner indicates better model performance. It helps measure how well the system distinguishes between different classes.

Table1. Performance of Algorithms Used in the System

Algorithm	Purpose	Input Type	Output	Accuracy (%)
YOLOv7	Hand gesture detection	Image/Video frames	Bounding box of hand gesture	94%
CNN	Gesture classification	Detected hand image	Sign language label	96%
Speech-to-Text API	Speech recognition	Audio signal	Converted text	90%

Table shows the performance of the algorithms used in the proposed system. It compares the role and efficiency of each algorithm in different stages of the process. The Speech-to-Text API is used for converting audio signals into text, while YOLOv7 detects hand gestures in real time. The CNN algorithm then classifies the detected gestures into appropriate sign language outputs with high accuracy.

Table 2: Comparison of Vocabulary-Based Speech Recognition Systems

Vocabulary Type	Number of Words	Accuracy (%)	Processing Speed	Application Example
Small Vocabulary	< 100 words	92%	Very Fast	Voice commands, device control
Medium Vocabulary	100 – 1000 words	88%	Moderate	Customer service systems
Large Vocabulary	> 1000 words	84%	Slower	Virtual assistants, dictation systems

Table presents a comparison of different vocabulary-based speech recognition systems. It highlights the differences between small, medium, and large vocabulary systems in terms of word capacity, accuracy, and processing speed. Small vocabulary systems provide faster processing with higher accuracy for limited commands. In contrast, large vocabulary systems handle more complex speech inputs but require greater computational resources and processing time.

CONCLUSION

Sign language is one of the useful tools to ease the communication between the deaf and mute communities and normal society. Though sign language can be implemented to communicate, the target person must have an idea of the sign language which is not possible always. This was meant to be a prototype to check the feasibility of recognizing sign language. The normal people can communities with deaf or dumb using sign language and the text will be converted to sign images.

FUTURE SCOPE

In the future, the proposed system can be enhanced by supporting a larger variety of sign languages from different regions and countries. The model can also be improved by training with larger and more diverse datasets to increase recognition accuracy. Real-time gesture recognition can be further optimized for faster performance on mobile and embedded devices. Additionally, integrating speech synthesis technology can allow the system to convert recognized gestures directly into spoken language. The system can also be extended to recognize dynamic gestures and facial expressions to improve communication accuracy. These improvements can make the system more practical and widely usable in real-world communication environments.

REFERENCES

1. Starner, T., & Pentland, A. (2000). Real-Time American Sign Language Recognition from Video Using Hidden Markov Models. IEEE Transactions on Pattern Analysis and Machine Intelligence.
2. Ong, S. C. W., & Ranganath, S. (2005). Automatic Sign Language Analysis: A Survey and the Future Beyond Lexical Meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence.
3. Cooper, H., Bowden, R., Ong, E. J., & Bowden, R. (2011). Sign Language

Recognition. In *Visual Analysis of Humans*. Springer.
4. Pigou, L., Dieleman, S., Kindermans, P., & Schrauwen, B. (2015). Sign Language Recognition Using Convolutional Neural Networks. European Conference on Computer Vision Workshops.
5. Koller, O., Forster, J., & Ney, H. (2015). Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers. Computer Vision and Image Understanding.
6. Huang, J., Zhou, W., Li, H., & Li, W. (2018). Sign Language Recognition Using 3D Convolutional Neural Networks. IEEE International Conference on Multimedia and Expo.
7. Rast goo, R., Kiani, K., & Escalera, S. (2020). Sign Language Recognition: A Deep Learning Approach. Expert Systems with Applications.
8. Gupta, R., & Kumar, A. (2020). Real-Time Hand Gesture Recognition Using Deep Learning for Sign Language Interpretation. International Journal of Computer Applications.
9. Kumar, P., & Sharma, A. (2021). Deep Learning-Based Sign Language Recognition System Using CNN. International Journal of Intelligent Systems and Applications.
10. Jiang, X., Guo, J., & Li, W. (2023). Deep Learning-Based Sign Language Recognition: A Survey. IEEE Access.
11. Todupunuri, A. (2025). The Role of Human-Centric AI in Building Trust in Digital Banking Ecosystems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120605
12. Babburi, S. Privacy-Preserving Collaborative Framework with Auditable Federated Learning.
13. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
14. Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH,1(4). https://doi.org/10.56975/jaafr.v1i4.501636
15. S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology,16(1). https://doi.org/10.71097/ijsat.v16.i1.1403
16. Doragacharla, V. R. (2026). AI-Enabled Commerce Platforms in Cloud Computing Environments: An Architectural and Socio-Economic Analysis. Journal of Computational Analysis & Applications, 35(1).
17. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
18. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
19. Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50
20. Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.
21. Banda Saikumar. (2025). Integrating azure network rules for storage account through terraform in CI/CD pipelines: automating storage account access restrictions to public IP. Journal of Scien+B112ce & Technology, 10(2), 15–22. https://doi.org/10.46243/jst.2025.v10.i02.pp15-22
22. Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
23. Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782
24. Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.