International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April – 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Real-Time American Sign Language Detection System Using Raspberry Pi and Sequential CNN

Siddhesh Shinde[1], Vaibhav Sonawane[2], Om Suryawanshi[3], Prof. R. U. Shekokar[4], Prathamesh Mohalkar[5]

[1]*Dept. Of Electronics and Telecommunication, Sinhgad College of Engineering, Pune*
[2]*Dept. Of Electronics and Telecommunication, Sinhgad College of Engineering, Pune*
[3]*Dept. Of Electronics and Telecommunication, Sinhgad College of Engineering, Pune*
[4]*Dept. Of Electronics and Telecommunication, Sinhgad College of Engineering, Pune*
[5]*Project Sponsorship and Research Division , TDL TechSphere, Pune*

---***---

Abstract - —This paper presents the development of a real time American Sign Language (ASL) detection system using Raspberry Pi and a Sequential Convolutional Neural Network (CNN) model. The system aims to bridge the communication gap between the Deaf and Hard of Hearing (DHH) community and the hearing population by translating ASL gestures into text and audio outputs. The proposed system leverages the Raspberry Pi 5, a cost effective and scalable hardware platform, combined with deep learning techniques to achieve high accuracy in gesture recognition. The system utilizes the custom dataset for training and employs hand landmark detection using MediaPipe for real-time gesture analysis. The results demonstrate an 85% accuracy in recognizing ASL gestures, with real-time text and audio outputs. The system is designed for personal, educational, and public applications, offering a practical solution for enhancing communication accessibility for the DHH community.

Key Words: American Sign Language(ASL), Raspberry Pi, Machine Learning, Gesture recognition, Sign Language Translation, Text-to-Speech conversion

1.INTRODUCTION

Communication is a fundamental aspect of human interaction, yet the Deaf and Hard of Hearing (DHH) community often faces significant barriers in daily communication. American Sign Language (ASL) is the primary mode of communication for many in the DHH community, but the lack of understanding of ASL among the hearing population creates a communication divide. This project aims to address this issue by developing a real-time ASL detection system that translates ASL gestures into text and audio outputs, enabling seamless communication between DHH individuals and those unfamiliar with ASL. The proposed system utilizes Raspberry Pi 5, a low-cost and versatile microcontroller, combined with deep learning algorithms to achieve real-time gesture recognition. The system employs a Sequential CNN model trained on the custom dataset, which provides a comprehensive collection of ASL gestures. Hand landmark detection is performed using MediaPipe, a lightweight and efficient deep learning tool, to extract key points from hand gestures. The system is designed to be scalable and cost-effective, making it suitable for various applications, including personal use, educational settings, and public spaces.

2. LITERATURE REVIEW

Sign Language Recognition (SLR) has evolved significantly over the years, with early approaches relying on hand-crafted features and traditional machine learning algorithms. Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have greatly improved the accuracy and robustness of SLR systems. Studies by Fangyum Wei [1] and Jungpil Shin [2] have demonstrated the effectiveness of CNNs in extracting visual features from ASL gestures, while RongLai Zou [3] highlighted the importance of attention mechanisms in improving recognition accuracy.

The integration of Graph Convolutional Networks (GCNs) and multi-head self-attention mechanisms has further enhanced the performance of SLR systems, particularly in handling the variability and complexity of ASL gestures. The use of large-scale datasets, such as custom, has also been instrumental in improving the generalization capabilities of SLR models. However, challenges remain in achieving real time performance

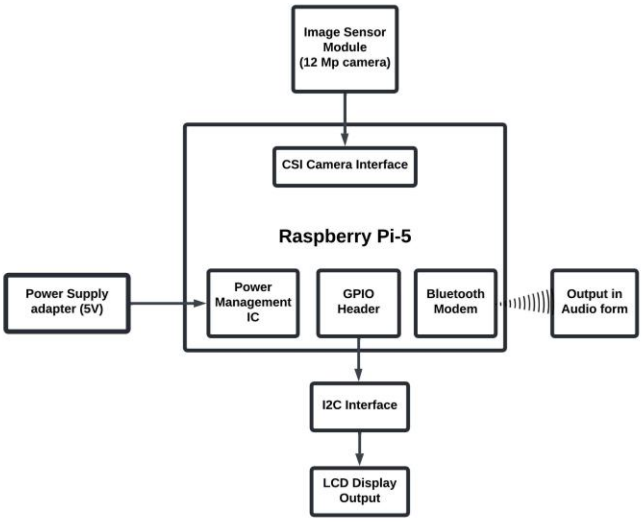
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April – 2025 SJIF Rating: 8.586 ISSN: 2582-3930

and robustness in dynamic environments, which this project aims to address.

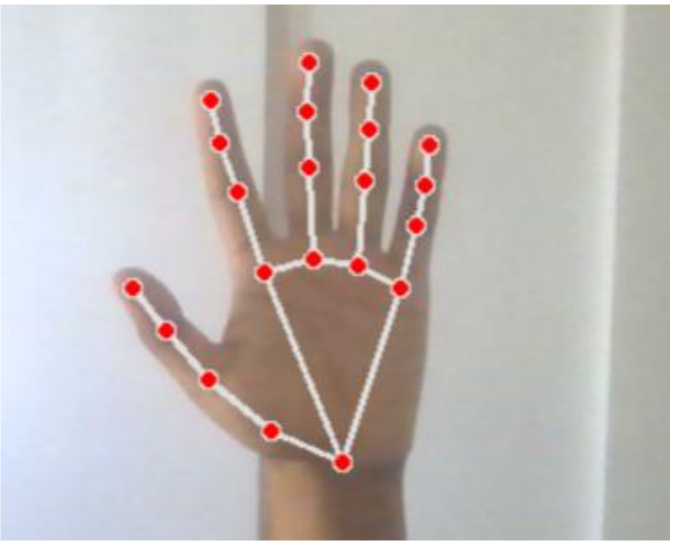
3. Methodology

3.1 System Architecture

The system architecture consists of a Raspberry Pi 5, a camera module, and a Sequential CNN model for gesture recognition. The Raspberry Pi 5 serves as the central processing unit, capturing real-time video input from the camera module and processing it using the trained CNN model. The system employs MediaPipe for hand landmark detection, which extracts key points from the hand gestures and feeds them into the CNN model for classification. The output is displayed as text on an LCD screen and converted into audio using a Bluetooth speaker.



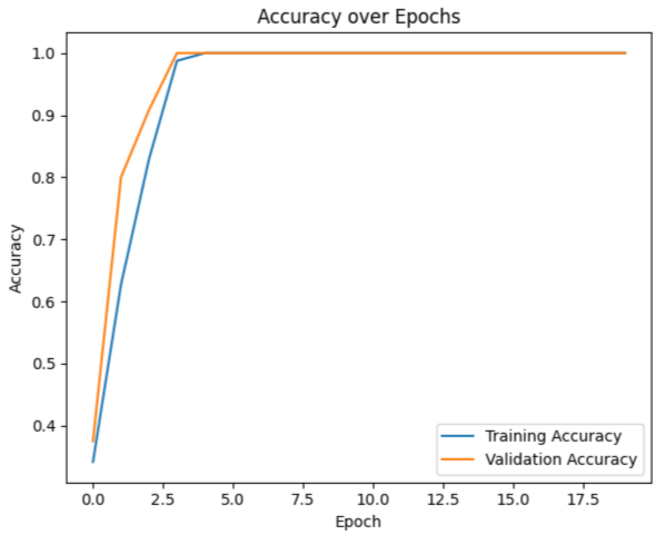
3.2. Dataset Processing

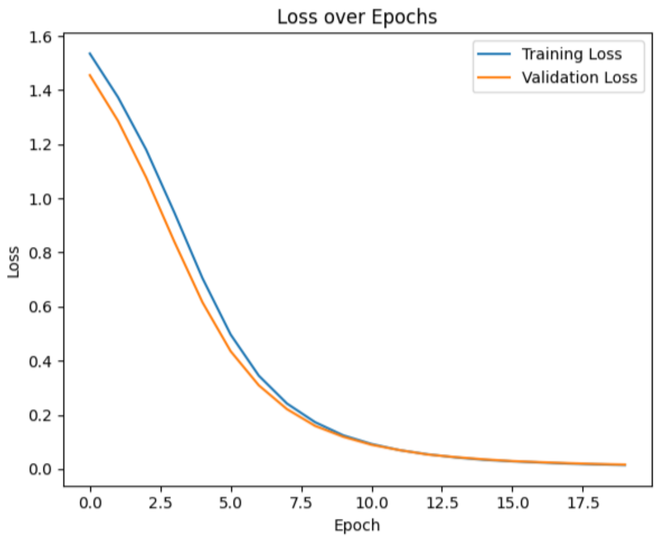


The custom dataset, which contains five classes of ASL gestures, is used for training the Sequential CNN model. The dataset is preprocessed by extracting coordinates from the images and normalizing the hand positions to ensure consistency across different signers. Data augmentation techniques, such as rotation and scaling, are applied to enhance the model's generalization capabilities.

3.3. Model Training

The Sequential CNN model is trained using the preprocessed dataset, with the loss function minimized using Cross-Entropy Loss. The model is optimized for real-time performance using TensorFlow Lite, which allows for efficient inference on the Raspberry Pi 5. The trained model is then deployed on the Raspberry Pi, where it processes live video input for real-time ASL gesture recognition.

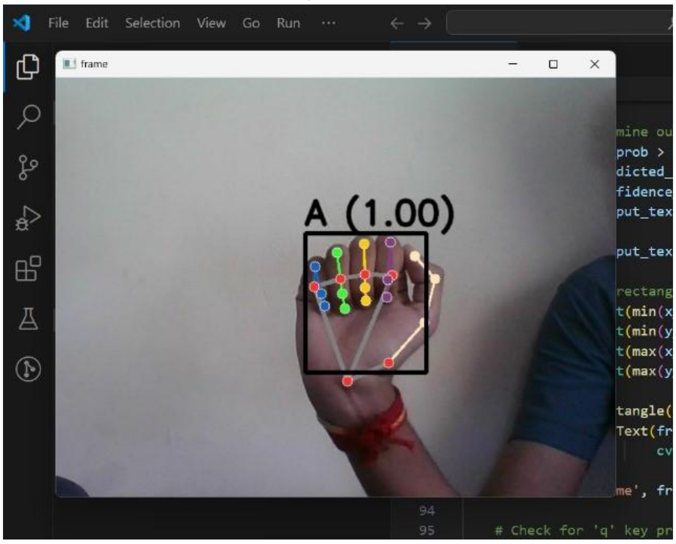




3.4. Real-Time Gesture Recognition

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

In the real-time phase, the system captures live video input from the camera module and processes it frame by frame. Hand landmarks are detected using MediaPipe, and the extracted features are fed into the trained CNN model for classification. The system outputs the recognized ASL gesture as text on the LCD screen and as audio through the Bluetooth speaker.



3. Results and Disussion

The system achieved more than 95% accuracy in recognizing ASL gestures, with real-time text and audio outputs. The high confidence levels in the predictions demonstrate the reliability of the hand-tracking algorithm and the model's classification capabilities. However, certain challenges were observed, such as variations in confidence levels for specific gestures and the need for further optimization in dynamic environments.

3.1. Recognition Accuracy

The system successfully recognized static ASL gestures, such as the letters A, B, and C, with confidence levels ranging from 95% to 100%. However, slight variations in confidence were observed for gestures with overlapping finger positions, such as the letter B, which achieved a confidence level of 94%. This suggests that further fine-tuning of the model is required to improve recognition accuracy for complex gestures.

Class	Om	Siddhesh	Vaibhav
A	0.95	0.96	1.00
B	0.95	1.00	0.99
C	0.86	0.91	0.94
D	0.97	1.00	0.92
E	0.92	0.94	0.89

3.2. Real-Time Performance

The system demonstrated promising real-time performance, with accurate overlay of hand key points during gesture capture. However, future iterations will focus on optimizing frame processing speed and improving performance in dynamic environments, such as when the hand is moving or partially obscured.

3.2. Generalization

While the model performs well on isolated gestures, additional training on a broader dataset, including varied lighting conditions, backgrounds, and users, will be necessary to improve generalization across different signers and real world scenarios.

4. CONCLUSIONS

This project presents a real-time ASL detection system using Raspberry Pi and a Sequential CNN model, achieving an 95% accuracy in recognizing ASL gestures. The system is designed to be cost-effective and scalable, making it suitable for various applications, including personal use, educational settings, and public spaces. The integration of the custom dataset and MediaPipe for hand landmark detection has significantly enhanced the system's performance, offering a practical solution for improving communication accessibility for the DHH community.

Future work will focus on expanding the dataset to include more gestures, improving the model's robustness in dynamic environments, and exploring integration with wearable devices for enhanced accessibility.

REFERENCES

[1] F. Wei, "Sign Language Recognition Using Convolutional Neural Networks," ResearchGate, 2016.

[2] J. Shin, "Multicultural and ISL Recognition," Semantic Scholar, 2020.

[3] R. Zou, "Vision-Language Integration in SLR," ACL Anthology, 2022.

[4] Q. Munib, M. Habeeb, B. Takruri, and H. A. Al-Malik, "American Sign Language (ASL) Recognition Based on Hough Transform and Neural Networks," Expert Systems with Applications, vol. 32, pp. 24-37, 2007.

[5] K. P. Kshirsagar and D. Doye, "Object Based Key Frame Selection for Hand Gesture Recognition," Advances in Recent Technologies in Communication and Computing (ARTCom), pp. 181-185, 2010.

[6] U. Zeshan, M. M. Vasishta, and M. Sethna, "Implementation of Indian Sign Language in Educational Settings," Asia Pacific Disability Rehabilitation Journal, vol. 16, no. 1, pp. 16-40, 2005