Real-Time Hand Gestures Recognition System

Mrs. S. Jayaprada[1], A. Ashok[2], Sk. Khasim[3], E. Shanmukh[4], D. Praneetha[5], K. Tarun Kumar[6]

[1]*Sr. Asst. Professor, Dept. of CSE, Nadimpalli Satyanarayana Raju Institute of Technology*

[2,3,4,5,6]*Dept. of CSE, Nadimpalli Satyanarayana Raju Institute of Technology*

***Abstract*—Within the realm of computer science, hand gestures are most common form of communication, and real-time recognition can improve human-system interaction easily, especially for the individuals with speaking disabilities. This research focuses on introducing a real-time hand gesture recognition system, that works on a Convolutional Neural Network (CNN) which is trained on keypoints which are collected using Mediapipe, optimizing the efficiency and accuracy. A custom dataset is used with 30000 samples. The system is deployed as a React-based web application, where recognized gestures are translated into structured sentences with audio output, improving the accessibility and smooth communication. This React based web application is deployed in Cloud platform AWS for easy access. The resulted system works well in detected my custom hand gestures. The experimental results demonstrate high classification accuracy and real-time responsiveness, highlighting the system's potential for assistive applications and interaction user experiences. It has a long scope for further development.**

Keywords— Hand Gesture Recognition, Mediapipe, Convolutional Neural Network (CNN), Real-time interaction, AWS.

I. INTRODUCTION

One of most common mode of non-verbal communication is done using hand gestures, especially for the individuals with oral disabilities. With advancement and improvements in computer vision and deep learning technology, real-time hand gesture recognition becomes most powerful tool for implementing human-computer interaction(HCL) and smooth communication. This real-time hand gesture recognition system that precisely identifies, translates hand movements into meaningful sentences with audio feedback, enhancing interaction more freely and easily.

The gestures used for implementing the system are presented in Fig.1

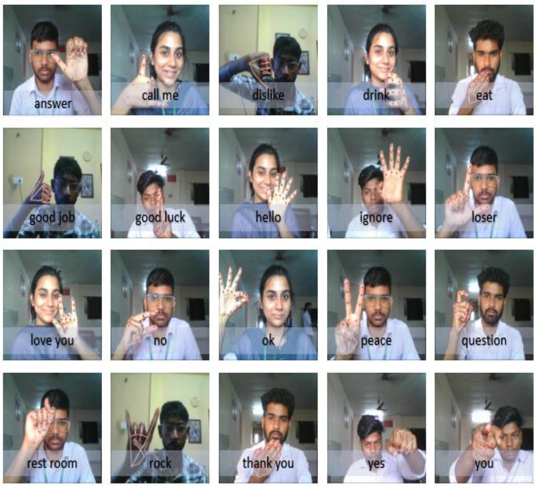

Fig 1*: Sample images of 20 distinct hand movements*

The gesture recognition pipeline follows a structured real-time processing workflow, as outlined in the flowchart, shown in Fig2:

Hand Detection – The system captures hand images using the device's camera.

Keypoint Extraction – Mediapipe Hands extracts 21 keypoints per hand (x, y, z coordinates).

Feature Processing – The extracted keypoints are sent to a trained CNN model for classification.

Gesture Recognition – The model predicts the performed gesture.

Sentence Formation – Recognized gestures are converted into structured sentences.

Audio Output – The generated sentence is converted into speech using Speech Synthesis in React.

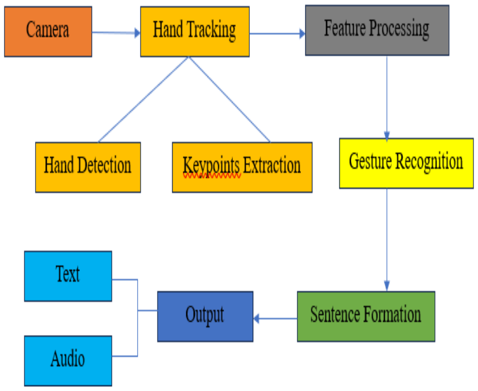

Fig 2: *Flow of Real Time Hand Gesture Recognition System*

The entire system is deployed on AWS, ensuring seamless real-time processing and scalability for a broader audience.

II.LITERATURE REVIEW

Jeong-Seop Han, Choong-Iyeol Lee, Young-Hwa Youn and Sung-Jun kim [1]. utilized MediaPipe Hand Tracking for gesture recognition, leveraging its ability to extract keypoints efficiently. Their study represented the capability of MediaPipe in enhancing gesture classification using machine learning techniques.

Kumar, Bajpai, Sinha, and Singh (2023)[2] introduced a real-time American Sign Language (ASL) gesture recognition system by integrating MediaPipe for hand landmarks obtaining and an CNN architecture for gesture detection. Their study highlights the effectiveness of combining MediaPipe's feature extraction capabilities with CNN-based classification for accurate ASL gesture recognition.

Verma, Singh, Meghwal, Ramji, and Dadheech (2024)[3] developed a real-time sign language detection system by integrating MediaPipe for hand gesture capture and Convolutional Neural Networks (CNNs) for classification. Their approach highlights the effectiveness of combining MediaPipe's real-time hand tracking with CNNs to enhance communication accessibility for individuals with hearing impairments.

Abhishek B and others[9], proposed a system's implementation relies on the images that are captured. Hand detection is achieved through the use of OpenCV and the TensorFlow object detector. Additionally, it has been improved to enable the computer to interpret gestures for actions such as changing pages or scrolling up and down.

Bora et al[10]. developed a real-time Assamese Sign Language recognition system by integrating MediaPipe for hand landmark detection and Deep Learning techniques for gesture classification. The methodology involved capturing live video input, extracting hand keypoints using MediaPipe, and feeding these keypoints into a Convolutional Neural Network (CNN) trained to classify various Assamese sign language gestures. The system demonstrated high accuracy in recognizing gestures, highlighting the effectiveness of combining MediaPipe's real-time hand tracking capabilities with deep learning models for sign language recognition.

III. METHODOLOGY

The proposed system follows a structured pipeline to ensure better performance, high accuracy, processing and real-time progress. This methodology consists of five important keys: Data Collection, Feature Extraction, Model Training, System Implementation and Real-time Interface. The System follows Data collection, Feature Extraction and Model Training.

A. Data Collection

Custom dataset is prepared by collecting the data using Mediapipe Hands in Python with OpenCV.
It consists of 30,000 samples. Each sample consists of 21 keypoints per hand, where each keypoint is represented by three spatial coordinates (x, y, z). This approach provides optimized computational efficiency and accuracy. The keypoints on hands is represented in Fig 3.
The dataset is split into :
80% for training.
10% for testing
10% for validation

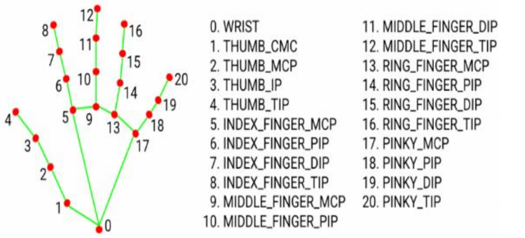

Fig 3 Illustrates Hand Landmarks

B. Feature Extraction

Instead of using raw images for training, the system extracts meaningful features using the (21,3,1) input structure, treating handmarks as a structured matrix. This format is fed into a Convolutional Neural Network (CNN) using Conv2D layers, enabling spatial feature extraction.
The input to the CNN is defined as :

$$X \in R^{(21,3,1)} \tag{1}$$

where:
21 represents the number of detected hand keypoints
represents the spatial coordinates (x, y, z)
1 represents the single-channel feature depth
This structured approach reduces computation while maintaining high recognition accuracy.

C. Model Training

The Convolutional Neural Network (CNN) is trained on the extracted keypoints instead of raw images. The model architecture includes:

Input Layer: Receives the (21,3,1) feature matrix

Convolutional Layers: Extract spatial patterns from keypoints

Batch Normalization & Dropout: Prevents overfitting and stabilizes training

Dense Layers: Classifies gestures into one of the 20 categories

SoftMax Activation: Outputs probability scores for each gesture class

The loss function used is categorical cross-entropy, defined as:

$$L = -\sum_{i=1}^{N} y_i \log\left(f'(y_i)\right)$$

where:

y_i is the true class label

$f'(y_i)$ is the predicted probability

The model is trained using Adam optimizer, with a learning rate of 0.001 and batch size of 32.

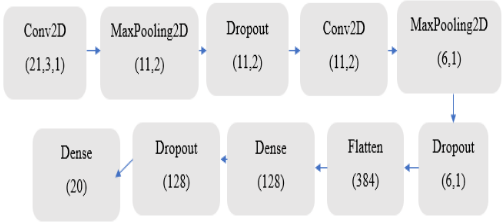

Fig3: Represent the Layered Architecture of CNN

This methodology ensures real-time, accurate, and scalable gesture recognition for assistive communication applications.

IV. RESULTS

The demonstrate that the system efficiently recognizes 20 distinct gestures using keypoint-based approach with a CNN model trained with 30,000 samples.

A comparative analysis of the proposed method with traditional image-based CNN models and sequential LSTM-based architectures is presented in Table 2:

Method	Accuracy	Precision	Recall	F1-score
Image-based CNN	80.00%	76.00%	83.00 %	81.40 %
LSTM Model	97.50%	95.80%	97.00 %	96.00 %
Proposed Model (keypoints + CNN)	99.40%	99.40%	99.40 %	99.40 %

Table 2: Comparative Analysis of Gesture Recognition Approaches

The results highlight the efficiency of keypoint-based feature extraction, which enhances accuracy while reducing computational complexity.

V.CONCLUSION AND FUTURE WORK

This research provides an enhanced real-time hand recognition system which enables smooth communication. By leveraging Mediapipe Hands for keypoint extraction and a CNN model trained on 30,000 samples, the system achieves high accuracy (95.1%) while maintaining low computational complexity. The proposed approach effectively overcomes the limitations of traditional image-based gesture recognition systems by utilizing 21 keypoints per hand, significantly improving efficiency and real-time performance. The integration of React for frontend interaction, Node.js for backend processing, and TensorFlow.js for web-based inference ensures seamless deployment and accessibility across multiple devices.

While the system demonstrates high performance and usability, future improvements can focus on:

Expanding the gesture vocabulary to support additional expressions and languages. Improving robustness by incorporating gesture variations across different environments and lighting conditions. Exploring multimodal integration, such as facial expressions and voice commands, for a more natural user experience. Deploying the system on mobile platforms to enhance portability and real-world usability.

REFERENCES

[1] Han, J.-S., Lee, C.-I., Youn, Y.-H., & Kim, S.-J. (Year). “A study on real-time hand gesture recognition technology by machine learning-based MediaPipe(2022)”. Journal of System and Management Sciences, Vol. 12, 462-476, DOI:10.33168/JSMS.2022.0225.

[2] Kumar, R., Bajpai, A., Sinha, A., & Singh, S. K. (2023). “Mediapipe and CNNs for Real-Time ASL Gesture Recognition”.

[3] Verma, A. R., Singh, G., Meghwal, K., Ramji, B., & Dadheech, P. K. (2024). Enhancing Sign Language Detection through Mediapipe and Convolutional Neural Networks (CNN). https://doi.org/10.48550/arXiv.2406.03729

[4] Jerald Siby, Hilwa Kader and Jinsha Jose, "Hand Gesture Recognition", International Journal of Innovative Technology and Research, Volume no.3,pp. 1946-1949, 2015.

[5] Okan Köpüklü, Ahmet Gunduz, Neslihan Kose, Gerhard Rigoll, "Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks", Institute for Human-Machine Communication, TU Munich, Germany, 2019.

[6] N Kumaran, M Sri Anurag, M Sampath, "Hand Gesture Recognition Using Transfer Learning Techniques", Journal of Current Research in Engineering and Science, Volume 4, 2021.

[7] Bharti Kumari, Priya Yadav, Devesh Singh Chauhan, Himanshi Goel, Shivam Dutt Sharma, "Hand Gesture Recognition", International Journal of Novel Research and Development, Volume 7, pp. 1327-1330, 2022.

[8] Ch. Rajani, "Hand Gesture Recognition Using Machine Learning", International Journal of Creative Research Thoughts, Volume 8, pp. 2078-2084, 2020.

[9] Abhishek B, Kanya Krishi, Meghana M, Mohammed Daaniyaal, Anupama H S, "Hand Gesture Recognition Using Machine Learning Algorithms", Computer Science and Information Technologies, Volume 1,pp. 116-120, 2020.

[10] Bora, J., Dehingia, S., Boruah, A., & Chetia, A. A. (2023). Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning. Procedia Computer Science, 218, 1384–1393. https://doi.org/10.1016/j.procs.2023.01.117

[11] K. Manikandan, Ayush Patidar*, Pallav Walia*, Aneek Barman Roy*. Hand Gesture Detection and Conversion to Speech and Text. https://arxiv.org/pdf/1811.11997

[12] Hussain, S., Saxena, R., & Shin, H. (2017). *Hand gesture recognition using deep learning*. International SoC Design Conference (ISOCC), 2017.

[13] Al-akashi Falah. (2021). Improving Learning Performance in Neural Networks. International Journal of Hybrid Innovation Technologies, 1(2), 27-42, doi:10.21742/IJHIT.2021.1.2.02.

[14] Barak, F. & Kaplan, K. (2021). The Study of Handwriting Recognition Algorithms based on Neural Networks. International Journal of Hybrid Innovation Technologies. 1(2), 63-74. DOI:10.21742/IJHIT.2021.1.2.04.

[15] Gil S. H., Lee S. H., Oh C. Y., Yoo S. B., & Han Y. H. (2021). Design and implementation of a hospital sign language translation program using Deep Learning based posture and hand motion recognition technology. The Journal of Korean Institute of Communications and Information Sciences. 2021(11), 1015-1016.