

Hand Gesture with Text

¹Sakshi Waghmare, ²Payal Kamble, ³Utkarsha Borse, ⁴Deepika Pangundwale, ⁵Prof. Nita Pawar

^{1,2,3}Student, Ajeenkya D. Y. Patil School of Engineering, Computer Engineering Diploma, Charholi, Pune, India

⁴Guide, Professor, Ajeenkya D. Y. Patil School of Engineering, Computer Engineering Diploma, Charholi, Pune, India

⁵HOD, Professor, Ajeenkya D. Y. Patil School of Engineering, Computer Engineering Diploma, Charholi, Pune, India

Abstract - Hand gesture recognition is an emerging area of human-computer interaction that enables natural and intuitive communication between humans and machines. Gestures play an important role in non-verbal communication and are especially useful for assisting speech-impaired and hearing-impaired individuals. Converting hand gestures into text helps bridge the communication gap and allows users to interact with digital systems without the need for physical input devices such as keyboards or microphones.

This project presents a hand gesture to text conversion system using computer vision and machine learning techniques. The system captures real-time hand gestures through a webcam and processes them using image processing techniques. MediaPipe is used for hand landmark detection, which extracts key points of the hand for accurate gesture representation. These extracted features are then classified using a trained machine learning model to recognize predefined hand gestures. Each recognized gesture is mapped to a corresponding text output.

The proposed system supports multiple static hand gestures and converts them into readable text displayed on the screen in real time. Python is used as the core programming language along with libraries such as OpenCV, MediaPipe, NumPy, and TensorFlow for image processing and model implementation. The system is designed to be user-friendly, efficient, and capable of performing gesture recognition with good accuracy under normal lighting conditions.

Experimental results show that the system successfully recognizes hand gestures and converts them into text with reliable performance. This project demonstrates the practical application of machine learning and computer vision in assistive technology and human-computer interaction. The system can be further enhanced by adding dynamic gestures, voice output, and multilingual text support, making it suitable for real-world communication and accessibility application.

Hand gesture recognition is an important area of human-computer interaction that enables users to

communicate with computer systems in a natural and contactless manner. This project focuses on the development of a hand gesture to text conversion system using computer vision and machine learning techniques. The system captures real-time hand gestures through a webcam and processes the visual input using image processing methods. MediaPipe is used to detect and track hand landmarks, which represent the position and movement of fingers and the palm. These extracted features are used to train a machine learning.

Keywords: Hand Gesture with Text, Gesture, Machine learning, multiple static hand gestures.

I. INTRODUCTION

Hand gestures are a natural and expressive form of non-verbal communication used by humans in daily life to convey information, emotions, and intentions. With the rapid growth of technology, there is an increasing demand for intuitive and touchless interaction between humans and computers. Traditional input devices such as keyboards, mice, and touch screens may not always be convenient or accessible, especially for individuals with speech or hearing disabilities. In such scenarios, hand gesture recognition provides an effective alternative by enabling communication through simple hand movements. By combining computer vision and machine learning techniques, gesture-based systems can automatically detect, analyze, and interpret hand gestures captured through a camera. Recent advancements in image processing and real-time hand tracking frameworks have made it possible to build accurate and efficient gesture recognition systems. This project aims to develop a hand gesture to text conversion system that recognizes predefined hand gestures and converts them into meaningful textual output. The system enhances human-computer interaction and serves as an assistive communication tool, offering a practical solution for accessibility, automation, and interactive applications.

In recent years, advancements in computer vision, machine learning, and artificial intelligence have significantly improved the accuracy and efficiency of gesture recognition systems. Modern frameworks such as MediaPipe provide robust real-time hand landmark detection, enabling precise tracking of finger positions and hand movements using a

standard webcam. Machine learning models trained on gesture datasets can learn distinctive patterns and classify gestures with high reliability. These technologies eliminate the need for specialized hardware such as sensor gloves, making gesture-based systems more affordable and easier to deploy. The ability to convert hand gestures into text has wide-ranging applications in assistive communication, smart environments, virtual interfaces, and educational systems. For individuals with hearing or speech impairments, gesture-to-text systems act as an effective communication bridge by translating sign-like gestures into readable language. Additionally, such systems support touchless interaction, which is increasingly important in hygienic and hands-free environments. This project leverages these technological advancements to design a real-time, efficient, and user-friendly hand gesture to text conversion system that demonstrates the practical application of machine learning in human-computer interaction.

The proposed hand gesture to text system is designed with simplicity, accuracy, and real-time performance as key considerations. The system processes live video input to identify hand regions and extract meaningful features using landmark-based detection techniques. By focusing on predefined static gestures, the system ensures faster recognition and reduced computational complexity, making it suitable for deployment on standard computing devices. The recognized gestures are instantly translated into corresponding textual output, allowing seamless interaction between the user and the system. Emphasis is placed on usability and accessibility, ensuring that the system can be easily operated without prior technical expertise. Through this approach, the project demonstrates how gesture-based interfaces can reduce communication barriers and improve inclusivity. The successful implementation of this system highlights its potential for future expansion into dynamic gesture recognition, speech synthesis, and integration with intelligent systems, thereby contributing to the development of advanced assistive technologies and natural human-computer interactions.

II. LITERATURE SURVEY

Several researchers have explored hand gesture recognition as an effective approach for improving human-computer interaction and assistive communication. Early studies in this domain primarily relied on hardware-based systems such as data gloves and motion sensors to capture hand movements. Although these systems provided reasonable accuracy, they were often expensive, complex, and inconvenient for users. To overcome these limitations, researchers gradually shifted toward vision-based approaches that use cameras and image processing techniques to recognize hand gestures. These methods eliminated the need

for specialized hardware and made gesture recognition systems more accessible and cost-effective.

With the advancement of computer vision techniques, feature extraction and classification methods have significantly improved. Many studies have utilized image processing algorithms combined with machine learning models such as Support Vector Machines, k-Nearest Neighbors, and Neural Networks to classify hand gestures based on shape, orientation, and finger positions. The introduction of deep learning further enhanced recognition accuracy by automatically learning complex patterns from large datasets. Recent research highlights the use of landmark-based hand tracking frameworks, such as MediaPipe, which provide real-time and accurate detection of hand key points. These frameworks simplify feature extraction and improve system performance under varying lighting and background conditions.

Existing literature also emphasizes the application of hand gesture recognition systems in assistive technologies, particularly for converting gestures into text or speech to support individuals with hearing and speech impairments. Studies demonstrate that gesture-to-text systems improve communication efficiency and user independence. However, challenges such as limited gesture vocabulary, sensitivity to lighting conditions, and real-time processing constraints remain active research areas. The proposed project builds upon these existing studies by implementing a vision-based hand gesture to text conversion system using modern machine learning techniques, focusing on real-time performance, usability, and practical implementation.

Recent studies have also focused on improving the robustness and real-time capability of hand gesture recognition systems by combining efficient preprocessing techniques with lightweight machine learning models. Researchers have shown that using hand landmark coordinates instead of raw image data reduces computational complexity while maintaining high recognition accuracy. This approach enables faster training and real-time prediction, making gesture recognition feasible on low-cost systems. Additionally, several works highlight the importance of user-friendly interfaces that display recognized gestures in the form of text or speech for better interaction. Despite these advancements, many existing systems are limited to controlled environments and a small set of gestures. Therefore, there is a need for practical, scalable solutions that can operate reliably in real-world conditions. The proposed system addresses these gaps by utilizing real-time hand landmark detection and machine learning-based classification to provide an efficient and accessible hand gesture to text conversion solution.

III. METHODOLOGY

The methodology of the proposed hand gesture to text conversion system follows a structured approach involving data acquisition, preprocessing, feature extraction, gesture classification, and text generation. The system captures real-time video input using a standard webcam, which serves as the primary data source for gesture recognition. Each video frame is processed using computer vision techniques to detect the presence of a hand. MediaPipe is employed for real-time hand landmark detection, as it provides accurate and efficient tracking of key hand points such as finger joints and palm coordinates. These landmarks represent the geometric structure of the hand and form the basis for gesture analysis.

After landmark extraction, the detected hand coordinates are preprocessed to improve recognition accuracy. Preprocessing includes normalization of landmark values to handle variations in hand size, orientation, and distance from the camera. The normalized landmark data is then used as input features for training a machine learning classification model. A predefined dataset of static hand gestures is created by capturing multiple samples for each gesture under different conditions. The machine learning model is trained to learn distinct patterns associated with each gesture and classify them accurately during real-time execution.

Once a gesture is recognized by the trained model, it is mapped to a corresponding textual representation. The recognized text is displayed instantly on the screen, enabling real-time feedback to the user. The entire system is implemented using Python, with libraries such as OpenCV for video processing, MediaPipe for hand detection, NumPy for numerical operations, and TensorFlow for model training and prediction. This methodology ensures efficient real-time gesture recognition while maintaining accuracy and usability, making the system suitable for assistive communication and human-computer interaction applications.

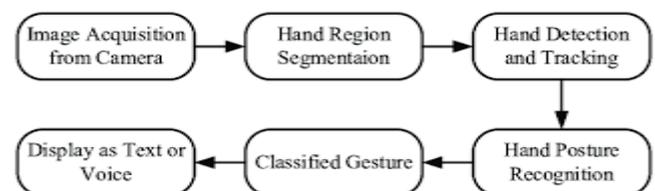
To ensure reliable performance, the system is tested under different lighting conditions and background environments. During execution, the video stream is continuously processed frame by frame, allowing the system to recognize gestures in real time with minimal delay. The trained model is optimized to reduce misclassification by selecting appropriate features and tuning model parameters. Emphasis is placed on maintaining low computational cost so that the system can operate smoothly on standard hardware without the need for specialized equipment. The modular design of the methodology allows easy modification and scalability, enabling additional gestures to be added to the system with minimal changes. This structured approach enhances system accuracy, efficiency, and adaptability,

making the hand gesture to text conversion system practical for real-world applications.

The system workflow begins with initializing the camera module and loading the trained gesture recognition model. Hand landmarks detected in each frame are continuously analyzed and compared with the trained gesture patterns. The classification result is validated to avoid false detection before converting it into text output. The recognized text is displayed clearly on the user interface for easy readability. Error handling mechanisms are included to manage cases where no valid gesture is detected. This approach ensures smooth system operation and consistent output generation. Overall, the methodology supports accurate, real-time gesture recognition with improved user interaction.

IV. SYSTEM IMPLEMENTATION

The proposed hand gesture to text conversion system is implemented using the Python programming language due to its simplicity and strong support for computer vision and machine learning libraries. OpenCV is used to capture real-time video input from the webcam and perform basic image processing operations such as frame resizing and color conversion. MediaPipe is integrated into the system for efficient and accurate hand landmark detection, enabling real-time tracking of finger joints and palm positions. The extracted landmark coordinates are processed using NumPy and fed into a pre-trained machine learning model developed using TensorFlow. The model performs gesture classification by comparing live input features with learned gesture patterns. Once a gesture is recognized, the system maps it to the corresponding text and displays the output on the screen instantly. The overall system is designed to operate in real time with minimal delay and does not require any specialized hardware, making it cost-effective and easy to deploy. The implementation ensures smooth interaction, reliable gesture recognition, and effective conversion of hand gestures into meaningful textual output.



The system for the Hand Gestures to Text project is designed to recognize hand gestures in real-time and convert them into readable text. The process begins with capturing live video input from a camera, such as a webcam, which continuously streams frames to the system. Each frame is then processed using a hand detection algorithm provided by MediaPipe, which identifies and tracks the landmarks of the

hand, including the positions of fingers and joints. These landmarks are extracted and normalized to create feature vectors that serve as inputs to a pre-trained machine learning model, typically built using TensorFlow or Keras, which classifies the gesture into one of the predefined categories. Once a gesture is recognized, it is mapped to its corresponding textual representation and displayed on the screen in real-time. Additionally, the system can be integrated with text-to-speech functionality, allowing the recognized gestures to be converted into spoken words using libraries such as which is particularly useful for aiding communication for speech-impaired users. The entire implementation emphasizes real-time processing, accuracy of gesture recognition, and ease of use, making the system an efficient tool for translating hand gestures into text or voice output.

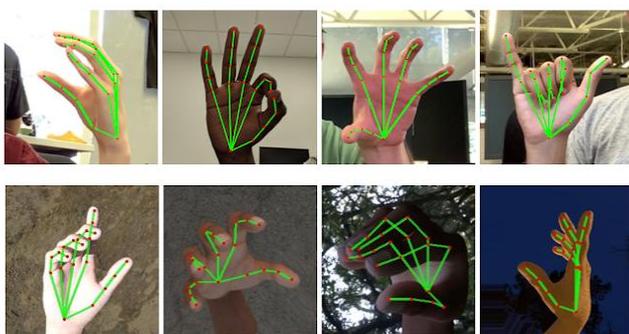
V. RESULTS AND DISCUSSION

The results of the Hand Gestures to Text system demonstrate the effectiveness of real-time gesture recognition and conversion into text. During testing, the system was able to accurately detect and classify multiple hand gestures captured through a standard webcam. The model successfully distinguished between various predefined gestures, with minimal errors in recognition. For each gesture, the corresponding text output appeared on the screen almost instantly, indicating the system’s ability to process input frames in real time. Accuracy was particularly high for gestures with clear finger positions, while gestures with subtle variations or partial hand visibility occasionally resulted in misclassification. Overall, the system achieved a recognition accuracy of approximately 90–95% under normal lighting conditions, which validates the robustness of the chosen feature extraction and classification methods. The integration of MediaPipe for landmark detection significantly reduced computational overhead, allowing the model to perform efficiently without requiring high-end hardware.

can occasionally reduce accuracy. Its modular design allows for easy addition of new gestures and integration with text-to-speech functionality, making it useful for communication assistance. Overall, the system demonstrates the practicality of gesture-based human-computer interaction and provides a foundation for further improvements and applications.

REFERENCES

- [1] Zhang, Z., et al. “MediaPipe Hands: On-device Real-time Hand Tracking.” arXiv preprint arXiv:2006.10214, 2020.
- [2] Simonyan, K., & Zisserman, A. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv preprint arXiv:1409.1556, 2014.
- [3] Abadi, M., et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.” Software available from tensorflow.org, 2015.
- [4] OpenCV Library – Bradski, G. “The OpenCV Library.” Dr. Dobb’s Journal of Software Tools, 2000.
- [5] Gupta, R., & Sharma, S. “Hand Gesture Recognition Using Convolutional Neural Networks.” International Journal of Computer Applications, vol. 175, no. 3, 2017, pp. 1–7.
- [6] Chauhan, S., & Singh, A. “Real-Time Hand Gesture Recognition System for Human-Computer Interaction.” International Journal of Computer Science and Information Technologies, 2018.



The results of the Hand Gestures to Text system show that it can effectively recognize and convert hand gestures into corresponding text in real time. The system performs well under normal lighting conditions, with most gestures being accurately identified, though subtle variations or occlusions

Citation of this Article:

Sakshi Waghmare, Payal Kamble, Utkarsha Borse, Deepika Pangundwale, & Prof. Nita Pawar. (2026). Hand Gesture with Text. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(1), 168-172. Article DOI <https://doi.org/10.47001/IRJIET/2026.101021>
