IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

# Accuracy Enhancement of Hand Gesture Recognition Using CNN

**Dr. Monika Dhananjay Rokade[1], Dr. Sunil Sudam Khatal[2], Shradha Sadanand Wabale[3]**
[1, 2] Assistant Professor, Department of Computer engineering
[3] Student, Department of Computer engineering
Sharadchandra Pawar College of Engineering, Otur, India.
monikarokade4@gmail.com, khatalsunils88@gmail.com, shradhawabale2@gmail.com

**Abstract:** *Sign language serves as an essential communication medium for individuals with hearing and speech impairments, where hand gestures play a crucial role. However, learning sign language can be challenging due to limited availability of structured learning resources and dependence on peer-based interaction. To address this issue, this study presents an intelligent system capable of automatically recognizing hand gestures and converting them into textual output, thereby facilitating effective communication between impaired and non-impaired individuals.*

*The proposed approach utilizes a Convolutional Neural Network (CNN) for accurate gesture classification. To enhance model performance, preprocessing techniques such as hand segmentation and calibration of hand orientation and position are applied to generate reliable training and testing data. Since varying lighting conditions can significantly impact skin color detection, a Gaussian Mixture Model (GMM) is employed to effectively distinguish skin regions from the background. The processed and calibrated images are then used to train the CNN model for robust gesture recognition.*

*The system aims to provide a user-friendly, efficient, and real-time solution for improving human-computer interaction and supporting inclusive communication..*

**Keywords:** Convolutional Neural Network (CNN), Hand Gesture Recognition, MobileNet, Sign Language, Human-Computer Interaction

## I. INTRODUCTION

Communication is a fundamental process that enables the exchange of information between individuals, forming the basis of human interaction and societal development. Effective communication requires a shared medium or language that can be easily understood by all participants. However, for individuals with hearing and speech impairments, conventional verbal communication is not feasible, making alternative communication methods essential. Sign language has emerged as one of the most effective and widely adopted forms of communication among the deaf community, utilizing hand gestures, facial expressions, and body movements to convey meaning [1].

Sign language primarily relies on hand gestures, which involve specific shapes, positions, and movements of the hands. These gestures represent alphabets, words, or complete sentences, enabling users to express thoughts and emotions. Despite its importance, learning sign language can be challenging due to limited availability of structured learning resources and dependency on human instructors or peer groups. This creates a communication gap between hearing-impaired individuals and the general population, highlighting the need for technological solutions that can bridge this divide [2].

Human-Computer Interaction (HCI) focuses on improving the interaction between humans and machines by making systems more intuitive, responsive, and user- friendly. One of the key goals of HCI is to develop interfaces that can understand natural human inputs such as speech, gestures, and facial expressions. Gesture-based interaction, in particular, has gained significant attention as it provides a natural and contactless way of communication, making it highly suitable for assistive technologies [3].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

A gesture in sign language can be defined as a deliberate movement or positioning of the hand with a specific shape and orientation. Gesture recognition involves identifying and interpreting these movements using computational techniques. It is a multidisciplinary field that combines concepts from computer vision, pattern recognition, and machine learning. Gesture recognition systems are not limited to hand gestures alone but can also be extended to recognize head movements, body posture, and other forms of non-verbal communication [4].

With the advancement of computer vision technologies, it has become possible to detect and analyze hand gestures using digital cameras or webcams. These systems capture real-time images or video streams and process them to identify gestures. Traditional approaches for gesture detection involve techniques such as contour detection, edge detection, and feature extraction. While these methods have shown promising results, they often struggle in complex environments with varying lighting conditions and backgrounds [5].

To overcome the limitations of traditional methods, deep learning techniques have been increasingly adopted for gesture recognition tasks. Among these, Convolutional Neural Networks (CNNs) have proven to be highly effective in image and video analysis. CNNs automatically learn hierarchical features from input data, eliminating the need for manual feature extraction. This capability makes them particularly suitable for recognizing complex patterns in hand gestures [6].

CNN-based models have been widely used in various computer vision applications such as image classification, object detection, facial recognition, and semantic segmentation. In the context of hand gesture recognition, CNNs can process large datasets of gesture images and learn distinguishing features that help in accurate classification. Their ability to generalize across different variations in gestures makes them a powerful tool for developing robust recognition systems [7].

However, one of the major challenges in gesture recognition is the variation in lighting conditions, which can significantly affect the detection of skin regions. To address this issue, advanced preprocessing techniques such as skin color segmentation are employed. Gaussian Mixture Models (GMM) are commonly used for modeling skin color distributions, enabling the system to effectively separate hand regions from the background even under varying illumination conditions [8].

In addition to preprocessing, the selection of an efficient CNN architecture plays a crucial role in system performance. MobileNetV2 is a lightweight deep learning model designed for mobile and embedded devices. It uses depthwise separable convolutions to reduce computational complexity while maintaining high accuracy. This makes it highly suitable for real-time gesture recognition applications where computational resources are limited [9].

In this work, a CNN-based hand gesture recognition system is proposed to facilitate smart human-computer interaction. The system aims to automatically detect hand gestures and convert them into textual output, thereby assisting individuals with hearing and speech impairments in communicating effectively. By integrating advanced preprocessing techniques and a lightweight deep learning model, the proposed system strives to achieve high accuracy, efficiency, and real-time performance, contributing to the development of inclusive and intelligent interaction systems [10].

## II. PROBLEM STATEMENT

Effective communication remains a significant challenge for individuals with hearing and speech impairments, as they primarily rely on sign language, which is not widely understood by the general population. This creates a communication gap between impaired and non-impaired individuals, limiting social interaction, education, and access to services. Although sign language serves as a powerful medium, its learning process is often difficult due to the lack of structured learning resources, trained instructors, and accessible tools. Existing gesture recognition systems either depend on costly hardware such as sensors and gloves or fail to perform accurately in real-world conditions due to variations in lighting, complex backgrounds, and differences in hand orientation and positioning. Traditional image processing techniques also struggle to extract reliable features, resulting in lower accuracy and poor adaptability. Furthermore, many systems are not capable of real-time processing or efficient deployment on lightweight devices. Therefore, there is a need to develop an intelligent, cost- effective, and real-time hand gesture recognition system using

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

advanced deep learning techniques such as Convolutional Neural Networks (CNN), which can accurately recognize gestures and convert them into meaningful text output, thereby enhancing human-computer interaction and enabling seamless communication for differently-abled individuals.

## III. OBJECTIVE

1. To design and develop a hand gesture recognition system using Convolutional Neural Networks (CNN) for accurate classification of gestures.
2. To implement an efficient preprocessing technique, including hand segmentation and noise removal, for improving input image quality.
3. To develop a real-time system that can capture hand gestures through a webcam and process them instantly.
4. To convert recognized hand gestures into meaningful text output to facilitate communication for hearing and speech impaired individuals.
5. To enhance system performance and accuracy by using an optimized deep learning model such as MobileNetV2 for lightweight and efficient execution.

## IV. LITERATURE SURVEY

1. Hand Gesture Recognition Using Convolutional Neural Network
Authors: Vanapalli Durga Prasanth et al.
Year: 2024
Publication: International Journal of Intelligent Systems and Applications in Engineering
Journal: IJISAE
This paper proposes a deep learning-based framework using Convolutional Neural Networks (CNN) for recognizing hand gestures in complex environments. The authors highlight that traditional approaches rely heavily on manual feature extraction, which limits system performance. In contrast, their CNN-based method automatically learns features from raw images, eliminating the need for preprocessing steps like foreground segmentation. The system is capable of handling variations in hand size, orientation, and background clutter, making it more robust and adaptable.
The experimental results demonstrate that the proposed CNN model significantly improves gesture recognition accuracy compared to traditional machine learning techniques. The system performs effectively even under challenging lighting conditions and noisy backgrounds. The study concludes that deep learning-based approaches are highly suitable for real-time applications such as virtual reality, sign language interpretation, and smart interaction systems .

2. Human Hand Gesture Recognition with Convolutional Neural Networks
Authors: J. Wang et al.
Year: 2020
Publication: Pattern Recognition (Elsevier)
Journal: ScienceDirect
This research introduces a CNN-based model designed for recognizing hand gestures in educational environments. The authors utilize infrared imaging to improve gesture detection under low-light conditions. The proposed model consists of multiple convolution layers that extract spatial features from hand images, enabling accurate gesture classification. The system focuses on recognizing pointing gestures to analyze teaching behavior in classroom scenarios.
The results show that the model achieves over 92% accuracy in gesture recognition, demonstrating its effectiveness in real-time applications. The use of infrared imaging enhances robustness by reducing dependency on lighting conditions. The study highlights the potential of CNN-based systems in behavior analysis, smart classrooms, and human-computer interaction applications .

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

3. D-CNN Based Dynamic Gesture Recognition for Indian Sign Language
Authors: D.K. Singh et al.
Year: 2021
Publication: Procedia Computer Science (Elsevier)
Journal: ScienceDirect
This paper focuses on recognizing dynamic gestures in Indian Sign Language (ISL) using a 3D Convolutional Neural Network. Unlike static gesture recognition systems, this approach captures temporal information by analyzing sequences of frames. The model is trained on gesture datasets representing commonly used ISL signs, allowing it to recognize both motion and spatial features simultaneously.
The study demonstrates that the 3D-CNN model effectively converts gestures into meaningful outputs, improving communication for hearing-impaired individuals. The system achieves high accuracy and supports natural language output generation. The authors conclude that incorporating temporal features significantly enhances gesture recognition performance, especially for dynamic gestures used in real-world communication .

4. Dynamic Hand Gesture Recognition Based on Short- Term Sampling Neural Networks
Authors: Wenjin Zhang, Jiacun Wang
Year: 2021
Publication: IEEE/CAA Journal of Automatica Sinica
Journal: IEEE
This paper presents a novel deep learning architecture for dynamic hand gesture recognition that combines short-term and long-term feature extraction. The system processes video input and captures both spatial and temporal patterns using multiple neural network modules. This hybrid approach allows the model to understand gesture sequences more effectively while reducing computational complexity. The results indicate that the proposed system achieves improved accuracy and efficiency compared to traditional methods. By integrating multiple feature extraction strategies, the model can handle complex gesture patterns and variations in motion. The study highlights the importance of temporal modeling in gesture recognition and its application in robotics, smart interfaces, and human- computer interaction systems .

5. Hand Gesture Recognition Using Convolutional Neural Network (IEEE Conference)
Authors: Ragapriya Saravanan et al.
Year: 2021
Publication: IEEE International Conference (i-PACT)
Journal: IEEE Xplore
This study presents a CNN-based hand gesture recognition system implemented using MATLAB and deep learning toolkits. The authors use image preprocessing techniques such as grayscale conversion and thresholding to prepare input data for the model. The CNN architecture includes multiple layers for feature extraction and classification, enabling the system to recognize gestures from real-time camera input.
The system demonstrates efficient performance with reasonable processing time and accuracy. The authors emphasize that CNN models are highly effective for gesture classification tasks due to their ability to learn hierarchical features. The study also discusses limitations related to hardware dependency and dataset size, suggesting improvements through larger datasets and optimized architectures .

6. Deep Learning-Based Hand Gesture Recognition: A Review
Authors: C. Cui et al.
Year: 2025
Publication: PMC / Journal Review

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

Journal: Biomedical & Computer Vision Review

This paper provides a comprehensive review of deep learning techniques used in hand gesture recognition. It discusses various models such as CNN, RNN, and hybrid architectures, highlighting their strengths and limitations. The review identifies key challenges such as background complexity, motion blur, and computational cost, which affect system performance in real-world applications.

The study also explores future research directions, including the integration of multimodal data, lightweight models for mobile devices, and improved datasets. It concludes that deep learning approaches, particularly CNN-based models, offer significant advantages in terms of accuracy and automation. However, further optimization is required to achieve real-time performance and scalability in practical applications

## V. PROPOSED SYSTEM

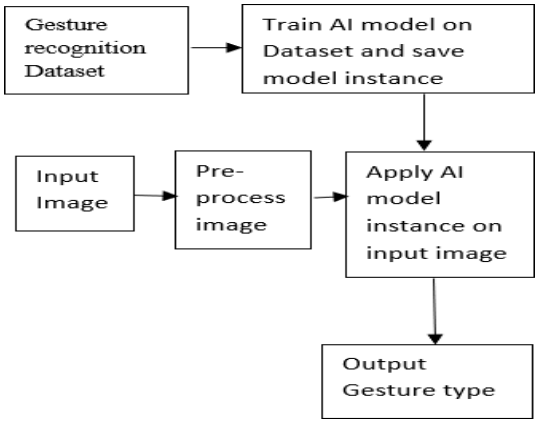


Fig 1: Block Diagram

The proposed system is designed to develop an intelligent and real-time hand gesture recognition framework using Convolutional Neural Networks (CNN) for smart human- computer interaction. The system captures hand gestures through an input device, processes the images, and classifies them into predefined gesture categories. The recognized gesture is then converted into meaningful output such as text, enabling effective communication for hearing and speech- impaired individuals. The system is divided into several functional modules to ensure accuracy, efficiency, and real- time performance.

### A. Dataset Collection and Preparation

The first step in the proposed system is the collection of a suitable hand gesture dataset. The dataset consists of various gesture classes such as open hand, fist, peace sign, and other commonly used signs. Images are collected either from publicly available datasets or captured using a webcam. Each class contains a sufficient number of images to ensure proper training of the model. The dataset is then divided into training and testing sets, typically in a 70:30 ratio, to evaluate model performance effectively.

### B. Image Acquisition

In this stage, the system captures real-time images using a webcam or camera device. The input image serves as the primary data for gesture recognition. The system continuously captures frames to ensure smooth and real-time interaction. Proper positioning of the hand in front of the camera is necessary to improve detection accuracy and reduce noise from the background.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

### C. Image Preprocessing

Preprocessing plays a crucial role in improving the quality of input data. The captured images undergo several preprocessing steps such as resizing, normalization, and noise removal. Techniques like background subtraction and skin color segmentation are applied to isolate the hand region from the background. A Gaussian Mixture Model (GMM) is used to detect skin regions effectively under varying lighting conditions. Data augmentation techniques such as rotation, shifting, and zooming are also applied to increase dataset diversity and prevent overfitting.

### D. Feature Extraction and Model Training

In this stage, a Convolutional Neural Network (CNN) is used to automatically extract important features from the processed images. The system utilizes the MobileNetV2 architecture, which is lightweight and suitable for real-time applications. The model is trained using the prepared dataset, where it learns to identify patterns and distinguish between different gestures. Optimization techniques such as Stochastic Gradient Descent (SGD) are used to minimize loss, and performance is evaluated using metrics like accuracy and mean squared error.

### E. Model Deployment and Gesture Recognition

Once the model is trained, it is saved and deployed for real-time use. The system takes a new input image, applies preprocessing, and feeds it into the trained CNN model. The model analyzes the image and predicts the corresponding gesture class. This process is performed in real-time, enabling instant recognition of hand gestures.

### F. Output Generation

After classification, the recognized gesture is converted into a meaningful output such as text. This output can be displayed on the screen or used for further applications such as speech generation. The system ensures that the output is accurate and easy to understand, thereby improving communication between users.

### G. System Workflow

The overall workflow of the proposed system follows a sequential process:

1. Dataset collection and model training
2. Image capture using webcam
3. Image preprocessing and feature extraction
4. Application of trained CNN model
5. Gesture classification and output generation

This structured workflow ensures efficient processing and accurate gesture recognition.

## VI. SYSTEM DESIGN

The system design defines the overall architecture, components, and data flow of the proposed hand gesture recognition system. It explains how different modules interact with each other to achieve accurate and real-time gesture recognition using CNN. The design follows a modular approach to ensure scalability, efficiency, and ease of implementation.

The system architecture consists of multiple interconnected modules, including input acquisition, preprocessing, model training, and output generation. The flow begins with capturing gesture images through a camera, followed by preprocessing to enhance image quality. The processed image is then passed through a trained CNN model for classification. Finally, the predicted gesture is converted into a meaningful output such as text. This layered architecture ensures smooth data processing and accurate results.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

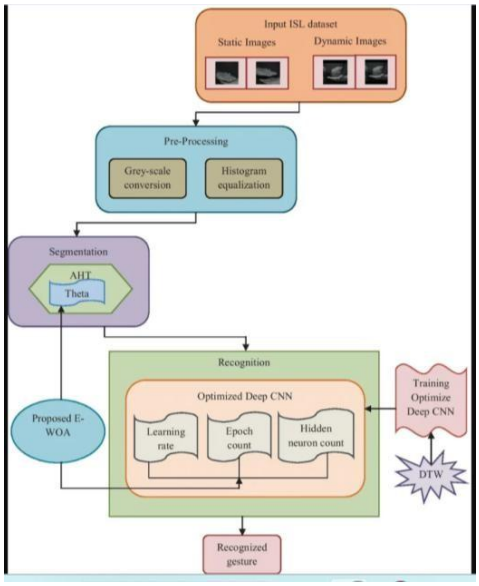


Fig 2: System Overview

**B. Input Module Design**

The input module is responsible for capturing real-time hand gesture images using a webcam or camera. The system continuously collects frames to ensure smooth interaction. The quality of input images plays a significant role in system performance; hence, proper lighting and positioning of the hand are essential. The module is designed to handle dynamic inputs and provide a steady stream of images to the preprocessing stage.

**C. Preprocessing Module Design**

The preprocessing module enhances the captured images to make them suitable for model input. It includes several operations such as image resizing, normalization, noise reduction, and background removal. Skin color segmentation is performed using a Gaussian Mixture Model (GMM) to isolate the hand region from the background. Additionally, data augmentation techniques such as rotation, scaling, and shifting are applied to increase dataset diversity and improve model generalization.

**D. Feature Extraction Module**

Feature extraction is performed automatically by the Convolutional Neural Network (CNN). Unlike traditional methods, where features are manually defined, CNN extracts hierarchical features such as edges, textures, and shapes directly from the input images. This module plays a critical role in identifying unique characteristics of each gesture and improving classification accuracy.

**E. CNN Model Design (MobileNetV2)**

The core of the system is the CNN model based on the MobileNetV2 architecture. MobileNetV2 is a lightweight and efficient deep learning model designed for mobile and embedded applications. It uses depthwise separable convolutions to reduce computational complexity while maintaining high accuracy. The model consists of multiple layers, including convolutional layers, activation functions, pooling layers, and fully connected layers. This design enables fast processing and real-time performance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

### F. Training and Validation Design

The system uses a supervised learning approach where the dataset is divided into training and validation sets. The training phase involves feeding labeled images into the CNN model, allowing it to learn patterns associated with different gestures. During validation, the model's performance is evaluated using unseen data. Metrics such as accuracy and loss are monitored to ensure optimal model performance and to avoid overfitting.

### G. Classification and Prediction Module

In this module, the trained CNN model is used to classify new input images. The model processes the preprocessed image and predicts the gesture class based on learned features. The classification is performed in real-time, enabling instant recognition of gestures. The system ensures high accuracy by using optimized model parameters and preprocessing techniques.

### H. Output Interface Design

The output module displays the recognized gesture in the form of text on the screen. This module can also be extended to generate audio output for enhanced usability. The interface is designed to be simple and user-friendly, ensuring that users can easily interpret the results without confusion.

### I. System Integration

All modules are integrated to form a complete system. The data flows sequentially from input acquisition to output generation. Proper synchronization between modules ensures smooth execution and real-time performance. The system is designed to be scalable, allowing future enhancements such as adding more gesture classes or integrating voice output.

### J. Advantages of System Design

The proposed system design offers several advantages, including real-time gesture recognition, high accuracy using CNN, lightweight implementation using MobileNetV2, and robustness against varying lighting conditions. It eliminates the need for expensive hardware and provides an efficient solution for assistive communication.

## VII. RESULT

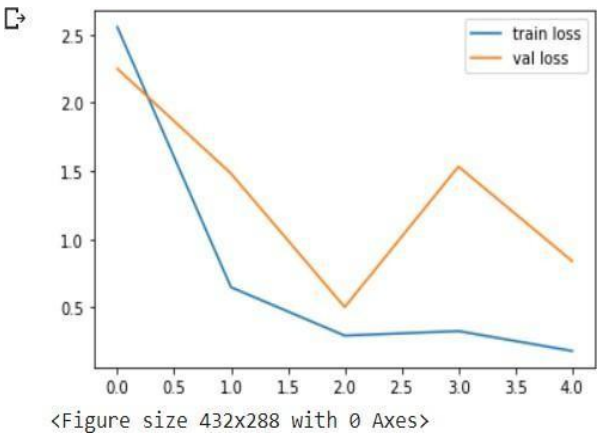


Fig 3: Graph 1

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

### A. Dataset Distribution and Experimental Setup

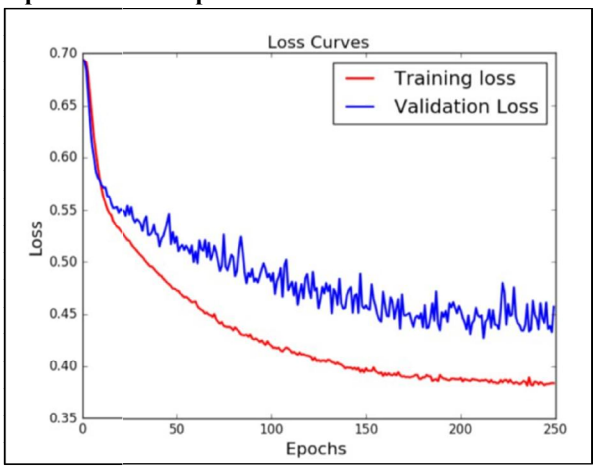


Fig 5: Loss covers

The dataset used for the proposed hand gesture recognition system was divided into two subsets: training and testing. Approximately 70% of the total dataset was used for training the CNN model, while the remaining 30% was utilized for testing and validation. This split ensures that the model learns patterns effectively during training and is evaluated on unseen data to assess its generalization capability. The dataset includes multiple gesture classes such as open hand, fist, peace sign, and others, with a balanced number of samples in each class to improve classification performance.

### B. Training and Validation Performance

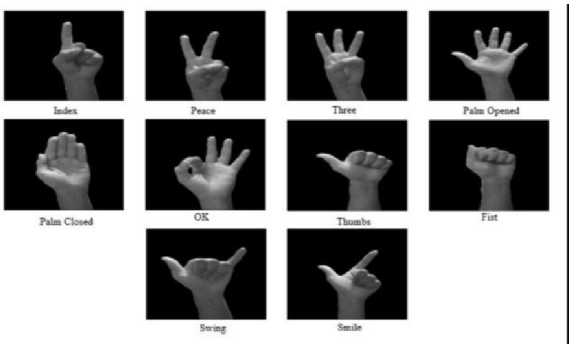


Fig 6: Performance

The performance of the model was evaluated using training loss and validation loss metrics. The graph shows a consistent decrease in both training and validation losses as the number of epochs increases. Initially, the loss values are high, but they gradually reduce as the model learns the underlying patterns in the dataset. This indicates that the CNN model is effectively optimizing its parameters and improving its predictive capability over time.

Although the validation loss shows slight fluctuations at certain stages, the overall decreasing trend confirms that the model is not significantly overfitting and maintains a good balance between bias and variance. The convergence of both loss curves demonstrates that the model is stable and capable of generalizing well to new data.

### C. Gesture Prediction Results

The trained model was tested using real-time hand gesture images captured through a camera. The system takes the input image, preprocesses it, and feeds it into the trained CNN model for classification. The model successfully predicts the gesture class for most of the test images, demonstrating its effectiveness in recognizing predefined gestures.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

For example, gestures such as open hand and fist were correctly classified by the model, indicating that it has learned the distinguishing features of each gesture. However, in some cases where the background is complex or contains multiple objects, the prediction accuracy slightly decreases. This shows that background noise and environmental conditions can affect the performance of the system.

### D. System Performance and Observations

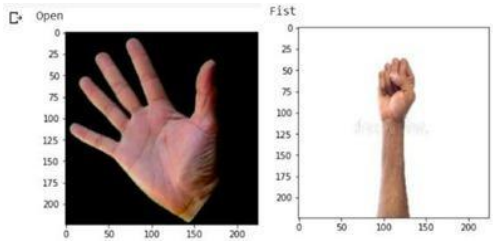


Fig 7: System Performance

Overall, the proposed system achieves satisfactory performance in recognizing hand gestures with good accuracy. The use of CNN and MobileNetV2 architecture enables efficient feature extraction and fast processing, making the system suitable for real-time applications. The model performs well under controlled conditions with proper lighting and minimal background noise.

However, certain limitations were observed, such as reduced accuracy in cluttered backgrounds and sensitivity to lighting variations. These challenges can be addressed in future work by improving preprocessing techniques, increasing dataset size, and applying advanced models. Despite these limitations, the system demonstrates strong potential for practical applications in assistive communication and smart human-computer interaction.

## VIII. CONCLUSION

The proposed system successfully demonstrates the design and implementation of a CNN-based hand gesture recognition framework for smart human-computer interaction. By leveraging deep learning techniques, particularly the MobileNetV2 architecture, the system is able to automatically extract meaningful features from input images and accurately classify different hand gestures. The integration of preprocessing techniques such as skin segmentation and data augmentation further enhances the robustness of the model under varying conditions. The results obtained show that the system performs effectively in real-time scenarios, with a consistent reduction in training and validation loss and accurate prediction of gestures such as open hand and fist. This validates the capability of the model to generalize well on unseen data and provide reliable outputs.

Despite achieving promising results, certain limitations such as sensitivity to background complexity and lighting variations were observed, which can slightly affect the prediction accuracy. These challenges indicate the need for further improvements in dataset diversity and preprocessing methods. Future enhancements may include the use of more advanced deep learning models, integration of real-time video processing, and incorporation of speech output for better accessibility. Overall, the proposed system provides a cost-effective, efficient, and user-friendly solution that contributes significantly to assistive communication technologies and advances the field of intelligent human- computer interaction.

## IX. FUTURE SCOPE

The proposed hand gesture recognition system offers significant potential for further enhancement and real-world applications. In future work, the system can be improved by incorporating a larger and more diverse dataset to increase accuracy and robustness under varying environmental conditions such as complex backgrounds and different lighting scenarios. Advanced deep learning models such as EfficientNet or transformer-based architectures can be explored to achieve higher performance and better generalization. The system can also be extended to recognize dynamic gestures

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

and continuous sign language sentences using video-based processing and recurrent neural networks (RNNs) or LSTM models. Additionally, integrating speech synthesis can convert recognized gestures into audio output, making the system more interactive and accessible. Deployment on mobile and embedded devices can further enhance portability and real-time usability. Furthermore, combining gesture recognition with other modalities such as facial expression and voice recognition can lead to the development of a more comprehensive and intelligent human-computer interaction system, ultimately contributing to inclusive communication technologies.

## REFERENCES

1) G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.
2) K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICLR), 2015.
3) Y. LeCun et al., "Gradient-Based Learning Applied to Document Recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
4) O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
5) D. Zhang, "Hand Gesture Recognition Using Computer Vision Techniques," IEEE Transactions on Industrial Electronics, 2018.
6) R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision (ICCV), 2015.
7) J. Redmon et al., "You Only Look Once: Unified, Real- Time Object Detection," CVPR, 2016.
8) Szegedy et al., "Going Deeper with Convolutions," CVPR, 2015.
9) S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
10) M. Abavisani et al., "Deep Learning-Based Hand Gesture Recognition," IEEE Access, 2019.
11) V. Prasanth et al., "Hand Gesture Recognition Using CNN," IJISAE, 2024.
12) J. Wang et al., "Human Hand Gesture Recognition Using CNN," Pattern Recognition, Elsevier, 2020.
13) K. Singh et al., "3D CNN Based Gesture Recognition for Indian Sign Language," Procedia Computer Science, 2021.
14) W. Zhang and J. Wang, "Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks," IEEE/CAA Journal of Automatica Sinica, 2021.
15) R. Saravanan et al., "Hand Gesture Recognition Using CNN," IEEE International Conference, 2021.
16) Cui et al., "Deep Learning-Based Hand Gesture Recognition: A Review," Biomedical Signal Processing and Control, 2025.
17) S. Mitra and T. Acharya, "Gesture Recognition: A Survey," IEEE Transactions on Systems, Man, and Cybernetics, 2007.
18) G. R. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.
19) Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," CVPR, 2017.
20) M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for CNNs," ICML, 2019.
21) Goodfellow et al., Deep Learning, MIT Press, 2016.
22) Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012.
23) P. Molchanov et al., "Hand Gesture Recognition with 3D CNNs," CVPR Workshops, 2015.
24) H. Starner and A. Pentland, "Real-Time American Sign Language Recognition," IEEE Transactions on Pattern Analysis, 1998.
25) T. Starner et al., "Visual Recognition of American Sign Language," IEEE Conference, 1995.
26) Koller et al., "Deep Hand Gesture Recognition for Sign Language," ICCV Workshops, 2016.
27) S. Escalera et al., "Multi-Modal Gesture Recognition Challenge," ICMI, 2013.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

28) J. Carreira and A. Zisserman, "Quo Vadis, Action Recognition? A New Model," CVPR, 2017.
29) L. Pigou et al., "Beyond Temporal Pooling: Recurrence for Gesture Recognition," CVPR Workshops, 2015.
30) N. Neverova et al., "ModDrop: Adaptive Multi-Modal Gesture Recognition," IEEE Transactions on Pattern Analysis, 2016