iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.83301

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

A Prototype Deep Learning-Based System for Indian Sign Language Alphabet Gesture Recognition

Jeetumoni Barman[1], Nafisa Zaman[2], Nayana Das[3]
[1, 2, 3]*Department of Computer Application, Girijananda Chowdhury University, Guwahati, Assam, India*

Abstract: ***Indian Sign Language (ISL) is a communication medium used by hearing and speech-impaired individuals. In this paper a prototype deep learning-based Indian Sign Language alphabet gesture recognition system using a MobileNetV2-based Convolutional Neural Network (CNN) model has been proposed. The proposed system recognizes static ISL alphabet gestures and converts them into readable text in real time. TensorFlow and Keras with transfer learning techniques have been used in developing the model. Image preprocessing and data augmentation methods were applied to improve model generalization and prediction accuracy. The MediaPipe Python framework and OpenCV were used for real-time hand detection, gesture extraction, and webcam-based prediction.The experimental results showed stable training and validation performance with minimal overfitting. During testing, the trained model achieved an accuracy of 99.90%. The developed system performed effectively under proper lighting conditions and clear hand positioning, demonstrating its suitability for real-time static ISL gesture recognition.***
Keywords: Indian Sign Language (ISL), Deep Learning, Gesture Recognition, MobileNetV2, Convolutional Neural Network (CNN), Transfer Learning, MediaPipe, Real-Time Recognition.

I. INTRODUCTION

Indian Sign Language (ISL) is visual-gesture based language widely used by individuals with hearing and speech impairments for communication through hand gestures and visual movements. Since many people are not familiar with sign language, communication between ISL users and non-sign users often becomes difficult in everyday situations, education, and social interaction. Due to these communication difficulties, many researchers are now working on automated sign language recognition systems.

Recent advancement in the field of Artificial Intelligence, Deep Learning, and Computer Vision has improved the performance of image-based recognition systems. Among different type of deep learning approaches, Convolutional Neural Networks (CNNs) are commonly used for gesture recognition system. The visual features of an image can be effectively extracted using CNN. These techniques are now being applied in various real-time recognition and assistive technology applications.

The proposed system captures hand gestures using a webcam and predicts the corresponding ISL alphabet as readable text output. TensorFlow and Keras were used for model development and training, while OpenCV and MediaPipe were used for hand detection and real-time frame processing. The dataset used for this work was divided into training, validation, and testing sets for proper performance evaluation. Image preprocessing and augmentation methods were also applied during training to improve model stability and reduce overfitting. Model testing showed stable learning performance and high prediction accuracy under normal lighting conditions and proper hand gestures.

The developed prototype shows that how deep learning and computer vision techniques can be applied in adapting communication systems. Although the current implementation focuses only on static alphabet recognition, it provides a foundation for future work related to dynamic gesture recognition and sentence-level ISL translation systems.

II. REVIEW AND RELATED WORK

Many researchers have worked on Indian Sign Language (ISL) recognition system using both machine learning and deep learning techniques.

Rokade et al. [1] worked on static ISL gesture recognition using image processing methods where the features were extracted using skin color segmentation, Fourier descriptors, Euclidean distance transformation, and Hu moments. The authors have used ANN and SVM for classification, where the ANN model achieved 94.37% accuracy while the SVM model achieved 92.12% accuracy.

Patil et al. [2] proposed Indian Sign Language alphabet recognition system using CNN where the authors have applied preprocessing methods such as grayscale conversion, masking, and dilation before classification and the model achieved nearly 95% accuracy.

Katoch et al. [3] developed a real-time ISL recognition system for alphabets and digits using SURF feature extraction together with SVM and CNN classifiers. The system converts recognized gestures into text and speech output. The CNN model achieved an accuracy of 99.64%.

With the increase in need and popularity of Indian Sign Language (ISL) and its uses,researchers started focusing more on deep learning models for better recognition performance result.

Goyal et al. [4] used MediaPipe Holistic with CNN and LSTM models for recognizing both static and dynamic gestures. The study showed that CNN models worked better for static gestures system, while LSTM models performed better for dynamic gesture recognition system.

Kadwade et al. [5] developed an ISL recognition system using CNN, LSTM, and GRU models. The system converts recognized gestures into text and speech output and achieved around 98% accuracy for alphabet and number recognition and nearly 96% accuracy for word-level recognition.

Rawat et al. [6] proposed a Sequential LSTM-based recognition model integrated with MediaPipe Holistic. The system extracted hand, face, and body landmarks for real-time gesture classification and achieved 96.97% accuracy.

Vashisth et al. [7] developed a CNN-based ISL hand gesture recognition system using a custom dataset. Different preprocessing and augmentation techniques such as resizing, HSV conversion, zooming, and flipping were applied during training. The designed CNN model achieved an accuracy of 99% with a loss value of 0.0178. Recent studies focus more on improving real-time performance and using advanced deep learning architectures.

Awalkar et al. [8] developed an ISL recognition system using MediaPipe Hands and CNN-based classification. The authors also added pyttsx3 python library for text-to-speech conversion. The system achieved approximately 92% accuracy during real-time testing.

Khetam et al. [9] proposed an ISL translation system combining computer vision and Natural Language Processing (NLP) techniques. The system proposed by the authors used CNN, LSTM, and Transformer models for sequence learning and real-time gesture translation.

Rastogi et al. [10] developed a real-time ISL recognition framework called YOLOv10-ST by combining Swin Transformer with the YOLOv10 architecture. Their model was trained using a dataset containing 15,000 images and 35 videos of ISL gestures. Experimental results showed 97.5% precision, 98.1% recall, and 97.62% mAP for gesture recognition performance.

III. PROPOSED METHODOLOGY

The proposed system was developed for recognizing static ISL alphabet gestures using deep learning and computer vision techniques. The proposed model architecture using MobileNetV2 is shown in Fig. 2. The system captures hand gestures through a webcam, processes the input image, and predicts the corresponding ISL alphabet in real time. The overall methodology includes dataset preparation, image preprocessing, deep learning model development, model training, and real-time gesture prediction implementation.

A. Dataset Collection

The dataset used was collected from a publicly available Kaggle Indian Sign Language (ISL) alphabet dataset [11]. The dataset contains static hand gesture images representing ISL alphabet classes from A–Z. Each image represents a specific hand gesture corresponding to an ISL alphabet symbol. The dataset was organized into separate folders for each alphabet as a class to simplify dataset management, training, and classification processes.

The dataset included gesture images captured under different hand positions, orientations, and lighting conditions, which helped improve model robustness and generalization capability. Images with different background conditions and gesture variations were also included to help the model learn diverse gesture patterns during training. Fig-1 shows the Indian Sign Language alphabet gesture images collected from the dataset. The dataset was divided into training, validation, and testing datasets using a custom Python script. The dataset splitting process was implemented using Python libraries such as os, shutil, and random. The training dataset was used to train the model, while the validation dataset was used to monitor model performance and reduce overfitting during training. The testing dataset was used for the final evaluation of prediction accuracy using unseen gesture images.

Proper dataset organization and splitting played an important role in maintaining balanced class distribution and improving the reliability of model evaluation. The organized dataset was further used for CNN model training and performance evaluation.

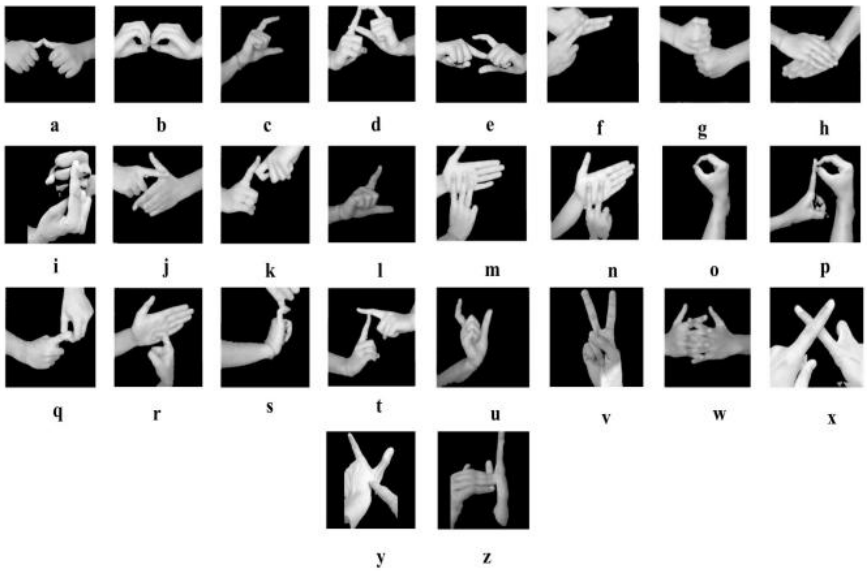


Fig-1 Indian Sign Language Alphabet Gesture Images

B. Image Preprocessing

Image preprocessing plays an important role in improving model performance, feature extraction capability, and prediction accuracy in deep learning-based gesture recognition systems. Before training, all dataset images were resized to 224×224 pixels to match the required input size of the MobileNetV2 architecture. Resizing helps maintain consistency among input images and reduces computational complexity during model training and prediction.

Image normalization and MobileNetV2 preprocessing functions were applied to improve feature extraction and compatibility with the pre-trained ImageNet weights used for transfer learning. Pixel values were normalized to improve training stability and maintain a stable learning process for the CNN model.

Additional preprocessing techniques such as grayscale conversion, Gaussian blur, and CLAHE (Contrast Limited Adaptive Histogram Equalization) enhancements were applied during real-time prediction to improve gesture visibility and reduce noise under different lighting conditions. Grayscale conversion helped reduce unnecessary color information, while Gaussian blur minimized background noise and minor image distortions. CLAHE enhancement improved local contrast and brightness distribution, allowing better visibility of hand gesture features under low-light and uneven illumination conditions.

MediaPipe hand tracking was used to detect hand landmarks and extract the Region of Interest (ROI) from the webcam frame before preprocessing. This helped reduce background interference and allowed the model to focus mainly on the hand gesture region for accurate prediction.

Data augmentation techniques were implemented using the TensorFlow and Keras ImageDataGenerator class to improve model generalization and reduce overfitting. The augmentation process included rotation, zooming, width shifting, height shifting, horizontal flipping, and brightness adjustment. These augmentation techniques helped the model learn gesture variations under different hand positions, orientations, and lighting conditionns. As a result, the trained model achieved improved robustness and stable performance during real-time gesture recognition.

C. CNN Architecture

The proposed gesture recognition model was developed using MobileNetV2-based Convolutional Neural Network (CNN) architecture with transfer learning. MobileNetV2 was selected because of its lightweight architecture, fast processing speed, and strong feature extraction performance for image classification tasks.

The MobileNetV2 model used pre-trained ImageNet weights, which helped improve feature extraction and reduce overall training time. The base MobileNetV2 model was used as the feature extraction layer, while additional custom layers were added for gesture classification.

The implemented CNN architecture included:

- MobileNetV2 base model

- Global Average Pooling Layer
- Batch Normalization Layer
- Dense Fully Connected Layer
- Dropout Layer
- Softmax output layer

The Batch Normalization and Dropout layers were used to improve model stability, reduce overfitting, and improve generalization performance. The final output layer predicts the probability score for each ISL alphabet gesture class.

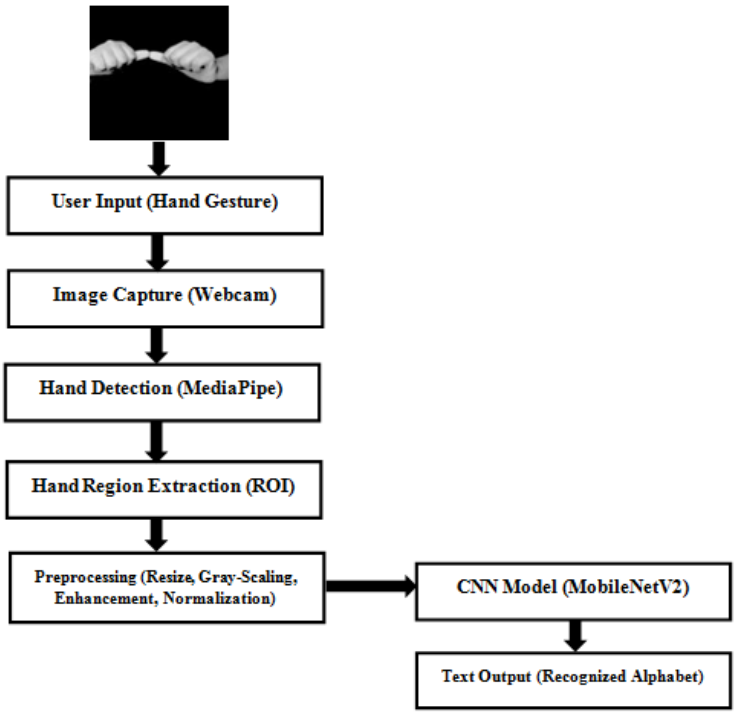


Fig-2 Proposed Model Architecture using **MobileNetV2**

D. Model Training and Evaluation

The model was developed and trained using TensorFlow and Keras frameworks in Python. The training process was carried out using the prepared training and validation datasets. During training, the model learned important feature representations of different ISL alphabet gestures from the input images.

Transfer learning with pre-trained ImageNet weights was applied in the MobileNetV2-based CNN model to improve feature extraction and reduce training time. The pre-trained feature extraction capability of MobileNetV2 helped the model identify important hand gesture patterns more effectively. Additional custom classification layers were also added to improve gesture classification performance for ISL alphabet recognition.

Regularization techniques such as Batch Normalization, Dropout, data augmentation, and Early Stopping were applied to improve model generalization and reduce overfitting during training. Batch Normalization helped stabilize the learning process, while Dropout reduced dependency on specific neurons and improved model robustness. Early Stopping automatically stopped the training process when validation performance no longer improved, helping reduce unnecessary training and overfitting.

The validation dataset was continuously monitored after each training epoch to evaluate model performance and learning behavior. Training and validation accuracy and loss values were recorded throughout the training process to analyze model stability and optimization performance.

The final trained model was evaluated using a separate testing dataset containing unseen gesture images. Experimental results showed that the trained model achieved approximately 99% training and validation accuracy with stable learning performance. During testing, the model obtained an accuracy of 99.90% with a loss value of 0.0292, demonstrating effective gesture classification capability and strong generalization performance.

The trained model was saved in Keras format for real-time deployment and later integrated with the webcam-based gesture prediction system for real-time ISL alphabet recognition.

E. Real-Time Gesture Prediction System

The real-time gesture recognition module was developed using Python libraries such as OpenCV, MediaPipe, TensorFlow, and CustomTkinter. The webcam continuously captures live video frames for gesture recognition and prediction.

OpenCV was used for webcam handling, frame processing, and displaying prediction results during execution. MediaPipe hand tracking was applied to detect hand landmarks and isolate the hand region from the webcam frame. This helped the system focus mainly on the gesture area and reduced unnecessary background interference during prediction.

After extracting the hand region, the image was processed using resizing, grayscale conversion, Gaussian blur, CLAHE enhancement, normalization, and MobileNetV2 preprocessing techniques before being passed to the trained CNN model. These image processing methods helped improve gesture visibility and maintain consistent input quality during real-time prediction.

The trained MobileNetV2 model predicts the corresponding ISL alphabet gesture and generates a confidence score for each prediction. The recognized output and confidence values were displayed through a graphical user interface (GUI) developed using CustomTkinter. Fig-3 shows the User Interface for capturing hand gesture

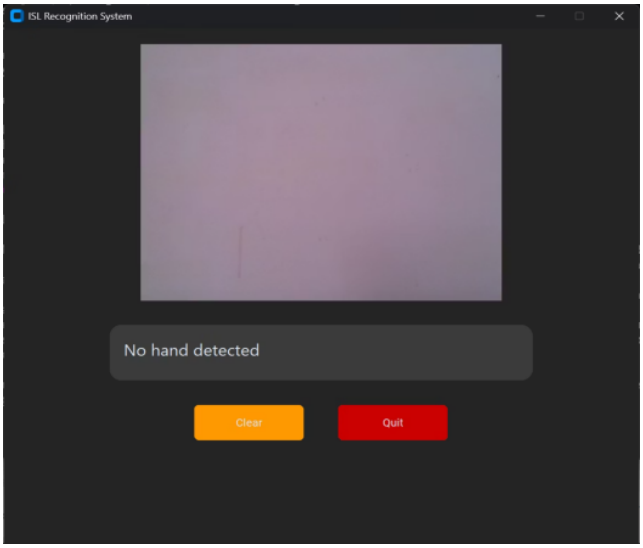


Fig-3 User Interface for Capturing Hand Gesture

During real-time execution, the interface continuously displayed webcam frames, predicted gesture outputs, and confidence scores. The developed system performed effectively under proper lighting conditions and clear hand positioning. The combined use of OpenCV, MediaPipe, and the trained CNN model enabled smooth real-time gesture recognition with stable performance and quick response time.

IV. RESULTS

The proposed Indian Sign Language (ISL) Alphabet Gesture Recognition System was successfully developed and tested using training, validation, and testing datasets. The MobileNetV2-based Convolutional Neural Network (CNN) model showed stable learning performance and achieved high prediction accuracy for static ISL alphabet recognition.

During training, the model learned important gesture features from the training dataset. The validation dataset was used to monitor model performance after each training epoch and help reduce overfitting during training. Techniques such as Batch Normalization, Dropout, Early Stopping, and data augmentation were applied to improve model generalization and maintain stable learning behavior.

The training and validation accuracy graph showed a gradual improvement in accuracy across multiple epochs, indicating successful model learning and effective feature extraction. The training and validation accuracy values remained closely aligned throughout the training process, showing minimal overfitting and stable performance. The use of MobileNetV2 transfer learning also helped improve classification performance and reduced overall training time.

The training and validation loss graph showed decreasing loss values with only small variations during training. This reduction in loss indicated improved prediction capability and effective optimization of the CNN model.

The small gap between training loss and validation loss further showed that the model generalized well on unseen gesture samples and maintained stable validation performance. Fig-4 shows the Training and Validation Accuracy graph and Fig-5 shows the Training and Validation Loss Graph of the proposed model.

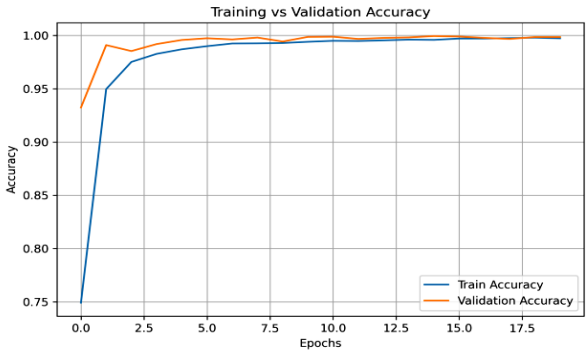


Fig-4 Training and Validation Accuracy Graph

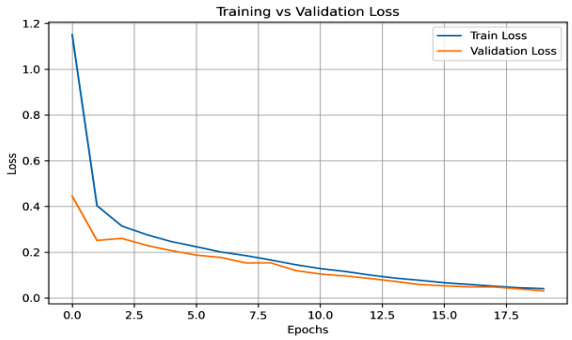


Fig-5 Training and Validation Loss Graph

The final trained model was evaluated using a separate testing dataset containing unseen gesture images. Experimental evaluation showed approximately 99% training and validation accuracy. During testing, the trained model achieved 99.90% accuracy with a loss value of 0.0292, showing the effectiveness of the proposed MobileNetV2-based architecture for static ISL alphabet recognition. The high prediction accuracy indicates that the CNN model was able to extract important gesture features from the input images effectively. Transfer learning, image preprocessing, and data augmentation techniques also helped improve prediction performance and model generalization. The trained model classified gesture images efficiently with minimal overfitting during training and validation.

The developed system was further tested with real time input using webcam. MediaPipe hand tracking successfully detected the hand region from live video frames, while the MobileNetV2 model predicted the corresponding ISL alphabet gesture accurately. The recognized alphabet and confidence score were displayed through the graphical user interface (GUI) developed using CustomTkinter.

Real-time testing showed stable prediction performance with fast response time under normal operating conditions. The system recognized ISL alphabet gestures more accurately when the hand region was clearly visible to the webcam. Smooth frame processing and effective integration between MediaPipe, OpenCV, and the trained CNN model also contributed to stable real-time gesture recognition performance.The system performed more effectively under proper lighting conditions and clear hand positioning. However, prediction accuracy may decrease under poor lighting conditions, cluttered backgrounds, partial hand visibility, or incorrect hand orientation. Some visually similar gestures may occasionally produce incorrect predictions during real-time testing.

The current implementation mainly focuses on static ISL alphabet recognition and serves as a foundation for future improvements such as dynamic gesture recognition, word-level and sentence-level translation, speech synthesis integration, and real-time assistive communication systems for hearing and speech-impaired individuals. Fig-6 shows the user interface of the proposed system to capture real-time gesture and predict the alphabet with percentage of confidence.

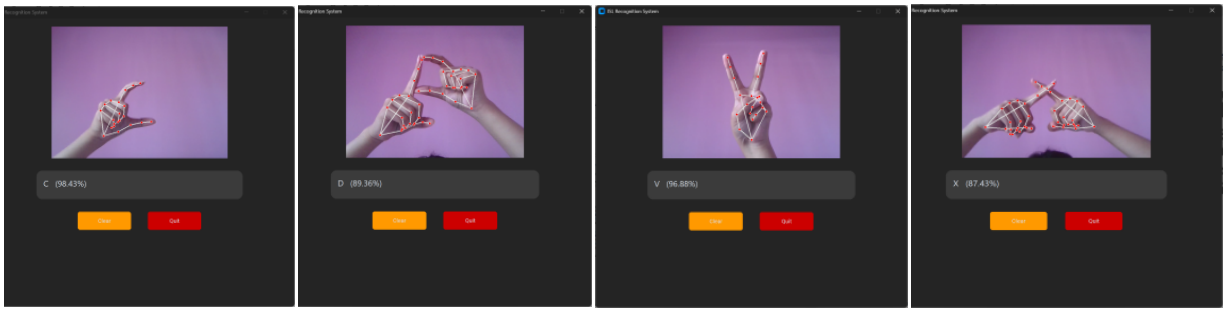


Fig-6 Sample of Some Real-Time Gesture Prediction Outputs

V. CONCLUSION

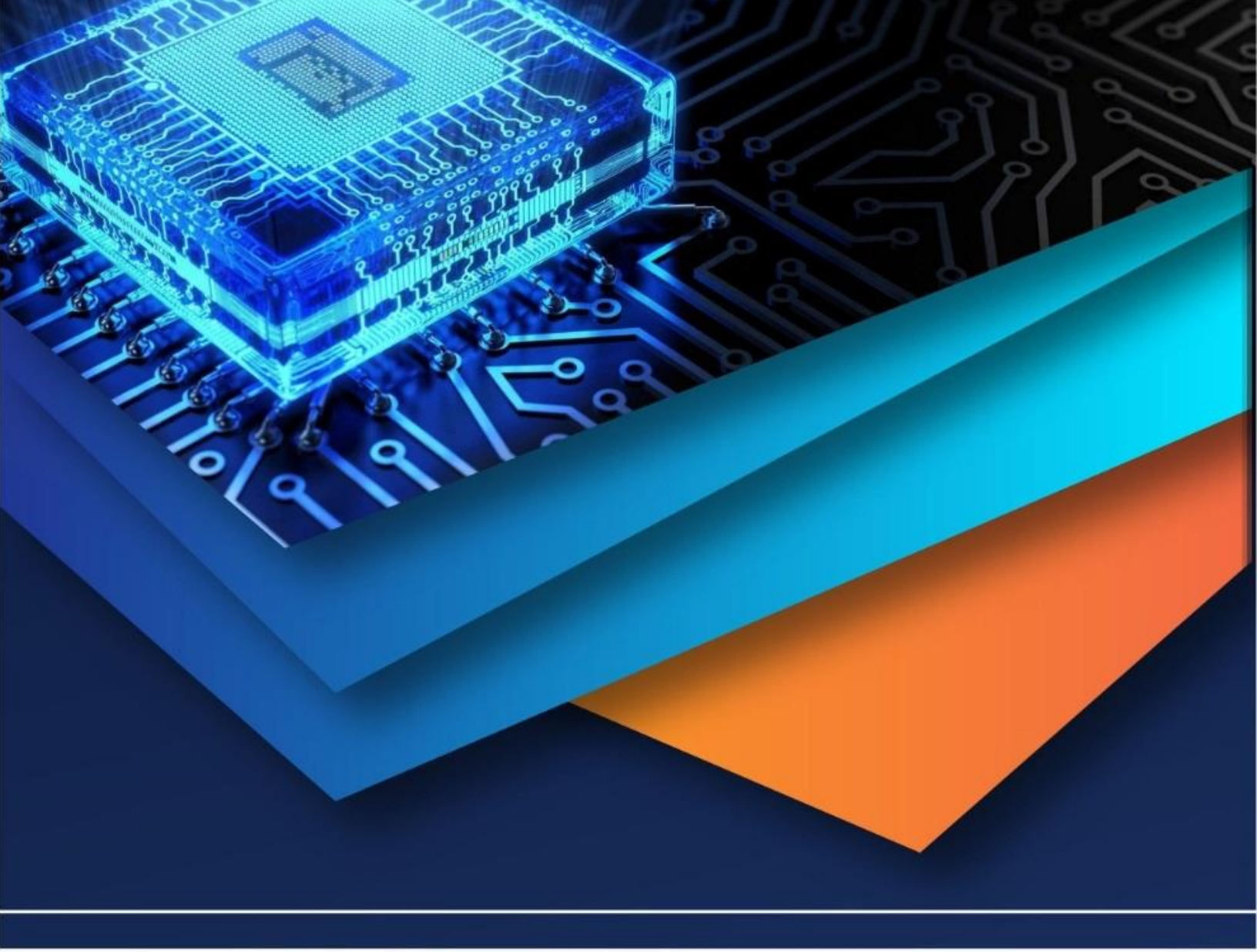
This research journal paper presented a real-time static Indian Sign Language (ISL) alphabet gesture recognition system using Deep Learning and Computer Vision techniques. The developed system successfully recognizes static ISL alphabet gestures using a webcam input device and converts them into readable text output. The proposed MobileNetV2 model showed stable learning performance and achieved high prediction accuracy during training and testing. Experimental evaluation demonstrated effective classification performance with minimal overfitting and efficient feature extraction capability. Real-time testing also showed accurate gesture recognition under proper lighting conditions and clear hand positioning. MediaPipe hand tracking with webcam-based prediction improved the practical usability and efficiency of the system. Image preprocessing and data augmentation techniques also helped improve model generalization and recognition performance under different gesture variations and environmental conditions. The proposed prototype aims to reduce communication barriers between hearing and speech-impaired individuals and non-sign language users by providing a simple and efficient real-time ISL recognition solution. Although the system achieved high accuracy for static ISL alphabet recognition, the current implementation is limited to recognizing only static hand gestures. Dynamic gesture recognition, continuous sentence interpretation, and complete sign language translation are not currently supported. Future improvements can be done by modeling a dynamic gesture recognition system using sequence-based deep learning models such as LSTM and Transformer architectures, sentence-level ISL translation, speech synthesis integration, mobile application deployment, and larger dataset support for improved robustness and real-time communication performance. Additional enhancements may also include multilingual output support and improved performance under complex real-world environments.

VI. DECLARATION

All the authors have declared that no conflict of interest exists

REFERENCES

[1] Rokade, Yogeshwar & Jadav, Prashant. (2017). Indian Sign Language Recognition System. International Journal of Engineering and Technology. 9. 189-196. 10.21817/ijet/2017/v9i3/170903S030.
[2] Patil, Rachana & Patil, Vivek & Bahuguna, Abhishek & Datkhile, Gaurav. (2021). Indian Sign Language Recognition using Convolutional Neural Network. ITM Web of Conferences. 40. 03004. 10.1051/itmconf/20214003004.
[3] Katoch, Shagun & Singh, Varsha & Tiwary, Uma Shanker. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array. 14. 100141. 10.1016/j.array.2022.100141.
[4] G, Dr & Goyal, Kaushal. (2023). Indian Sign Language Recognition Using Mediapipe Holistic. 10.48550/arXiv.2304.10256.
[5] Rupali Kadwade, Akanksha Tangade, Neha Pakhare, Samiksha Kolhe, Hajara Waikar, S. J. Wagh, 2023, Indian Sign Language Recognition System, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 12, Issue 05 (May 2023),
[6] Kumar, Anuj & Rawat, Prachi & Tamta, Vivek & Kumar, Papendra. (2025). A Comprehensive Approach to Indian Sign Language Recognition: Leveraging LSTM and MediaPipe Holistic for Dynamic and Static Hand Gesture Recognition. EAI Endorsed Transactions on AI and Robotics. 10.4108/airo.8693.
[7] Vashisth, Harsh & Tarafder, Tuhin & Aziz, Rehan & Arora, Mamta. (2023). Hand Gesture Recognition in Indian Sign Language Using Deep Learning. Engineering Proceedings. 59. 96. 10.3390/engproc2023059096.
[8] R. Awalkar, A. Sah, R. Barahate, Y. Kharche, and A. Magar, 'Silent expressions: Two-handed Indian Sign Language recognition using MediaPipe and machine learning', International Journal of Innovative Science and Research Technology (IJISRT), no. IJISRT25MAR598, pp. 587–595, Mar. 2025.
[9] Sahil Sagar Khetam, Komal Murkute, Sagar Surve, Ruturaj Bhunje, Pratham Raut, and Anurag Kumar, "Indian Sign Language to Text/Speech Translation: A Deep Learning", Int. J. Sci. Inno. Eng., vol. 2, no. 4, pp. 11–15, Apr. 2025, doi: 10.70849/IJSCI.
[10] Rastogi, Umang & Mahapatra, Rajendra & Kumar, Sushil. (2025). Advanced gesture recognition in Indian sign language using a synergistic combination of YOLOv10 with Swin Transformer model. Scientific Reports. 15. 10.1038/s41598-025-18496-8.
[11] Rushil Verma."Indian Sign Language Alphabet Dataset" Kaggle. https://www.kaggle.com/datasets/rushilverma07/indian-sign-language-alphabet-dataset



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY