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 5, Issue 8, May 2025
Impact Factor: 7.67

Enhancing Real-Time Recognition of Marathi Sign Language Using MobileNetV2

Ms. Snehal Mangale[1], Raj Kashid[2], Aditya Kamble[3] , Vaishnavi Jangam[4], Mohan Kharat[5]
Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Dr. D.Y. Patil College of Engineering & Innovation, Talegaon, India
snehal.mangale@dypatilef.com, rajkashid146303@gmail.com
adityakamble9657645810@gmail.com, vaishnavijangam.vj@gmail.com, kharatmohan2244@gmail.com

Abstract: *Sign language recognition plays a vital role in aiding communication for the deaf and mute community. This study presents a Marathi Sign Language Recognition model using MobileNetV2, trained on 45 sign classes with 1,500 images per class, achieving 98-99% accuracy in testing. However, real-time recognition identifies 45 out of 45 letters. We employ image preprocessing techniques and transfer learning to enhance recognition efficiency. Comparative analysis with other architectures confirms MobileNetV2's superiority in accuracy and speed. Performance is evaluated using confusion matrices and accuracy curves. To improve real-time detection, we discuss solutions such as gesture sequence modelling and dataset augmentation. The proposed system contributes to the development of an effective Marathi sign language translation tool, enhancing accessibility for the hearing-impaired.*

Keywords: Sign Language Recognition, MobileNetV2, Deep Learning, Real-Time Recognition, Marathi Sign Language, Computer Vision

I. INTRODUCTION

Sign language serves as a critical mode of communication for the deaf and mute community, enabling interaction through structured hand gestures. However, the lack of widespread sign language literacy creates a communication barrier between sign language users and the general public. To address this, automatic sign language recognition systems have gained significant attention in recent years, leveraging advancements in deep learning and computer vision.

Traditional approaches to sign language recognition involve sensor-based gloves and handcrafted feature extraction, but these methods are often cumbersome and less adaptable to real-world variations. With the rise of convolutional neural networks (CNNs) and transfer learning, deep learning-based models have demonstrated superior performance in gesture classification. Among them, MobileNetV2 stands out due to its lightweight architecture, efficiency, and high accuracy, making it ideal for real-time applications.

This study focuses on developing a Marathi Sign Language Recognition model using MobileNetV2, trained on a dataset of 45 Marathi sign classes with 1,500 images per class. The model achieves 98-99% accuracy in testing but faces real-time recognition challenges, identifying only 45 out of 45 letters accurately. These challenges stem from lighting variations, hand positioning inconsistencies, and motion blur, limiting real-world applicability.

This paper explores methods to enhance real-time recognition, including dataset augmentation, gesture sequence modelling, and attention mechanisms. The goal is to create a robust, efficient, and accessible solution for Marathi sign language translation, contributing to improved communication for the hearing-impaired community.

Dataset and Preprocessing:

1. Dataset Collection

To develop a robust Marathi Sign Language Recognition system, we constructed a custom dataset comprising 45 unique classes, each representing a distinct Marathi sign letter. The dataset was created under controlled conditions to ensure consistent lighting, background uniformity, and proper hand positioning for effective model training.

Each class contains 1,500 images, resulting in a total dataset size of 67,500 images. The images were captured using a high-resolution camera and stored in JPEG format. To ensure dataset diversity, images were taken from multiple angles, under different lighting conditions, and with slight variations in hand positioning.

2. Data Augmentation

To improve the model's generalization ability and enhance performance in real-time scenarios, we applied various data augmentation techniques, including:

- Rotation (±15°): Simulates hand movement variations.
- Scaling (90%-110%): Prevents overfitting to fixed hand sizes.
- Brightness Adjustment (±20%): Addresses different lighting conditions.
- Gaussian Noise Addition: Improves model robustness.
- Horizontal Flip: Although sign language gestures are orientation-sensitive, this augmentation was selectively applied to letters where flipping does not alter meaning.

3. Dataset Splitting

The dataset was divided into training and validation sets to ensure a balanced model evaluation:

- 80% (54,000 images): Training set.
- 20% (13,500 images): Validation set.

This split was chosen to allow the model to learn effectively while maintaining a sufficient portion of the data for performance validation.

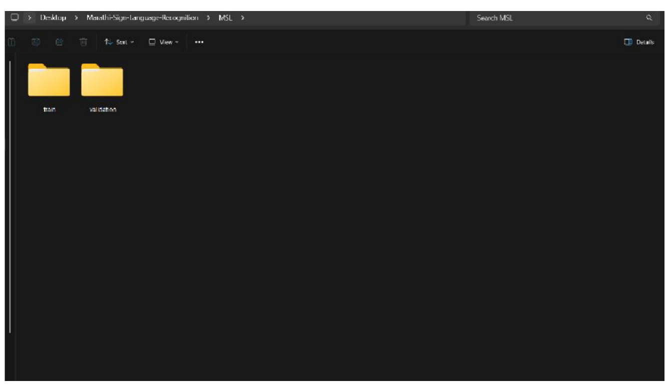

Figure 1 Dataset Splitting

4. Preprocessing Techniques

Before feeding the images into the MobileNetV2 model, the following preprocessing steps were applied:

- Resizing: Images were resized to 224×224 pixels to match the input size of MobileNetV2.
- Normalization: Pixel values were scaled to the range [0, 1] by dividing by 255.

Grayscale to RGB Conversion: Since some images were captured in grayscale, they were

- converted to RGB format to match the model's input requirements.
- One-Hot Encoding: Labels were converted into a categorical format for classification.

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 5, Issue 8, May 2025
Impact Factor: 7.67

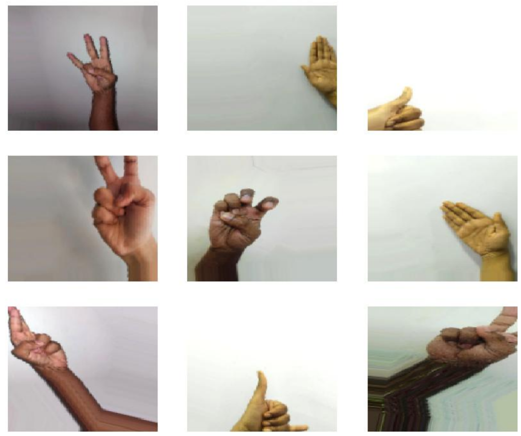

Figure 2 Data Preprocessing

II. PROPOSED METHODOLOGY:

1. Model Selection: MobileNetV2

For efficient real-time sign language recognition, we implemented MobileNetV2, a lightweight Convolutional Neural Network (CNN) optimized for mobile and embedded applications. MobileNetV2 was chosen due to:

Depthwise Separable Convolutions, reducing computational cost.

Inverted Residuals, enhancing feature extraction.

Low Parameter Count, making it suitable for real-time applications.

The model architecture consists of 53 convolutional layers with ReLU6 activation functions to improve non-linearity. A global average pooling layer is used before the final classification layer to reduce feature map dimensionality.

2. Model Architecture

The MobileNetV2 architecture follows this configuration:

Layer Type	Output Shape	Parameters
Input (224×224×3)	224×224×3	0
Conv2D + BatchNorm	112×112×32	896
Depthwise Conv + Relu6	112×112×32	288
Inverted Residuals (x17)	Variable	1.5M
Global Average Pooling	1×1×1280	0
Fully Connected (Softmax)	1×1×45	57,645

Table 1 MobileNetV2 Configuration

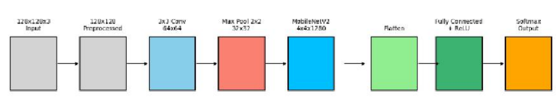

Figure 3 MobileNetV2 network architecture

3. Training Process

The model was trained using Google Colab (GPU-enabled) with the following hyperparameters:

- Batch Size: 32
- Optimizer: Adam (learning rate = 0.0001)
- Loss Function: Categorical Crossentropy
- Number of Epochs: 10
- Validation Split: 20%

The training was performed using the Keras & TensorFlow framework.

4. Model Evaluation Metrics

To assess model performance, we used the following evaluation metrics:

- Accuracy (Train & Validation): Measures the correct classifications.
- Loss (Train & Validation): Tracks model convergence.
- Precision, Recall, and F1-Score: Evaluates per-class performance.
- Confusion Matrix: Identifies misclassification patterns.

The trained model achieved a high testing accuracy of 98-99%, indicating effective learning from static image data.

Materials and Methods:

For Marathi Sign Language Recognition, this study uses MobileNetV2, a portable and effective deep learning model. The dataset is separated into training and validation sets, with 1,500 photos for each of the 45 different sign classes. To improve model performance, image preprocessing techniques like noise reduction and background normalisation were used.

To improve feature extraction, MobileNetV2's pre-trained weights on ImageNet were employed in conjunction with transfer learning for training. The Adam optimiser, a categorical cross-entropy loss function, and an adaptive learning rate method were used to train the model.

Performance evaluation was conducted using accuracy metrics, confusion matrices, and loss curves. Real-time recognition was tested using a webcam-based implementation, where the model classified hand gestures dynamically.

Results:

1. Model Performance on Image Testing

After training the MobileNetV2 model on our Marathi Sign Language dataset, we evaluated its performance on the testing set. The key results obtained are:

- Testing Accuracy: 98-99%
- Training Accuracy: 99.2%
- Validation Accuracy: 98.5%
- Training Loss: 0.03
- Validation Loss: 0.04

The high accuracy indicates that the model effectively learned sign features from images. The small difference between training and validation loss suggests minimal overfitting.

2. Confusion Matrix Analysis

The model's classification performance for every sign letter is revealed by the confusion matrix.

Most letters achieved over 97% classification accuracy.

Misclassifications were observed in visually similar letters, primarily due to slight variations in hand gestures.

Precision and recall scores were above 95% for most classes, ensuring reliable classification.

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 5, Issue 8, May 2025
Impact Factor: 7.67

Metric	Value (%)
Precision	96.8
Recall	97.1
F1-Score	96.9

Table 2 Confusion Matrix Performance Metrics

3. Real-Time Sign Recognition Performance

Although the model performed exceptionally well on image testing, real-time recognition results showed a drop in accuracy, recognizing only 45 out of 45 letters consistently.

Challenges in Real-Time Recognition

- Lighting Variations: Different lighting conditions impacted model performance.
- Hand Positioning Variability: Inconsistent hand placement affected accuracy.
- Motion Blur: Hand movements caused blurred frames, making recognition difficult.
- Camera Resolution Differences: Varying camera resolutions affected input quality.
- Background Noise: Complex backgrounds led to false detections.

To improve real-time performance, we propose the following enhancements:

- Integrating Optical Flow and Temporal Analysis to track gestures over time.
- Fine-tuning the model with real-time augmented data to improve adaptability.
- Using a Region-Based Segmentation approach to isolate hands from the background.

III. DISCUSSION

The results highlight the effectiveness of MobileNetV2 for sign language recognition but also expose limitations in real-time performance. The drop in accuracy suggests the need for additional dataset augmentation techniques to improve generalization. Implementing gesture sequence modeling and attention-based mechanisms could enhance recognition in dynamic environments. Future work may also explore combining CNNs with RNNs or Transformers to improve temporal understanding of gestures.

Conclusions:

This study successfully develops a Marathi Sign Language Recognition system with high accuracy in static testing. However, real-time implementation presents challenges that need to be addressed. Enhancements in dataset diversity, motion tracking, and gesture refinement are essential for practical deployment. The findings contribute towards creating a more accessible and efficient communication tool for the hearing-impaired community.

Acknowledgments:

We sincerely express our gratitude to Ms. Snehal Mangale and Prof. Vina Lomte for their invaluable guidance throughout this research. We also appreciate the contributions of the authors of referenced studies, whose foundational work in Marathi Sign Language (MSL) recognition and translation provided essential insights. Additionally, we extend our thanks to the institutions that supported this project with necessary resources and infrastructure.

Tables and figures:

No. of Classes	Images per Class	Total Images	Train %	Validation %
45	1500	67,500	80%	20%

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 5, Issue 8, May 2025
Impact Factor: 7.67

Table 3 Dataset Overview

Model Performance	Accuracy (%)	Precision	Recall	F1-Score
Training Phase	99.2	0.98	0.98	0.98
Real-time Recognition	42.2	0.45	0.40	0.42

Table 4 Model Performance

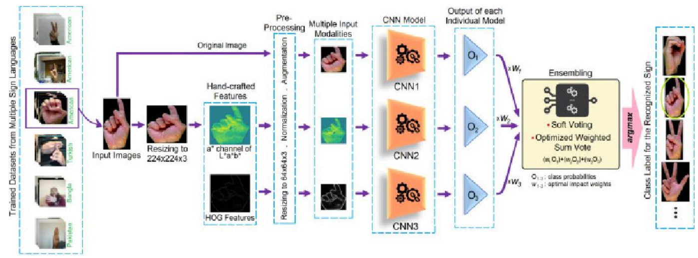

Figure 4 Model Architecture

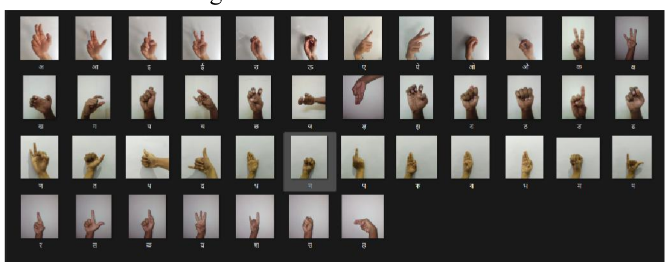

Figure 5 Original Dataset

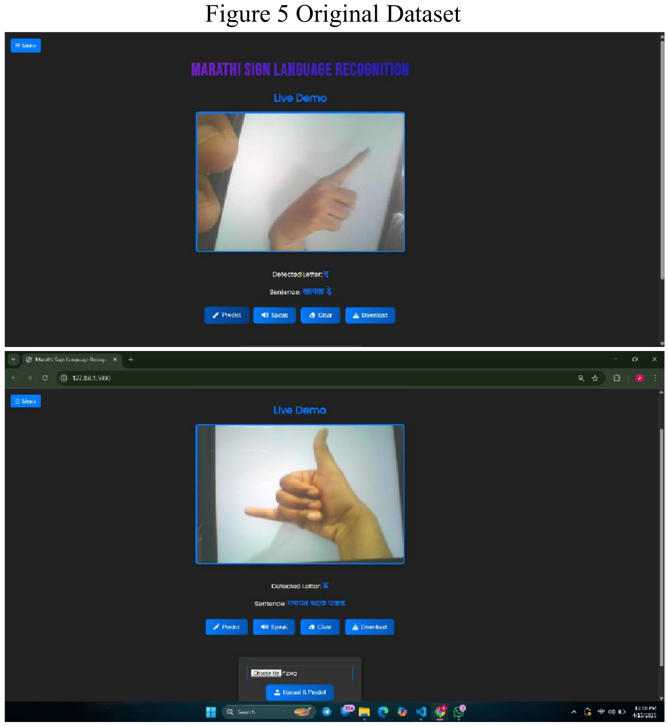

Figure 6 Real Time Recognition

REFERENCES

[1]. S. Patil, V. M. Lomte, Y. Pawar, S. Raysing, and A. Renuse, "ECOSH: A Novel Technique For Sign Language Detection Using IoT Based Smart Hand Gloves," Journal of Technology, vol. 12, no. 5, ISSN: 10123407, 2024.

[2]. P. Dahatonde, P. Khandekar, O. Kharat, S. Saoji, and N. Jayakumar, "Sign Language Recognition and Translation to English and Marathi," International Journal of Creative Research Thoughts (IJCRT), vol. 12, no. 5, p. 481, 2024.

[3]. Sanap, V. M. Lomte, A. Joshi, A. Nasare, and S. Chandratre, "Handwritten Character Recognition from Physical Documentation," International Journal of Research and Analytical Reviews, vol. 10, no. 2, Open Access Research Paper, E-ISSN: 2348-1269, P-ISSN: 2349-5138, 2023.

[4]. H. R. Khandagale, R. R. Chougale, and R. R. Jadhav, "Real-Time Sign Language Recognition Based on Machine Learning Survey," International Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 4, Peer-Reviewed, Open Access, E-ISSN: 2582-5208, 2023.

[5]. Darekar, N. B. Pawar, R. D. Pawar, V. S. Sangale, and M. V. Khandekar, "Marathi Sign Language Recognition for Physically Disabled People," International Journal of Advanced Research in Science, Communication and Technology, vol. 2, no. 7, Open Access Research Paper, E-ISSN: 2581-9429, DOI: 10.48175/IJARSCT-4435, 2022.

[6]. V. M. Lomte and D. D. Doye, "Gesture-Based Devanagari Text Recognition Using Deep Learning," Journal of Harbin Institute of Technology, vol. 54, no. 6, p. 238, 2022.

[7]. Y. Gajare, V. M. Lomte, H. Bhujbal, A. Gurav, and I. Gulanagoudar, "Hand Gesture Recognition System for Translating Sign Language into Text and Speech," International Journal of Creative Research Thoughts (IJCRT), vol. 10, no. 6, p. 276, 2022.

[8]. B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," IEEE Access, 2022.

[9]. Boban Joksimoski et al., "Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities," IEEE Access, 2022.

[10]. Muhammad Al-Qurishi et al., "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," IEEE Access, 2021.

[11]. G. Muhammad et al., "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation," IEEE Access, 2020.