

Feature Extraction and Hyperparameter Optimization Using MobileNet for Conversion from Sign Language to Text

Dr. R. Kavitha¹, S. Infant Triphina², K. Atchaya³, M. Roshini⁴, S. Roshini⁵

¹ Professor, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Email: pits.hod.cse@gmail.com

²UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Email: triphinaraj26@gmail.com

³UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu 613006, India
Email: atchayakarthika04@gmail.com

⁴UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Email: roshiniradhika22@gmail.com

⁵UG Student, Department of Computer Science and Engineering, Parisutham Institute of Technology and Science, Thanjavur, Tamil Nadu – 613006, India
Email: roshini200411@gmail.com

Abstract:

Sign language is an important communication method for individuals with hearing and speech impairments. However, communication between sign language users and people who do not understand sign language remains challenging. This paper presents a deep learning-based system for Sign Language to Text Conversion using the MobileNetV2 architecture for efficient feature extraction. Hand gesture images are first preprocessed to improve image quality and then passed through a MobileNetV2-based Convolutional Neural Network (CNN) to extract important features. To improve model performance, hyperparameter optimization techniques such as learning rate tuning, batch size adjustment, optimizer selection, dropout regularization and epoch variation are applied. The extracted features are classified using a Softmax classifier to recognize gestures and convert them into readable text. The system provides an efficient solution for assistive communication and accessibility applications.

Keywords — Sign Language Recognition, MobileNetV2, Feature Extraction, Hyperparameter Optimization, Softmax Classification, Assistive Communication.

I. INTRODUCTION

The growing need for inclusive communication has increased the importance of assistive technologies for individuals with hearing and speech impairments. Sign language is widely used by such individuals, but the lack of understanding among the general public creates

communication barriers. This highlights the need for automated systems that can interpret hand gestures and convert them into readable text. Recent advancements in computer vision and deep learning have enabled accurate image-based recognition systems. Convolutional Neural Networks (CNNs) are particularly effective in extracting meaningful features from gesture images. The proposed

system utilizes the MobileNetV2 architecture to develop a lightweight and efficient sign language recognition model. Hand gesture images are captured and processed through preprocessing techniques such as color space conversion, resizing, normalization, noise reduction and data augmentation. Deep feature extraction is performed using a pre-trained MobileNetV2 model with transfer learning. The extracted features are classified into alphabetic and numeric outputs using optimized parameters and a softmax layer. The system achieves high accuracy with low computational complexity, making it suitable for real-time assistive communication applications. Furthermore, the proposed system emphasizes scalability and ease of deployment on resource-constrained devices. It also ensures faster inference speed, making it practical for real-time usage in everyday communication scenarios. Overall, the system contributes to improving accessibility and bridging the communication gap between individuals with hearing and speech impairments and the general public.

II. RELATED WORK

Al-Hammadi M et al. [1] proposed a deep learning approach for sign language recognition using a 3D Convolutional Neural Network (3DCNN). Their work focuses on recognizing dynamic hand gestures by extracting spatial and temporal features from video sequences. The model analyzes motion patterns across consecutive frames for accurate gesture identification. The study also applies transfer learning techniques to improve performance with limited datasets. Experimental results show that the proposed method achieves higher accuracy compared to traditional and 2D CNN-based approaches.

Al Quirishi M et al. [2] presented a comprehensive review of sign language recognition systems using machine learning and deep learning techniques. Their study analyzes methods developed between 2014 and 2021, highlighting strengths and limitations. The research examines sensor-based, vision-based and hybrid data acquisition approaches. It also reviews models such as CNN, RNN and HMM for effective feature extraction from gesture data. The results indicate that deep learning-based methods achieve better accuracy and performance compared to traditional approaches, though challenges in real-time implementation still remain.

Aruna Bhat et al. [3] proposed a deep learning-based approach for converting sign language gestures into text using Convolutional Neural Networks (CNN). Their work focuses on recognizing hand gestures from images using preprocessing techniques like grayscale conversion, Gaussian blur, thresholding and resizing. The CNN model extracts spatial features for accurate classification. Experimental results show good recognition accuracy in controlled environments. However, the study mainly focuses on static gestures rather than dynamic gesture recognition.

Sumit Kumar et al. [4] explored a deep learning-based real-time sign language detection system for assisting communication of hearing and speech-impaired individuals. Their study focuses on recognizing Indian Sign Language (ISL) gestures using computer vision and deep learning techniques. A pre-trained VGG16 Convolutional Neural Network combined with an attention mechanism is used to improve classification accuracy. Hand gesture images are preprocessed and transfer learning is applied for feature extraction. Experimental results show that the model achieves high recognition accuracy with low latency, making it suitable for real-time applications.

Tangfei Tao et al. [5] presented a comprehensive review of sign language recognition techniques using both traditional machine learning and deep learning methods. Their study discusses various feature extraction techniques, classification models and neural network architectures such as CNN and RNN. The paper highlights challenges including signer variability, lighting conditions, real-time processing limitations and dataset imbalance. The results indicate that deep learning approaches significantly improve recognition accuracy and system robustness compared to traditional methods.

Bashear Al Abdullah A et al. [6] presented a detailed review of recent advancements in sign language recognition systems. Their study analyzes both traditional machine learning and deep learning approaches for recognizing hand gestures. The research highlights the importance of computer vision techniques and neural networks. It also examines challenges such as variations in hand shape, lighting conditions, background noise and signer dependency, along with comparisons between vision-based and sensor-based methods. The results indicate that deep learning models provide improved accuracy and robustness, but require large datasets and high computational resources.

Fatma Najib M [7] proposed a machine learning-based approach for multilingual sign language recognition using a system called Multi-Lingual Sign Languages Interpreter (MSLI). Their study focuses on recognizing hand gestures from multiple sign languages using a single trained model. Hand gesture images from different datasets are preprocessed and relevant features are extracted for classification. The model is trained using machine learning techniques and evaluated on eleven different sign language datasets. Experimental results show that the system achieves high recognition accuracy and efficient performance across multiple sign languages.

Mukhiddin Toshpulatov et al. [8] explored deep learning approaches used in automatic sign language processing. Their study focuses on recognition, translation, production and dataset development using modern deep learning techniques. The paper surveys various machine learning and deep learning methods and highlights the limitations of traditional approaches based on handcrafted features. Deep neural networks such as CNN, RNN, Transformers and Graph Neural Networks are analyzed for their ability to learn spatial and temporal features.

Experimental results indicate that deep learning models significantly improve recognition accuracy, though challenges such as dataset limitations and computational complexity remain.

Karkar A et al. [9] proposed a real-time sign language to text translation system using deep learning techniques. Their work compares the performance of Long Short-Term Memory networks and 3D Convolutional Neural Network. The LSTM model captures temporal dependencies in sequential data, while the 3D CNN extracts spatial features from video frames. Additionally, a hybrid CNN-LSTM model is introduced to combine both spatial and temporal learning. Experimental results demonstrate that the hybrid model achieves improved accuracy and better real-time performance compared to individual LSTM and 3D CNN models.

Sharath S R et al. [10] proposed a sign language to sign language translation system aimed at enabling communication across different sign languages. Their approach utilizes keypoint-based gesture extraction and converts gestures into an intermediate representation using HamNoSys notation. A sequence-to-sequence (Seq2Seq) model is then employed to translate the gestures into the target sign language. The system focuses on handling multiple sign languages such as ISL, ASL and FSL. Experimental results show that the proposed model achieves efficient translation with improved sequence accuracy.

III. PROPOSED METHODOLOGY

The proposed system is designed to provide an efficient and lightweight solution for image-based sign language to text conversion using deep learning techniques. The system integrates image acquisition, preprocessing, feature extraction using the MobileNetV2 architecture, hyperparameter optimization and classification within a unified framework. By utilizing a 2D Convolutional Neural Network instead of computationally intensive 3D models, the approach reduces complexity while maintaining high recognition accuracy. Hand gesture images are processed through resizing, normalization, noise reduction and data augmentation to ensure robustness under varying conditions. Deep feature extraction is performed using a pre-trained MobileNetV2 model with transfer learning to adapt to the sign language recognition task. The system further enhances performance through optimization of key hyperparameters and employs a softmax-based classification layer to map features to corresponding text outputs. Overall, the proposed system achieves accurate and fast recognition, making it suitable for real-time assistive communication applications while effectively reducing communication barriers.

A. Image Acquisition and Preprocessing

The first stage of the proposed system focuses on capturing hand gesture images and preparing them for further processing. The system uses a standard camera to acquire real-time gesture images or utilizes a pre-collected

dataset. These images are then subjected to preprocessing techniques to improve quality, consistency and robustness under varying environmental conditions. During this stage, several important operations are performed, including:

- 1) *Image Acquisition*: Hand gesture images are captured using a camera or loaded from a dataset for training and testing purposes.
- 2) *Color Space Conversion*: Input images are converted from BGR to RGB format to ensure correct color representation and compatibility.
- 3) *Image Resizing*: All input images are resized to a fixed dimension to ensure uniformity and compatibility with the model input requirements.
- 4) *Normalization*: Pixel values are normalized to a specific range to improve model convergence and performance.
- 5) *Data Augmentation*: Techniques such as rotation, flipping, zooming and shifting are applied to increase dataset diversity and reduce overfitting.
- 6) *Noise Reduction*: Basic filtering techniques are used to remove unwanted noise and improve image clarity.

B. Feature Extraction Using MobileNetV2

The system performs deep feature extraction using the MobileNetV2 architecture, which is a lightweight Convolutional Neural Network designed for efficient computation and high performance.

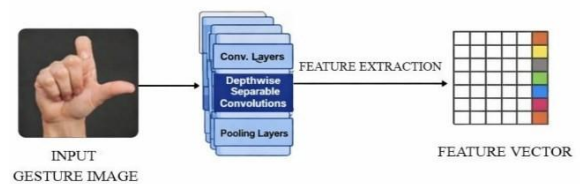


Fig 1. Feature Extraction

The model is pre-trained on the Image dataset and is adapted to the sign language recognition task using transfer learning. The following operations are performed during the feature extraction process:

- 1) *Convolutional Layer*: The input gesture image is passed through initial convolutional layers to extract basic visual features such as edges, textures, and shapes.
- 2) *Depthwise Separable Convolution*: The standard convolution is factorized into depthwise and pointwise convolutions, significantly reducing the number of parameters and computational cost while preserving feature quality.
- 3) *Batch Normalization*: Normalization is applied after convolutional operations to stabilize learning, accelerate training, and improve model performance.
- 4) *ReLU Activation Function*: The Rectified Linear Unit (ReLU) introduces non-linearity into the model, enabling it to learn complex patterns in hand gesture images.
- 5) *Feature Extraction*: Stacked convolutional blocks

progressively learn higher-level and more abstract features from the input images.

6) *Global Average Pooling (GAP)*: The feature maps are reduced into a compact feature vector by averaging spatial dimensions, minimizing overfitting and reducing model parameters.

7) *Feature Vector Generation*: The final output is a meaningful feature representation that captures important gesture characteristics and is passed to the classification stage.

C. Hyperparameter Optimization

This stage focuses on improving the performance and accuracy of the model by tuning important hyperparameters. Initially, the model was trained using default parameter settings, which resulted in moderate accuracy, slower convergence and slight overfitting. These limitations indicated the need for optimizing the training process to achieve better results. Hyperparameter optimization is performed to adjust critical training parameters such as learning rate, batch size, number of epochs, dropout rate and fine-tuning depth. By carefully selecting these values, the model is able to learn more effectively, reduce training errors and improve generalization on unseen data.

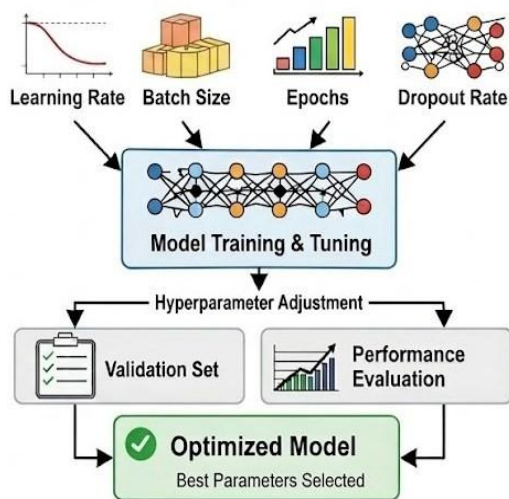


Fig. 2 Hyperparameter Optimization

Hyperparameter optimization includes:

1) *Learning Rate Tuning*: The optimal learning rate is selected to ensure stable and efficient training of the model. Learning rate set to $1e-5$ to ensure stable and gradual convergence during training.

2) *Batch Size Selection*: The number of samples processed per training iteration is adjusted to balance memory usage and training speed. It fixed at 32 to balance computational efficiency and training stability.

3) *Epoch Optimization*: The number of training epochs is

determined to achieve maximum accuracy without overfitting. The model is trained for 15 epochs to achieve optimal learning.

4) *Dropout Regularization*: Dropout layers are used to prevent overfitting by randomly disabling neurons during training. A value of 0.5 is applied to prevent overfitting by randomly disabling neurons during training.

5) *Fine-Tuning Depth*: Specific layers of the MobileNet model are fine-tuned to improve task-specific performance.

6) *Performance Evaluation*: The model is evaluated using validation data to select the best set of hyperparameters.

D. Classification and Text Conversion

This stage focuses on converting the extracted features into meaningful class predictions and generating corresponding text outputs. The refined feature vectors obtained from the MobileNetV2 model are processed through classification layers to accurately identify hand gestures. This stage ensures effective mapping between learned features and output labels while maintaining high prediction accuracy.

1) *Fully Connected Layer*: Extracted features are passed through dense layers for classification.

2) *Softmax Activation*: Softmax function is used to assign probability scores to each class.

3) *Classification of gesture*: The system predicts the corresponding alphabet or number based on the highest probability.

4) *Text Conversion*: The predicted gesture is converted into readable text output for user interaction.

E. Output Generation

After completing the classification process, the system generates the final output by converting the predicted gesture into a readable text format. During this stage, the model produces raw output scores, which are passed through the Softmax function to obtain probability values for each gesture class. The class with the highest probability is selected using the argmax function, resulting in the predicted class index. This index is then mapped to its corresponding gesture label using a predefined class dictionary. The processed output maintains consistency with the trained dataset by accurately representing the recognized gesture. Finally, the system prepares and displays the recognized text in the user interface, enabling effective communication without manual interpretation.

IV. CONCLUSIONS

The proposed system addresses the communication challenges faced by hearing and speech-impaired individuals by providing an automated sign language recognition solution. The system utilizes deep learning techniques to interpret hand gestures and convert them into meaningful text. By integrating modules such as image acquisition,

preprocessing, feature extraction using the MobileNet architecture, hyperparameter optimization and classification, the system ensures efficient and accurate gesture recognition. The use of a lightweight model improves computational efficiency while maintaining reliable performance. Experimental results demonstrate that the system effectively recognizes sign language gestures and generates appropriate text outputs. Overall, the proposed solution contributes to enhancing accessibility and reducing communication barriers through the application of artificial intelligence. Future improvements may include expanding the dataset, supporting real-time continuous recognition and increasing system robustness in diverse environments.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Dr. R. Kavitha, Professor, Department of Computer Science and Engineering, for her valuable guidance, continuous support and encouragement throughout the development of this work.

The authors also extend their gratitude to their external guide, Mr. P. Boopathy Pandi, from Monzha Research Labs, for providing valuable insights, technical guidance and necessary resources that contributed to the successful completion of this work.

We also thank the faculty members of the department for their support and constructive suggestions. The authors acknowledge the institution for providing the necessary facilities and resources to successfully complete this research.

REFERENCES

1. Al-Hammadi M., Muhammad G., Abdul W., Alsulaiman M., Bencherif A., Mekhtiche M., "Hand Gesture Recognition for Sign Language Using 3DCNN," *IEEE Access*, vol. 8, pp. 79491–79509, 2020.
2. Al-Qurishi M., Khalid T., Souissi R., "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," *IEEE Access*, vol. 9, pp. 126917–126945, 2021.
3. Bhat A., Yadav V., Dargan V., Yash Y., "Sign Language to Text Conversion using Deep Learning," in *Proc. 3rd Int. Conf. on Emerging Technology (INCET)*, 2022.
4. Kumar S., Rani R., Chaudhari U., "Real-time Sign Language Detection: Empowering the Disabled Community," *MethodsX*, vol. 13, p. 102901, 2024.
5. Tao T., Zhao Y., Liu T., Zhu J., "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges," *IEEE Access*, vol. 12, pp. 75034–75060, 2024.
6. Al Abdullah B. A., Amoudi G. A., Alghamdi H. S., "Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects," *IEEE Access*, vol. 12, pp. 128871–128890, 2024.
7. Najib F. M., "A Multi-Lingual Sign Language Recognition System Using Machine Learning," *Multimedia Tools and Applications*, vol. 84, pp. 27988–28011, 2025.
8. Tsoplatov M., Wookyeol W., Jun J., Lee S., "Deep Learning Pathways for Automatic Sign Language Processing," *Pattern Recognition*, vol. 164, p. 111475, 2025.
9. Karkar A., Khot A., Andure A., Ghosh A., Magadam A., Bahadur A., Pol M., "Real-Time Sign Language to Text Translation using Deep Learning: A Comparative Study of LSTM and 3D CNN," *International Journal of Computer Applications*, vol. 187, no. 35, pp. 31–36, 2025.
10. Sharath S. R., Suraj S., Abishek Kumar G. M., P. Siddharth, NalinaDevi K., "Sign Language to Sign Language Translator," *Procedia Computer Science*, vol. 260, pp. 373–381, 2025.