ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr960

Indian Sign Language Alphabet Recognition Using a Hybrid CNN-PCA-PSO-SVM Framework

Dr. Girish Katkar[1]; Shalaka Gaikwad[2]; Dr. Ajay Ramteke[3]

[1]Associate Professor, Department of Computer Science, Taywade College, Koradi Indian
[2]Research Scholar, Department of Computer Science, Taywade College, Koradi Indian
[3]Assistant Professor, Department of Computer Science, Taywade College, Koradi Indian

Publication Date: 2026/04/27

Abstract: In this work, we propose an efficient hybrid framework for recognizing Indian Sign Language (ISL) alphabets by combining deep feature extraction with classical machine learning and optimization techniques. A pre- trained MobileNetV2 network is utilized to extract discriminative visual features from hand gesture images. These high-dimensional features are subsequently compressed using Principal Component Analysis (PCA) to eliminate redundancy and improve computational efficiency. Particle Swarm Optimization (PSO) is then employed to determine optimal hyper-parameters for a Support Vector Machine (SVM) classifier. The proposed system is evaluated on a 26-class ISL dataset and achieves an overall accuracy of 96.17%, along with consistently high precision, recall, and F1-scores. Further validation using confusion matrix and ROC analysis demonstrates the robustness and strong class separability of the model.

Keywords: *Indian Sign Language Recognition, MobileNetV2, PCA, PSO, SVM.*

How to Cite: Dr. Girish Katkar; Shalaka Gaikwad; Dr. Ajay Ramteke (2026) Indian Sign Language Alphabet Recognition Using a Hybrid CNN-PCA-PSO-SVM Framework. *International Journal of Innovative Science and Research Technology*, 11(4), 1954-1959. https://doi.org/10.38124/ijisrt/26apr960

I. INTRODUCTION

Automatic ISL recognition enables inclusive communication for the hearing- impaired community. While deep CNNs provide strong feature representations, hybrid pipelines can further improve classification efficiency. Our work proposes a CNN-PCA-PSO-SVM framework to balance accuracy and computational cost.

Communication between hearing-impaired individuals and the general population is often hindered by the lack of effective real-time translation tools. Indian Sign Language (ISL) is one of the primary modes of communication for the deaf community in India, and automated recognition of ISL alphabets can significantly enhance accessibility in education, social interactions, and digital applications.

Recent advancements in deep neural architectures have significantly improved visual gesture recognition accuracy in image-based recognition tasks due to their ability to automatically learn hierarchical features from raw image data. MobileNetV2, a lightweight CNN architecture, provides an efficient solution for feature extraction from ISL hand gesture images, making it suitable for real-time applications [1], [2].

However, CNNs typically produce high-dimensional feature vectors, which can increase computational complexity and reduce classifier efficiency. Principal Component Analysis (PCA) is an effective dimensionality reduction technique that projects CNN-extracted features onto a lower-dimensional subspace, preserving the most informative components while reducing noise and redundancy [3].

For classification, Support Vector Machines (SVMs) are robust and widely used, particularly in small- to medium-sized datasets. Non-linear SVMs, such as those using the Radial Basis Function (RBF) kernel, provide high accuracy by mapping input features into a higher-dimensional space. To further enhance SVM performance, Particle Swarm Optimization (PSO) is employed for hyperparameter tuning, efficiently searching for optimal values of kernel parameters and regularization coefficients [4].

The hybrid approach combining MobileNetV2 for feature extraction, PCA for dimensionality reduction, PSO for SVM hyperparameter optimization, and SVM for final classification leverages the strengths of each technique. This results in a computationally efficient, accurate, and scalable

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr960

framework for recognizing ISL alphabets in real-time, bridging the communication gap between the hearing-impaired and the general population.

II. RELATED WORK

Recent studies have shown that convolutional neural networks (CNNs) and transfer learning are effective for sign- language recognition. In particular, MobileNetV2 has become popular because its lightweight design makes it efficient on mobile and embedded devices [5]. CNN-based methods generally perform very well on gesture recognition. For example, some researchers applied CNNs directly to sign language classification and saw much better results than older approaches [6]. Similarly, techniques using pretrained models like VGG or ResNet (via transfer learning) have been shown to generalize well to new sign datasets [7].

Researchers have also explored hybrid models that combine CNNs with Support Vector Machines (SVMs). In these approaches, the CNN extracts features from the images, and the SVM uses those features to classify the signs. This often leads to stronger decision boundaries between classes [8]. In addition, optimization algorithms like Particle Swarm Optimization (PSO) have been used to automatically tune the model's hyperparameters, which can further improve accuracy [9].

Most recent work (around 2023–2025) focuses on creating very lightweight, hybrid systems that can run in real time. Despite these advances, challenges remain, such as accurately recognizing dynamic gestures (like motion- based letters) and distinguishing between signs that look very similar. Our proposed method builds on this prior work by putting together CNN feature extraction, PCA dimensionality reduction, PSO-based hyperparameter tuning, and SVM classification into a single framework, aiming to improve overall performance even further.

III. METHODOLOGY

➢ *Dataset and Pre-processing:*

The dataset used in this study consists of 26 classes corresponding to the Indian Sign Language (ISL) alphabets, with a total of 1560 images (60 images per class). All images are resized to 224×224 pixels and normalized to the [0,1] range. To ensure unbiased evaluation, the dataset is divided into three subsets: 80% for training, 10% for validation, and 10% for testing. Data augmentation techniques, including rotation and horizontal flipping, are optionally applied to enhance model generalization.

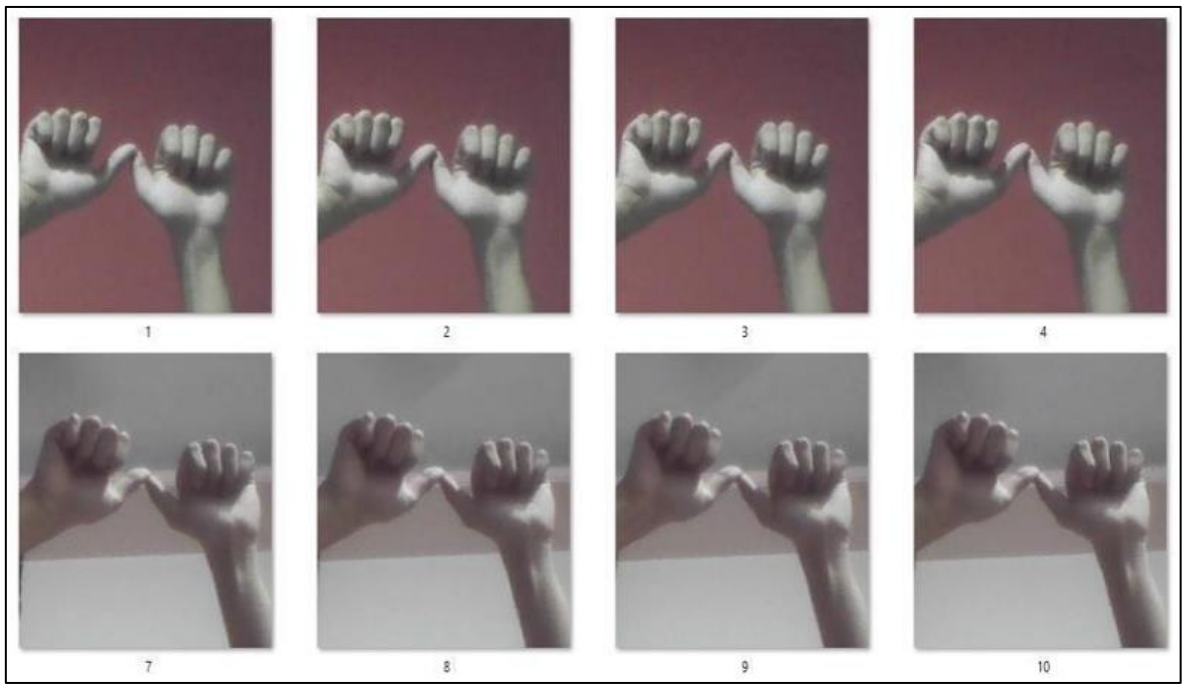

Fig 1 Indian Sign Language (ISLRTC referred)

➢ *Feature Extraction Using CNN:*

A pre-trained MobileNetV2 model is employed as a feature extractor. The fully connected layers are removed, and global average pooling is applied to generate a compact feature vector for each image. The CNN captures the spatial structure and shape characteristics of hand gestures, providing robust representations for subsequent classification. The extracted feature vectors serve as the input for dimensionality reduction and optimization.

➢ *Dimensionality Reduction Using PCA:*

To reduce feature redundancy and computational complexity, Principal Component Analysis (PCA) is applied to the CNN-extracted feature vectors. PCA projects the high-dimensional features onto a lower-dimensional subspace while preserving the majority of variance. In this study, the number of principal components is chosen to retain 95% of the cumulative variance. This step ensures efficient computation in the subsequent optimization and classification stages.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr960

➢ *Hyperparameter Optimization Using PSO:*

Particle Swarm Optimization (PSO), a soft computing technique inspired by the social behavior of bird flocks, is employed to optimize the hyperparameters of the SVM classifier, specifically the penalty parameter (C) and the kernel coefficient (γ) for the Radial Basis Function (RBF) kernel. Each particle in the swarm represents a candidate solution, and the fitness function is defined as one minus the classification accuracy on the validation dataset. PSO iteratively updates the particle positions and velocities to find the optimal hyperparameters that maximize classification performance. The swarm size is set to 10, and the maximum number of iterations is 5.

➢ *Classification Using SVM:*

The optimized feature subset obtained after PCA and PSO is fed into an SVM classifier with an RBF kernel. The SVM is trained on the training dataset (and optionally fine-tuned using validation data) and evaluated on the independent test set. The classifier's performance is measured using accuracy, precision, recall, and F1-score. Additionally, a confusion matrix is generated to analyze class-wise performance.

➢ *Overall Hybrid Framework:*

The proposed hybrid framework integrates deep learning, dimensionality reduction, and soft computing optimization to provide an efficient ISL recognition system. The workflow follows the sequence:

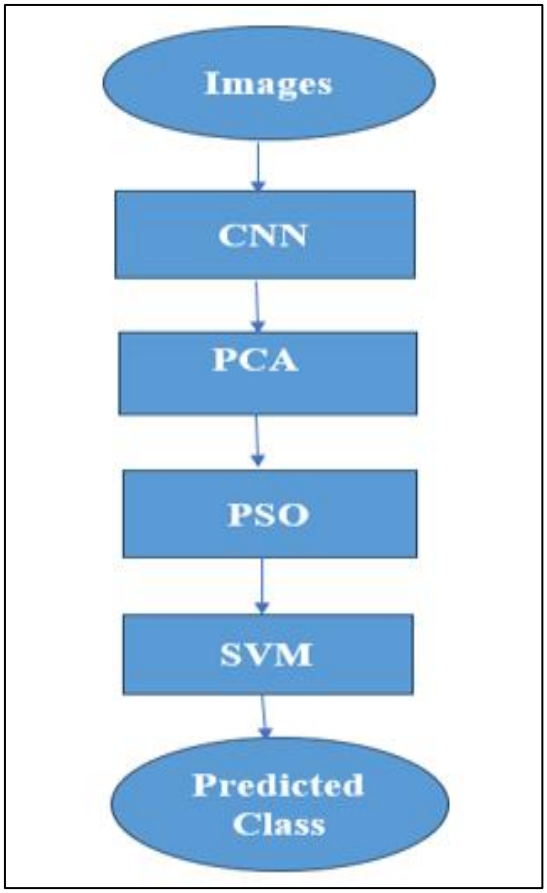

Fig 2 Block Diagram of the Proposed CNN-PCA-PSO-SVM Hybrid ISL Recognition Framework.

IV. COMPUTATIONAL COMPLEXITY ANALYSIS

Modern sign language recognition pipelines often combine deep feature extraction, dimensionality reduction, and optimized classification. In this work, MobileNetV2 is used for feature extraction, PCA for dimensionality reduction, SVM for classification, and PSO for hyperparameter optimization. The computational complexity, inference cost, and memory requirements of each module are discussed below.

➢ *MobileNetV2 CNN:*

Modern convolutional neural networks (CNNs) employ optimized convolution operations to reduce computational cost while maintaining high accuracy. For a standard 2D convolution on an input of size H x W x Cin with Cout filters of size k × k, the computational complexity is O (H. W. Cin. Cout.k^2). MobileNetV2 replaces standard convolutions with depth wise-separable convolutions, decomposing the operation into a depth wise convolution with complexity O(H·W·Cin·k^2) and a pointwise (1×1) convolution of complexity O(H·W·Cin·Cout) [1], [2]. Thus, the total complexity becomes O(H·W·Cin·(k^2+Cout)), approximately eight times more efficient for k=3. Denoting d=H·W· Cin and D=Cout, the complexity of one layer is O (d ·k^2+d ·D). Across N images and multiple layers, inference complexity is O (N·∑layers H·W·Cin·(k^2+Cout)). Training doubles this cost due to backpropagation, giving O(2·N·d·k^2). For example, a 224×224×3 (i.e., d≈150,000) with 20 layers requires several hundred million operations, highlighting the efficiency gains of MobileNetV2 over standard CNNs.

➢ *Principal Component Analysis (PCA):*

PCA reduces the dimensionality of extracted features by projecting them onto principal components. For an N×D feature matrix, computing the covariance matrix has complexity O (N D min (N, D), and eigen-decomposition requires O(D^3) [3]. Typically, when N > D, complexity approximates O (N D^2); when D > N, it is O (N^2 D). Projecting a single feature vector of dimension D onto m components costs O (D· m). Thus, PCA fitting requires O (N D^2+D^3), and PCA transformation requires O (N D_m). Memory requirements are modest, storing the D × m basis vectors and a mean vector.

➢ *Support Vector Machine (SVM):*

SVMs classify reduced features from PCA. Linear SVM training using specialized solvers such as LIBLINEAR scales from O (N D) to O(N^2), depending on optimization, whereas non- linear kernels (e.g., RBF) with LIBSVM have worst-case complexities between O(N^2) and O(N^3) [4]. Inference costs O(m) for a linear kernel and O (s ·m) for an RBF kernel, where s is the number of support vectors (s ≤ N).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr960

Memory usage depends on the kernel: linear SVMs store a weight vector of size O(m), while RBF SVMs store s support vectors and coefficients totaling O (s · m).

➢ *Particle Swarm Optimization (PSO):*

PSO optimizes SVM hyperparameters by iteratively updating a swarm of P particles for T iterations. Each iteration evaluates a fitness function here, the SVM accuracy at a cost C_{eval}. Consequently, PSO complexity is $O(P \cdot T \cdot C_{eval})$, which for RBF SVMs where $C_{eval} \sim O(N^2)$ results in total cost $O(PTN^2)$. As PSO is performed offline, it does not affect inference latency. Memory overhead is minimal, primarily storing particle positions and velocities.

V. RESULTS AND DISCUSSION

The proposed hybrid model (CNN + PCA + PSO + SVM) was evaluated on an unseen test dataset to assess its generalization capability and robustness. The model achieved an overall classification accuracy of 96.17%, demonstrating its effectiveness in recognizing Indian Sign Language (ISL) alphabets. This high accuracy indicates that the integration of deep feature extraction with dimensionality reduction and optimization techniques significantly enhances classification performance. The confusion matrix analysis reveals that most classes are correctly classified, as indicated by strong diagonal dominance. Only minor misclassifications were observed, primarily among visually similar hand gestures. For example, the class *'A'* achieved a recall of approximately 91%, indicating a small number of misclassified instances.

The confusion matrix indicates that the proposed model achieves high classification accuracy with minimal misclassification. Errors are primarily observed in visually similar and motion-dependent classes such as J, M–N, and W–Z. The misclassification of J highlights the limitation of static image-based recognition for dynamic gestures. Overall, the model demonstrates strong discriminative capability, with most classes achieving near-perfect accuracy.

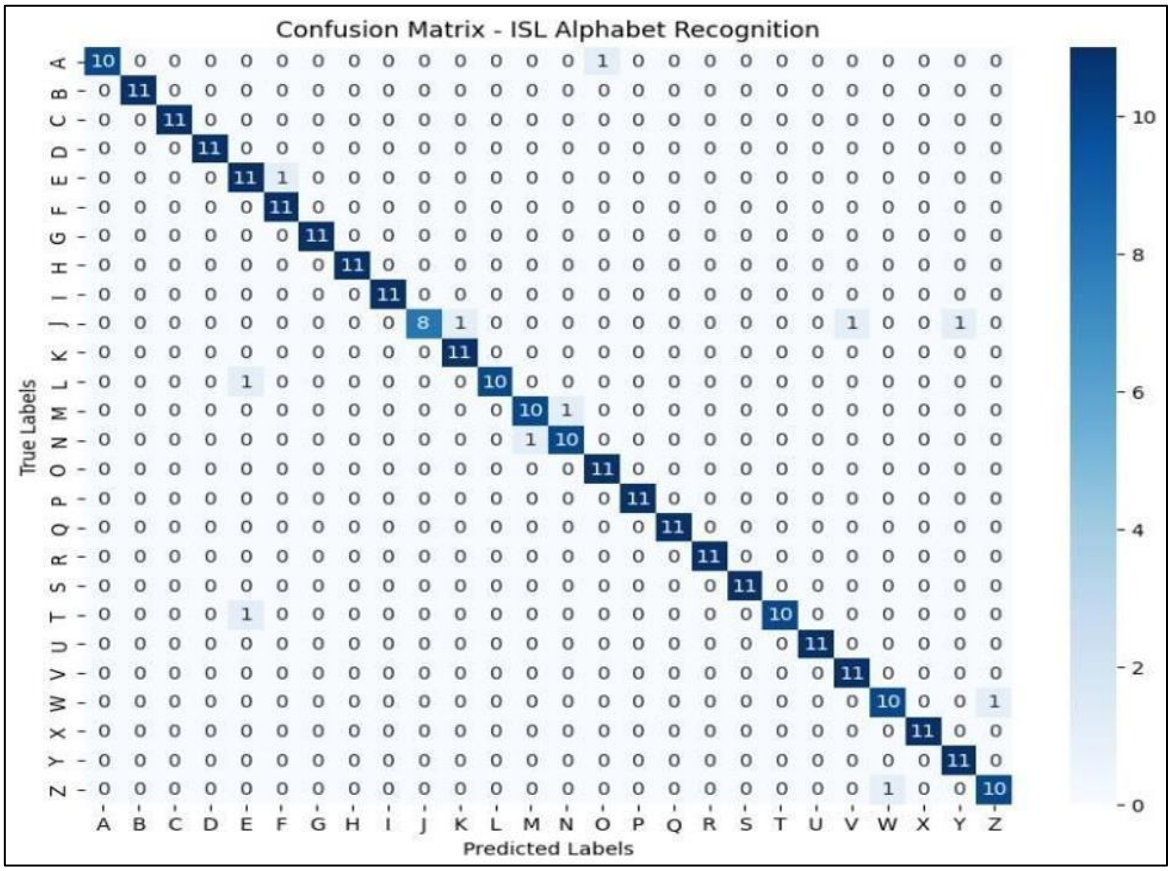

Fig 3 Confusion Matrix

The classification report further confirms the robustness of the model. A majority of the classes achieved perfect scores (Precision = 1.00, Recall = 1.00, F1-score = 1.00), indicating flawless classification. Even for the classes with slight misclassification, the performance metrics remained consistently high (above 0.85), demonstrating the stability and reliability of the proposed approach across all categories.

To evaluate the discriminative capability of the model, a Receiver Operating Characteristic (ROC) curve was plotted using a one-vs-rest strategy. The model achieved an Area Under the Curve (AUC) of 0.96, which indicates excellent class separability. The high AUC value confirms that the model maintains strong performance across different classification thresholds and is well-suited for real-world deployment.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr960

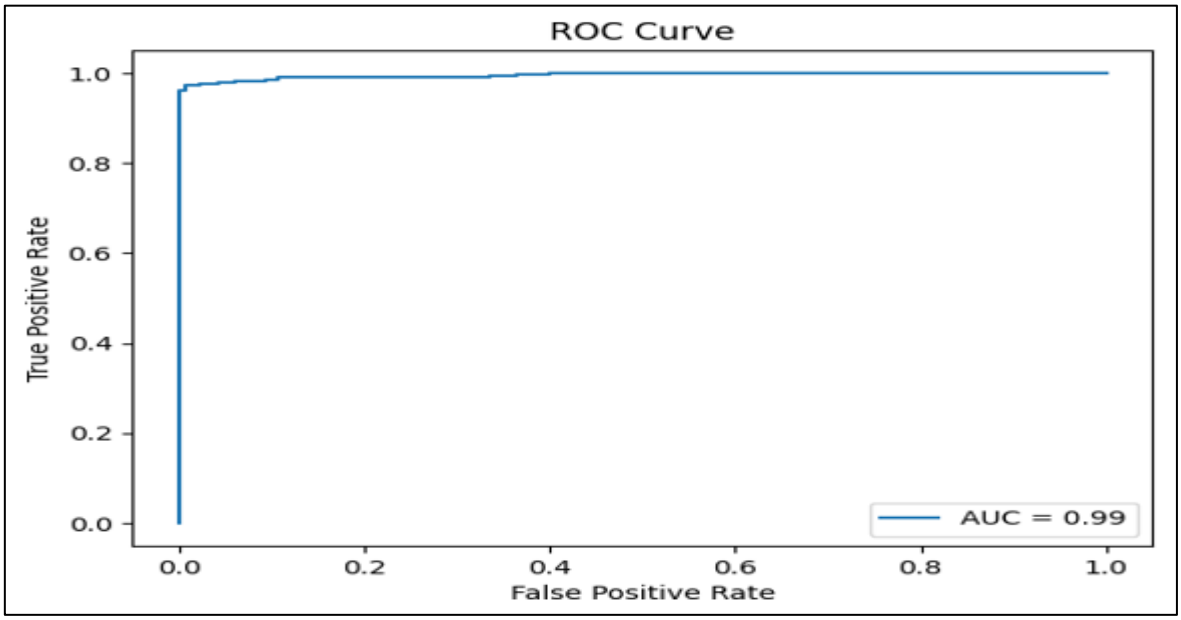

Fig 4 ROC Curve

Fig 5 Predicted Image

Overall, the experimental results validate that the proposed hybrid framework effectively combines the strengths of deep learning and machine learning techniques. The use of PCA reduces feature redundancy, while PSO optimizes SVM parameters, leading to improved classification accuracy and computational efficiency. This makes the model highly suitable for real-time ISL recognition systems.

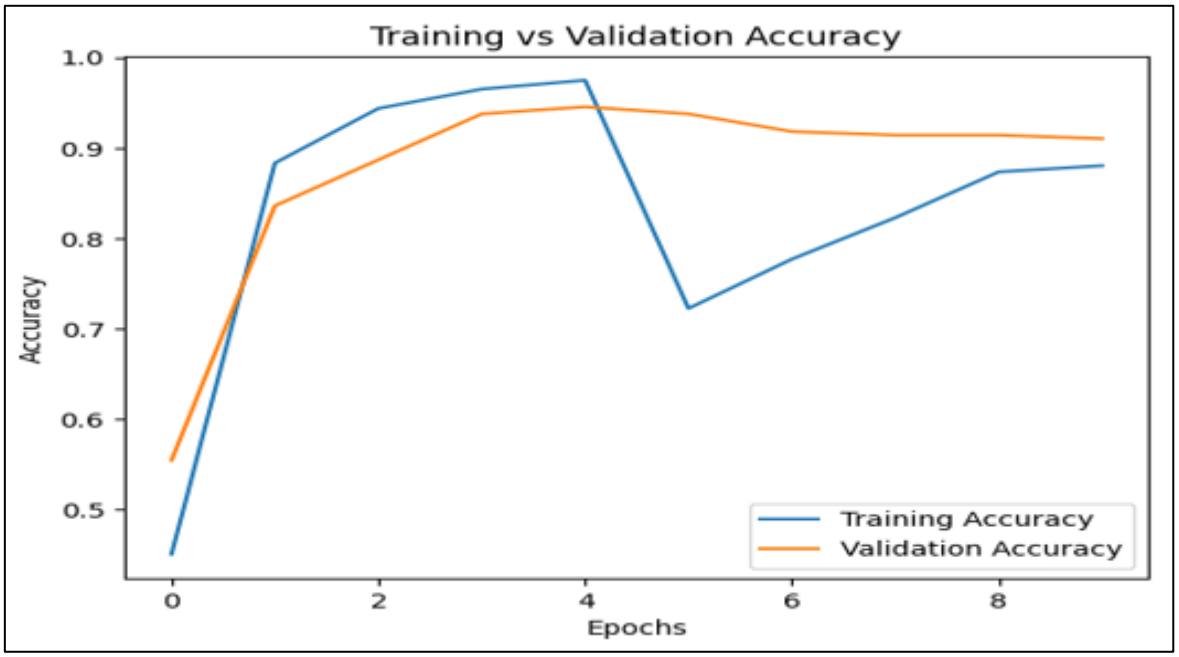

Fig 6 Training v/s Validation Accuracy

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr960

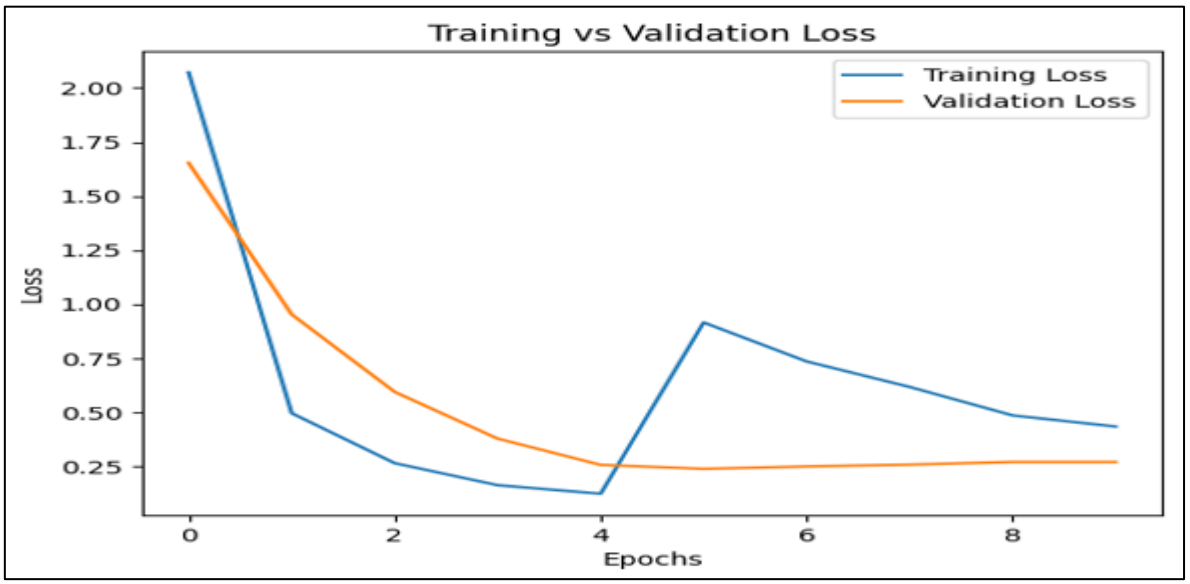

Fig 7 Training v/s Validation Loss

VI. CONCLUSION

This paper presented a novel hybrid approach for Indian Sign Language (ISL) alphabet recognition by integrating CNN-based feature extraction with PCA, PSO, and SVM. The proposed framework effectively combines the representational power of deep learning with the optimization capability of swarm intelligence and the classification strength of SVM. Experimental results demonstrated that the model achieved a high classification accuracy of 96.17%, along with strong precision, recall, and F1-scores across all classes. The confusion matrix and ROC analysis further confirmed the robustness and discriminative capability of the model.

FUTURE WORK

The current work focuses on static ISL alphabet recognition. In future, the framework can be extended to handle dynamic gestures by incorporating sequence-based models such as LSTM or Transformer architectures for video analysis. Additionally, integrating MediaPipe for hand landmark extraction can improve robustness against background variations and lighting conditions. Further, the model can be optimized for deployment on mobile and edge devices using lightweight frameworks like Tensorflow lite, enabling real-time ISL recognition in practical applications.

ACKNOWLEDGMENT

This research is supported by Mahatma Jyotiba Phule Research Fellowship an Autonomous Institute of The Other Backward Class Bahujan Welfare Department, Govt. of Maharashtra.

REFERENCES

[1]. Andrew G. Howard, et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv preprint arXiv:1704.04861, 2017.

[2]. Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in Proc. IEEE CVPR, 2018.

[3]. Ian T. Jolliffe, Principal Component Analysis, 2nd ed., Springer, 2002.

[4]. Chih-Chung Chang and Chih-Jen Lin, “LIBSVM: A Library for Support Vector Machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 1–27, 2011.

[5]. A. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv, 2017.

[6]. L. Pigou et al., "Sign Language Recognition Using Convolutional Neural Networks," ECCV Workshops, 2015.

[7]. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR, 2015.

[8]. M. Kumar et al., "Hybrid CNN-SVM Approach for Gesture Recognition," IEEE Access, 2023.

[9]. J. Kennedy and R. Eberhart, "Particle Swarm Optimization," IEEE ICNN, 1995.