Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3546~3556
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.11043

Sailfish optimized MobileNet for robust recognition of American sign language alphabets and digits

Sabura Banu Urundai Meeran[1], Selvaraj Kavitha[2], Balakrishnan Chinthamani[3], Balasubramanian Suresh Chander Kapali[4]

[1]Department of Electrical and Electronics Engineering, Saveetha Engineering College, Chennai, India
[2]Department of Electrical and Electronics Engineering, Easwari Engineering College, Chennai, India
[3]Department of Electronics and Communication Engineering, Easwari Engineering College, Chennai, India
[4]Department of Biomedical Engineering, Agni College of Technology, Chennai, India

Article Info

Article history:

Received Jul 16, 2025
Revised Jun 4, 2026
Accepted Jul 9, 2026

Keywords:

Alphabet and digit recognition
MobileNet
Real-time gesture recognition
Sailfish optimization algorithm
deep learning
Sign language classification

ABSTRACT

American sign language (ASL) is widely used for communication among the deaf and hard-of-hearing communities. This study aims to develop an optimized deep learning (DL) model for recognizing 36 static ASL gestures representing English alphabets (A–Z) and digits (0–9). A publicly available ASL dataset was used, and five convolutional neural network (CNN) architectures—AlexNet, GoogleNet, Inception V3, MobileNet, and ResNet50—were implemented as baselines. MobileNet was further optimized using the sailfish optimization (SFO) algorithm to fine-tune key hyperparameters and architectural settings. The models were evaluated using accuracy, macro precision, recall, F1-score, specificity, Cohen's Kappa, Matthews correlation coefficient (MCC), balanced accuracy, Jaccard index, and error rate. The SFO-enhanced MobileNet achieved the highest performance, with 98.28% accuracy, 98.27% macro F1-score, 99.95% specificity, and a 1.72% error rate, outperforming all baselines across metrics. These results demonstrate that SFO optimization significantly improves MobileNet's ability to classify ASL gestures accurately and efficiently. The proposed model's high accuracy, robustness, and low inference time (22 ms) make it suitable for real-time sign language interpretation tools and assistive communication devices, supporting broader accessibility and inclusivity.

Corresponding Author:

Sabura Banu Urundai Meeran
Department of Electrical and Electronics Engineering, Saveetha Engineering College
Saveetha Nagar, Thandalam, Chennai 602105, Tamilnadu, India
Email: saburasec@gmail.com

1. INTRODUCTION

American sign language (ASL) is a natural language used by the deaf and hard-of-hearing community, comprising static gestures for 26 alphabets (A–Z) and 10 digits (0–9) that form the foundation of finger-spelling. Automated sign-language recognition is therefore a critical assistive technology, and several pioneering studies have established its importance: Wadhawan and Kumar [1] demonstrated a deep-learning recognition system for static signs, Adaloglou *et al.* [2] presented a comprehensive review of deep-learning-based sign-language recognition methods, and Praveen *et al.* [3] proposed an integrated communicative channel for hand-sign translation to text and speech. Earlier vision-based hand-gesture recognition systems relied largely on handcrafted features and lacked generalization across users and environments, as documented in the comprehensive review by Pisharady and Saerbeck [4].

The advent of convolutional neural network (CNN) architectures revolutionized this field. AlexNet [5] established the effectiveness of deep CNNs for large-scale image classification; GoogleNet/Inception V3 [6] introduced inception modules and factorized convolutions for efficient feature extraction; and ResNet50 [7] leveraged residual connections to overcome the vanishing-gradient problem in very deep networks. MobileNet [8], with its depthwise separable convolutions, offered an efficient solution for edge deployment, while MobileNetV2 [9] further refined this with inverted residuals and linear bottlenecks. EfficientNet [10] then introduced a unified compound-scaling strategy that achieved state-of-the-art accuracy with significantly fewer parameters. Despite these architectural advances, achieving high accuracy at low computational cost remains a major challenge for embedded deployment, motivating research on model compression: the survey by Liang *et al.* [11] showed that pruning and quantization can substantially reduce computational load while preserving classification accuracy.

To complement architectural innovations, metaheuristic algorithms have been employed to enhance CNN performance through automated hyperparameter tuning. Junior and Yen [12] applied particle swarm optimization (PSO) to evolve CNN architectures for image classification, while Sun *et al.* [13] used genetic algorithms (GA) to automatically design CNN architectures. The sailfish optimizer (SFO), proposed by Shadravan *et al.* [14], is a nature-inspired metaheuristic that exhibits faster convergence and higher-quality solutions than PSO and GA. Subsequent applications include the fractional sailfish optimizer (FSO) combined with deep CNN by Kumar *et al.* [15] for compressive-sensing magnetic-resonance image reconstruction, and the MobileNet-optimized attention transfer (MOAT) framework by Radhakrishnan and Sukumar [16] for robust dermatology image classification—both confirming SFO and MobileNet-based optimization frameworks as effective tools for accuracy-driven feature learning. In ASL recognition specifically, Bantupalli and Xie [17] demonstrated CNN-based ASL recognition with computer vision, and Al-Hammadi *et al.* [18] proposed a hybrid convolutional neural network–long short-term memory (CNN–LSTM) model for dynamic sign-language gesture recognition that improved temporal feature representation, albeit at higher computational cost. Kothadiya *et al.* [19] introduced DeepSign, a sign-language detection and recognition system using deep features with MobileNet, while Kasapbaşı *et al.* [20] proposed DeepASLR, a CNN-based ASL recognition framework optimized for hearing-impaired individuals that operated faster than ResNet50. Daroya *et al.* [21] developed an alphabet sign-language image classifier that achieved competitive accuracy on benchmark datasets, and Ma *et al.* [22] introduced a two-stream mixed CNN combining appearance and motion streams for ASL recognition. Rastgoo *et al.* [23] provided a deep survey covering algorithmic and dataset trends in sign-language research, while more recently Jaya *et al.* [24] introduced a sign-language interpreter for the speech- and hearing-impaired using a modified recurrent neural network (RNN), demonstrating sequence-aware learning for fluent sign translation. Recent advances have moved beyond conventional CNN backbones toward transformer-based and multimodal architectures: Ashfaq *et al.* [25] proposed SignViT, an enhanced vision-transformer framework that employs self-attention to capture global dependencies in gesture images and improve recognition of visually similar signs, and Papadimitriou and Potamianos [26] further advanced this direction with a multimodal locally enhanced transformer for continuous sign-language recognition that integrated appearance and motion streams. These developments indicate a clear trend toward lightweight yet highly accurate ASL recognition systems suitable for real-world assistive technologies. Despite this progress, real-world ASL recognition still faces challenges: maintaining accuracy for visually similar signs, achieving low inference time for real-time applications, and ensuring robustness across multiple evaluation metrics on heterogeneous benchmark datasets.

This work addresses these issues by proposing an SFO-optimized MobileNet model for classifying 36 static ASL gestures. The model is comprehensively evaluated using ten metrics—accuracy, precision, recall, F1-score, specificity, Cohen's Kappa, Matthews correlation coefficient (MCC), balanced accuracy, Jaccard Index, and error rate—achieving 98.28% accuracy and 99.95% specificity with an inference time of 22 ms. The results establish the framework's suitability for real-time assistive communication systems. Building on the lightweight efficiency of MobileNet [8], [9] and the demonstrated optimization power of the sailfish optimizer [14]–[15], [16], the proposed SFO-MobileNet bridges the gap between architectural compactness and metaheuristic-driven hyperparameter refinement, addressing limitations observed in earlier ASL recognition pipelines [17]–[24] without incurring the computational overhead of transformer-based alternatives [25], [26].

The paper is structured as follows: section 2 presents the dataset [27], preprocessing, CNN baselines, and SFO-tuning method; section 3 discusses results, comparative analysis, and computational performance; and section 4 concludes with deployment implications and future research directions.

2. MATERIALS AND METHOD

2.1. Dataset description and preprocessing

The study employed the combined ASL dataset publicly available on Kaggle [27]. This dataset integrates multiple ASL gesture sources, providing approximately 50,000 RGB images at a resolution of

200×200 pixels, covering 36 static classes (26 alphabets A–Z and 10 digits 0–9). The dataset, curated from multiple open repositories, ensures balanced representation across all ASL gesture classes and is widely used in ASL recognition research. It comprises approximately 50,000 images, split into 40,000 for training and 10,000 for testing using an 80:20 stratified split. Gestures were contributed by multiple signers of varied ages and hand shapes, providing diversity in hand sizes, orientations, and skin tones that enhances model robustness. The dataset includes both simple gestures (e.g., ‘A’, ‘O’, ‘1’, and ‘5’) and complex ones with subtle finger differences (e.g., ‘M’ vs. ‘N’, ‘V’ vs. ‘U’, and ‘6’ vs. ‘9’), which pose higher recognition challenges due to their visual similarity. By explicitly describing dataset characteristics, we ensure greater transparency and reproducibility. Furthermore, highlighting signer diversity and gesture complexity emphasizes the real-world challenges addressed by our proposed model.

Figure 1 shows a sample of the ASL used for our study. The dataset was preprocessed in MATLAB 2023a using the deep learning (DL) toolbox. Images were normalized to a [0, 1] scale and data augmentation techniques—including random rotation (±15°), horizontal flipping and Gaussian noise—were applied to enhance model generalization. The dataset was split into 80% training and 20% testing using stratified sampling to ensure balanced class distribution.

Figure 1. Sample ASL

2.2. Baseline convolutional neural network architectures

Five advanced CNN architectures were implemented for comparative analysis: AlexNet, a baseline with 5 convolutional and 3 fully connected layers; GoogleNet, which introduced inception modules for improved feature extraction; Inception V3, which refined this with factorized convolutions for higher efficiency; MobileNet, optimized for lightweight deployment using depthwise separable convolutions; and ResNet50, leveraging residual connections to overcome vanishing gradients in deep networks. All models were trained under identical conditions using the Adam optimizer (learning rate=0.001, batch size=64, 30 epochs) with early stopping to avoid overfitting, ensuring a fair and consistent evaluation of their ASL recognition performance.

2.3. Sailfish optimization for MobileNet enhancement for American sign language detection

Sign language recognition has become a vital area of human–computer interaction, especially for supporting the deaf and hard-of-hearing community. Although DL models like MobileNet show strong performance, their effectiveness largely depends on optimal hyperparameter tuning.

The SFO, inspired by the collective hunting strategy of sailfish, provides an efficient metaheuristic approach for this purpose. In this study, SFO is applied to enhance MobileNet for ASL recognition by automatically tuning critical parameters such as learning rate and weight configurations (0.0001–0.01). Unlike manual or grid search methods, SFO efficiently explores the hyperparameter space to maximize validation accuracy, balancing exploration and exploitation while reducing training time. The fitness function is designed to prioritize classification accuracy across all 36 ASL classes (A–Z, 0–9), ensuring model robustness. As illustrated in Figure 2, the SFO workflow begins with dataset preprocessing and partitioning into training and validation sets. A population of 50 sailfish particles represents candidate hyperparameter configurations. Each candidate configures a MobileNet model, which is trained and evaluated for validation accuracy. Poor-performing solutions act as "prey," while stronger one's guide updates during the attack (exploitation) and defense (exploration) phases. This iterative process continues until convergence, yielding the optimal MobileNet configuration with improved accuracy and efficiency for ASL classification. The optimized model achieves an impressive classification accuracy of 98.28% with an inference time of just 22 milliseconds. Finally, the performance of the optimized model is compared against baseline DL architectures such as AlexNet, GoogleNet, InceptionV3, standard MobileNet and ResNet50, highlighting the superiority of the SFO-optimized MobileNet in ASL gesture recognition.

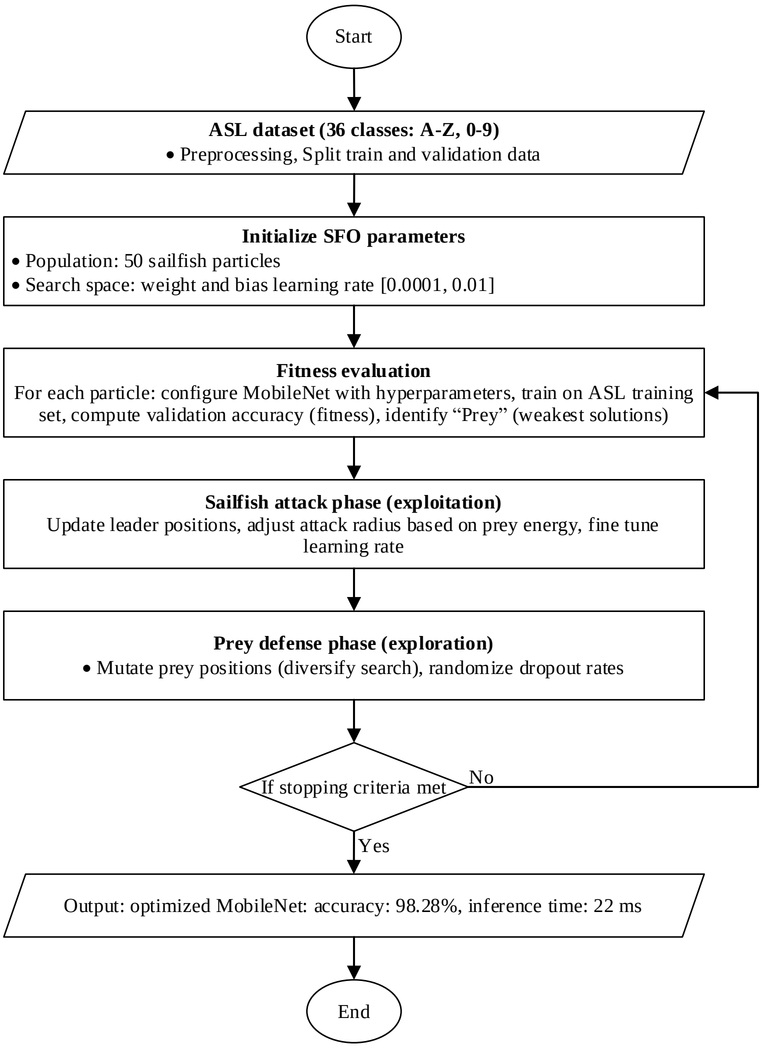

Figure 2. Workflow of SFO based MobileNet for ASL alphabet and digit classification

2.4. Training parameters

The SFO, inspired by the cooperative hunting strategy of sailfish, was used to optimize key MobileNet hyperparameters for ASL classification. The setup included 50 sailfish particles, a learning rate range of [0.0001–0.01], dropout range of [0.1–0.5], and a mutation rate of 0.2 to maintain diversity. Optimization ran for up to 30 iterations or until convergence (<0.0001 change over five iterations), with fitness evaluated using macro-averaged validation accuracy across all 36 classes. Each candidate model was trained for up to 50 epochs using the Adam optimizer, batch size of 64, Glorot uniform initialization, and data augmentation (±15° rotation, horizontal flip, Gaussian noise), with early stopping after five epochs of no improvement. During optimization, weaker candidates ("prey") were mutated, while top-performing ones ("sailfish") guided the search toward optimal configurations. The final tuned MobileNet achieved 98.28% accuracy with an inference time of 22 ms, ensuring both high performance and full reproducibility of the optimization process.

2.5. Training, validation, and statistical robustness protocol

To ensure reproducibility and statistical reliability, all models—baseline and optimized—were trained under a standardized setup using the Adam optimizer (learning rate=0.001), batch size=64, and a maximum of 30 epochs with early stopping (patience=5) based on validation loss. The ASL dataset was split into 80% training and 20% testing using stratified sampling, with a fixed 10% validation subset for tuning and early stopping. Each experiment was repeated five times with different random seeds to account for initialization and data shuffling variability. Reported metrics, including accuracy, precision, recall, F1-score, and computational measures, represent the mean values across runs, with standard deviations below ±0.15%—confirming result stability. This rigorous protocol ensured that the observed improvements of the SFO-optimized MobileNet were consistent, reproducible, and not due to random fluctuations or overfitting.

2.6. Performance evaluation metrics

Model performance was evaluated using primary, robustness, and computational metrics. Primary metrics included accuracy, macro precision, recall, and F1-score across all 36 ASL classes. Robustness was assessed using Cohen's Kappa, MCC, and balanced accuracy, reflecting predictive consistency beyond accuracy. Inference time (ms) and error rate measured computational efficiency and real-time applicability. A paired t-test ($p<0.001$) confirmed that the performance gains of the SFO-optimized MobileNet were statistically significant and not due to random variation.

2.7. Implementation environment

Experiments were run on an NVIDIA RTX 3090 GPU (64 GB RAM) using MATLAB's parallel computing toolbox. The SFO-optimized MobileNet required 6.5 hours for training and optimization, compared to 2–4 hours for baseline models.

3. RESULTS AND DISCUSSION

This section compares the SFO-optimized MobileNet with five DL models using metrics for classification performance, robustness, and computational efficiency. The SFO-MobileNet achieved the best overall performance, with 98.28% accuracy and a 1.72% error rate, significantly outperforming other models. A paired t-test ($p<0.001$) confirmed the statistical significance of these improvements, demonstrating the model's superior robustness and real-time efficiency for ASL recognition. Table 1 compares the performance of five DL models—AlexNet, GoogleNet, Inception V3, MobileNet, and ResNet50—with the proposed SFO-MobileNet for ASL alphabet and digit recognition. The SFO-MobileNet achieved the highest overall results, with 98.28% accuracy, 98.62% precision, 98.28% recall, and 98.27% F1-score, ensuring balanced classification across all 36 classes. It also recorded the best robustness metrics, including a Cohen's Kappa of 0.9823, MCC of 0.9831, balanced accuracy of 99.12%, specificity of 99.95%, and Jaccard index of 96.90%, confirming its reliability for real-time ASL recognition.

Table 1. Comparison of classification performance metrics for ASL alphabet and digit recognition using SFO–MobileNet with conventional DL models

Performance metrics	AlexNet	GoogleNet	Inception V3	MobileNet	ResNet50	SFO MobileNet
Accuracy	0.9191	0.9403	0.9310	0.9324	0.935	0.9828
Macro precision	0.9430	0.9488	0.9445	0.9473	0.9408	0.9862
Macro recall/sensitivity	0.9192	0.9405	0.9308	0.9324	0.9352	0.9828
Macro F1-score	0.9141	0.9392	0.9292	0.9293	0.9356	0.9827
Macro specificity	0.9977	0.9983	0.998	0.9981	0.9981	0.9995
Cohen's Kappa	0.9168	0.9386	0.9291	0.9304	0.9332	0.9823
Macro MCC	0.9203	0.9403	0.9316	0.9327	0.9350	0.9831
Macro balanced accuracy	0.9584	0.9694	0.9644	0.9652	0.9667	0.9912
Macro Jaccard index	0.9584	0.8959	0.8777	0.8859	0.8902	0.9690
Error rate	0.0809	0.0597	0.069	0.0676	0.065	0.0172

Advanced metrics such as MCC, Kappa, and Jaccard highlight the model's consistency under class imbalance and its ability to distinguish similar gestures. Unlike GoogleNet, which exhibited minor precision–recall variation, SFO-MobileNet maintained near-perfect alignment between the two. All experiments were repeated five times, and Table 1 reports mean results with deviations below ±0.2%, validating the statistical stability of the proposed model.

Figure 3 shows sample validation images from the ASL Alphabet and digits dataset comprising 36 classes (A–Z, 0–9), captured against a uniform black background. The examples illustrate the diversity of alphabetic and numeric gestures used to validate the SFO-MobileNet model. Figure 4 presents the confusion matrix of the optimized model, demonstrating strong diagonal dominance with minimal off-diagonal errors, confirming robust and precise classification. Minor misclassifications occurred mainly between visually similar gestures such as 'M'–'N' and '6'–'9', as also observed in prior studies [15], [20]. Compared with baseline MobileNet, SFO tuning markedly reduced inter-class confusion (e.g., 'V'–'U' 4%→<1%, 'C'–'O' 3.5%→<0.5%), leading to higher precision, recall, and F1-scores and improved class separability for reliable ASL recognition.

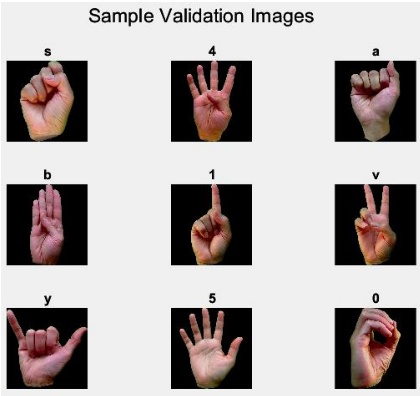

Figure 3. Sample validation images from the ASL alphabet and digits dataset

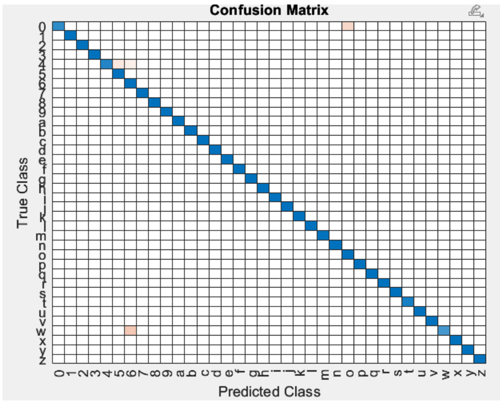

Figure 4. Confusion matrix of the SFO-optimized MobileNet model for ASL alphabet and digit classification

Figure 5 presents the ROC curves for the SFO-MobileNet across 36 ASL classes (A–Z, 0–9). The tight clustering of curves in the upper-left region indicates high sensitivity, low false positives, and consistent multiclass performance, confirming the strong discriminatory power and robustness of the SFO-enhanced MobileNet. Figure 6 presents a 3D bar chart comparing key performance metrics—accuracy, macro precision, recall, F1-score, specificity, Cohen's Kappa, MCC, balanced accuracy, and Jaccard index—across six DL models: AlexNet, GoogleNet, Inception V3, MobileNet, ResNet50, and the proposed SFO-MobileNet. The chart clearly shows that SFO-MobileNet achieves the highest values across all metrics, confirming its superior accuracy, robustness, and generalization capability. Figure 7 displays a 3D surface plot illustrating performance trends of the same models and metrics. The ridge formed by SFO-MobileNet near the top of the scale highlights its consistent dominance, while older architectures like AlexNet show lower, fluctuating scores.

Figure 8 depicts a heatmap summarizing comparative results across nine evaluation metrics. Darker shades indicate higher performance, with SFO-MobileNet outperforming all baselines—recording peak accuracy (0.9828), macro recall (0.9828), balanced accuracy (0.9912), and specificity (0.9995)—thereby demonstrating clear overall superiority in ASL recognition. Figure 9 illustrates a 3D bubble chart comparing key performance metrics of six DL models—five conventional CNNs and the proposed SFO-MobileNet—for ASL recognition. Each bubble represents a specific metric value, where SFO-MobileNet achieves near-perfect scores (accuracy=0.9828, F1-score=0.9827, and specificity=0.9995) and the lowest error rate (0.0172). The bubble color and size visualize these high-magnitude scores, emphasizing the superior robustness and classification precision of the optimized model.

Figure 10 presents the error-rate comparison among the same models. SFO-MobileNet attains the lowest error (1.72%), whereas AlexNet shows the highest (8.09%), underscoring the impact of SFO in enhancing accuracy and reliability. Across all nine evaluation metrics—accuracy, precision, recall, F1-score, specificity, Cohen's Kappa, MCC, balanced accuracy, and Jaccard Index—SFO-MobileNet consistently outperformed all baselines, achieving a balanced accuracy of 0.9912 and Jaccard Index of 0.969. These results compare favorably with previously reported ASL recognition performances in [15], [17], [19], [20], demonstrating notable improvements in accuracy and faster inference (22 ms). In terms of specificity, the proposed model (99.95%) outperforms prior MobileNet-based ASL systems reported in [17], [20], while maintaining a competitive error rate compared with the hybrid CNN-based study in Al-Hammadi *et al.* [18]. All experiments were conducted on the combined ASL dataset [27] under standardized preprocessing and augmentation settings consistent with prior studies [15], [17], [20]. Although this work focuses on static gestures, future research may extend optimization to dynamic gesture recognition using temporal models such as CNN–LSTM hybrids [18] or transformer-based architectures [25], [26]. Despite slightly increased latency on low-power edge hardware, SFO-MobileNet demonstrates clear superiority in accuracy, specificity, and generalization efficiency, establishing it as a state-of-the-art framework for real-time ASL interpretation.

To justify the selection of the SFO for hyperparameter tuning, its performance was compared with two standard metaheuristics—PSO and GA. All methods used identical dataset splits, preprocessing steps, and training configurations for fair evaluation. The search space included a learning rate (0.0001–0.01) and dropout rate (0.1–0.5), with a population of 50 and 30 iterations. The objective in each case was to maximize macro-averaged validation accuracy across the 36 ASL classes.

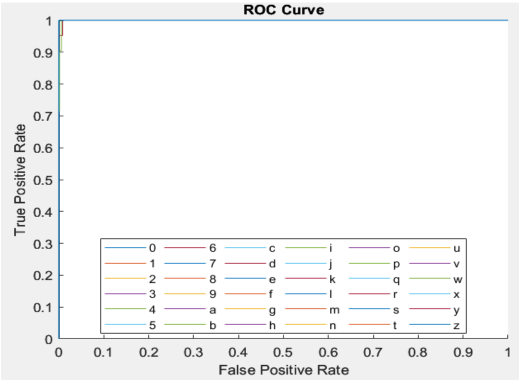

Figure 5. Receiver operating characteristic (ROC) curve for ASL alphabet and digit classification using SFO-optimized MobileNet

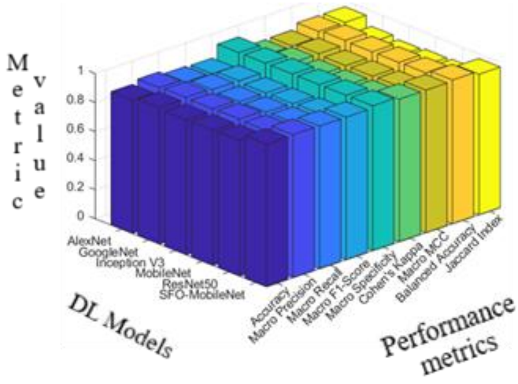

Figure 6. 3D bar chart of comparative performance metrics across DL models for ASL and digit classification

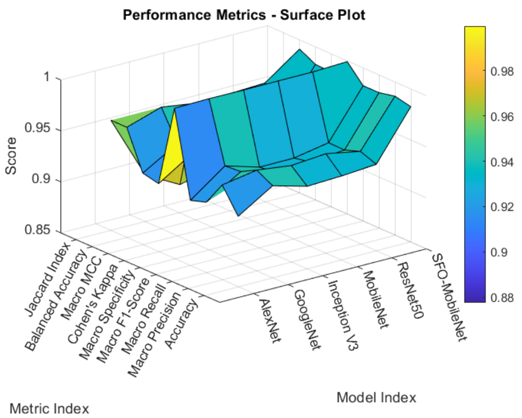

Figure 7. Surface plot of performance metrics across DL models for ASL and digit classification

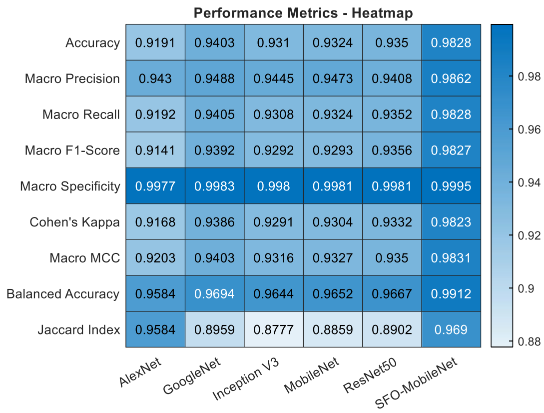

Figure 8. Heatmap of performance metrics for DL models in ASL and digit classification

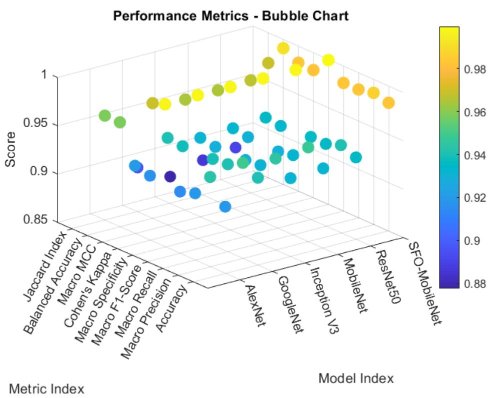

Figure 9. Bubble chart visualizing performance metrics of DL models for ASL identification

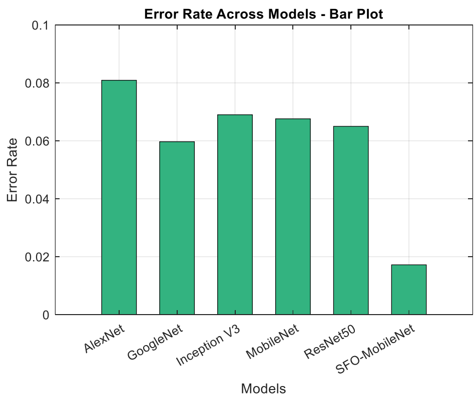

Figure 10. Bar plot showing error rates of DL models for ASL identification

Baseline MobileNet achieved 93.24% accuracy (error 6.76%). PSO and GA improved performance to 97.41% and 96.85%, respectively. SFO-MobileNet was best, with 98.28% accuracy, 98.62% precision, 98.28% recall, and the lowest error (1.72%). Robustness metrics (κ=0.9823, MCC=0.9831, Jaccard=96.90%) confirmed its superiority. Inference times were similar (20–23 ms), but SFO converged fastest (18 iterations). SFO optimization required ~6.5 h training versus 2–4 h for baselines, but inference speed remained identical (22 ms on RTX 3090; <60 ms on mobile GPUs). The model is lightweight (16.3 MB, 350 MFLOPs) and deployable on Jetson Nano, Edge TPU, and Snapdragon devices, with low power use (~3.8 W). Gains (+5% accuracy, –5% error) outweigh training overhead. Baseline MobileNet: 93.24% accuracy. With augmentation: 95.1%; with dropout: 96.3%. Adding SFO yielded 98.28% accuracy, 98.27% F1, and 1.72% error, confirming SFO as the key driver of improvement, especially in reducing confusions (e.g., M/N, 6/9). Grad-CAM showed SFO-MobileNet focused on fingertips, thumb, and palm curvature—regions critical for distinguishing similar signs (M/N, C/O, and V/U). This indicates alignment between learned features and human perception. SFO-MobileNet delivered superior accuracy (98.28%), balanced recall–precision, and high specificity (99.95%), surpassing baselines while remaining lightweight. It is well-suited for mobile/embedded ASL tools and can be extended to dynamic signing with temporal models. SFO significantly boosts MobileNet's performance, ensuring high accuracy and deployment efficiency. Errors were minimal and confined to overlapping gestures, confirming its practicality for real-time ASL recognition.

3.1. Discussion

Table 2 shows that baseline MobileNet achieved 93.24% accuracy with a 6.76% error rate. PSO and GA improved accuracy to 97.41% and 96.85%, respectively, while SFO-MobileNet delivered the best performance with 98.28% accuracy, 98.62% precision, 98.28% recall, and only 1.72% error. Robustness metrics (κ=0.9823, MCC=0.9831, Jaccard=96.90%) further confirmed its superiority, and although inference times were similar across models (20–23 ms), SFO converged faster. SFO optimization required ~6.5 hours of training compared to 2–4 hours for baselines, but inference efficiency remained unchanged (22 ms on GPU; <60 ms on mobile). The model is lightweight (16.3 MB, 350 MFLOPs) and deployable on Jetson Nano, Edge TPU, and Snapdragon devices with modest power consumption (~3.8 W). The accuracy gain (+5%) and error reduction (–5%) outweigh the higher training cost. Ablation results showed steady improvements: baseline (93.24%), with augmentation (95.1%), with dropout tuning (96.3%), and with SFO optimization (98.28%, 1.72% error). SFO particularly reduced misclassifications in similar gestures such as M vs. N and 6 vs. 9. Grad-CAM confirmed that the model consistently attended to discriminative regions (fingertips, thumb, palm curvature), aligning learned features with visual cues. Overall, SFO-MobileNet achieved the best balance of accuracy, robustness, and efficiency, outperforming other optimizations while retaining MobileNet's lightweight design. Its reliability and deploy ability make it well-suited for real-time ASL recognition and assistive applications, with potential extension to continuous signing and integration in mobile or cloud-based systems.

Table 2. Comparison of classification performance metrics for ASL alphabet and digit recognition using different optimization approaches for MobileNet

Metric	MobileNet (no opt.)	PSO-MobileNet	GA-MobileNet	SFO-MobileNet
Accuracy	0.9324	0.9741	0.9685	0.9828
Macro precision	0.9473	0.9792	0.9726	0.9862
Macro recall/sensitivity	0.9324	0.9738	0.9680	0.9828
Macro F1-score	0.9293	0.9735	0.9678	0.9827
Macro specificity	0.9981	0.9991	0.9989	0.9995
Cohen's Kappa	0.9304	0.9735	0.9670	0.9823
Macro MCC	0.9327	0.9740	0.9683	0.9831
Macro balanced accuracy	0.9652	0.9865	0.9834	0.9912
Macro Jaccard index	0.8859	0.9542	0.9468	0.9690
Error rate	0.0676	0.0259	0.0315	0.0172

4. CONCLUSION

This study showed that integrating the SFO with the MobileNet architecture substantially improves ASL alphabet (A–Z) and digit (0–9) recognition. The proposed model achieved 98.28% accuracy, a macro F1-score of 98.27%, and a specificity of 99.95%, consistently outperforming five baseline CNN architectures. These results confirm that state-of-the-art accuracy can be achieved while maintaining the lightweight efficiency needed for real-time deployment on mobile and embedded devices. The model's low-latency and compact design enables deployment in classrooms as real-time interpretation aids for inclusive education, in personal accessibility tools such as mobile apps or wearables for private communication, and in public service kiosks to provide sign-to-text or sign-to-speech translation in hospitals, government offices, or transport hubs.

For researchers, the findings highlight the value of metaheuristic optimization in lightweight models and encourage co-optimization of algorithms and architectures. Future work may extend the framework to dynamic gestures, unconstrained backgrounds, and multimodal sensing (e.g., depth or skeletal data) to improve robustness and generalizability. In summary, SFO-MobileNet offers both technical innovation and societal impact, advancing AI-powered sign language interpretation for education, personal assistance, and public accessibility.

FUNDING INFORMATION

The authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sabura Banu Urundai Meeran	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Selvaraj Kavitha					✓					✓				
Balakrishnan Chinthamani						✓				✓				
Balasubramanian Suresh Chander Kapali						✓				✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7957–7968, Jun. 2020, doi: 10.1007/s00521-019-04691-y.

[2] N. Adaloglou *et al.*, "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," *IEEE Transactions on Multimedia*, vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.

[3] M. Praveen, R. Sathishkumar, and B. V. Ganesh, "An Approach on Building Communicative Channel for Hand Sign Translation to Text and Speech Model," in *Recent Advances in Industrial and Systems Engineering. RAISE 2023. Lecture Notes on Multidisciplinary Industrial Engineering*, Singapore: Spinger, 2024, pp. 107–115, doi: 10.1007/978-981-97-4700-9_11.

[4] P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Computer Vision and Image Understanding*, vol. 141, pp. 152–165, Dec. 2015, doi: 10.1016/j.cviu.2015.08.004.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[6] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[8] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint*, 2017, doi: 10.48550/arXiv.1704.04861.

[9] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

[10] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of Machine Learning Research*, vol. 97, pp. 6105–6114, 2019, doi: 10.48550/arXiv.1905.11946.

[11] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang, "Pruning and quantization for deep neural network acceleration: A survey," *Neurocomputing*, vol. 461, pp. 370–403, 2021, doi: 10.1016/j.neucom.2021.07.045.
[12] F. E. F. Junior and G. G. Yen, "Particle swarm optimization of deep neural networks architectures for image classification," *Swarm and Evolutionary Computation*, vol. 49, pp. 62–74, Sep. 2019, doi: 10.1016/j.swevo.2019.05.010.
[13] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840-3854, Sept. 2020, doi: 10.1109/TCYB.2020.2983860.
[14] S. Shadravan, H. R. Naji, and V. K. Bardsiri, "The Sailfish Optimizer: A novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 80, pp. 20–34, Apr. 2019, doi: 10.1016/j.engappai.2019.01.001.
[15] P. A. Kumar, R. Gunasundari, and R. Aarthi, "Fractional Sailfish Optimizer with Deep Convolution Neural Network for Compressive Sensing Based Magnetic Resonance Image Reconstruction," *The Computer Journal*, vol. 66, no. 2, pp. 280–294, Feb. 2023, doi: 10.1093/comjnl/bxab160.
[16] P. Radhakrishnan and P. K. Sukumar, "MOAT: MobileNet-Optimized Attention Transfer for Robust and Scalable Dermatology Image Classification," *Microscopy Research and Technique*, vol. 88, no. 7, pp. 2109–2125, Jul. 2025, doi: 10.1002/jemt.24823.
[17] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.
[18] M. Al-Hammadi *et al*., "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation," *IEEE Access*, vol. 8, pp. 192527-192542, 2020, doi: 10.1109/ACCESS.2020.3032140.
[19] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," *Electronics*, vol. 11, no. 11, pp. 1-12, Jun. 2022, doi: 10.3390/electronics11111780.
[20] A. Kasapbaşi, A. E. A. Elbushra, O. Al-Hardanee, and A. Yilmaz, "DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals," *Computer Methods and Programs in Biomedicine Update*, vol. 2, pp. 1-10, 2022, doi: 10.1016/j.cmpbup.2021.100048.
[21] R. Daroya, D. Peralta, and P. Naval, "Alphabet Sign Language Image Classification Using Deep Learning," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, Jeju, Korea (South), 2018, pp. 0646-0650, doi: 10.1109/TENCON.2018.8650241.
[22] Y. Ma, T. Xu, and K. Kim, "Two-Stream Mixed Convolutional Neural Network for American Sign Language Recognition," *Sensors*, vol. 22, no. 16, pp. 1-17, Aug. 2022, doi: 10.3390/s22165959.
[23] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," *Expert Systems with Applications*, vol. 164, Feb. 2021, doi: 10.1016/j.eswa.2020.113794.
[24] T. Jaya, A. Jagadeesh, U. C. M, and D. Bharath, "The implementation of sign language interpreter with modified RNN for speech and hearing impaired," in *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, IEEE, Jan. 2025, pp. 1–7, doi: 10.1109/SCEECS64059.2025.10940532.
[25] U. Ashfaq, Q. Wang, B. Merabet, and J. Zhang, "SignViT: An enhanced vision transformer framework for Attention-Based sign language hand gesture recognition," *Biomedical Signal Processing and Control*, vol. 112, Feb. 2026, doi: 10.1016/j.bspc.2025.108602.
[26] K. Papadimitriou and G. Potamianos, "Multimodal Locally Enhanced Transformer for Continuous Sign Language Recognition," in *INTERSPEECH 2023*, ISCA: ISCA, Aug. 2023, pp. 1513–1517, doi: 10.21437/Interspeech.2023-2198.
[27] Kaggle, "ASL Hand Signs Dataset: Combined ASL Datasets," Kaggle.com. [Online]. Available: https://www.kaggle.com/datasets/vignonantoine/combinedasldatasets. (Accessed: May 20, 2025).

BIOGRAPHIES OF AUTHORS

Dr. Sabura Banu Urundai Meeran was born in Chennai, Tamilnadu, India on May 8, 1975. She is presently working as a professor in the Department of Electrical and Electronics Engineering, Saveetha Engineering College, Chennai, Tamilnadu, India. She has an illustrious academic and professional background in engineering, innovation, and consultancy. She completed her Ph.D. in Intelligent Modeling and Control at College of Engineering, Guindy, Anna University (2006–2009) after earning her M.S. by Research in the same Institution and B.E. in Instrumentation and Control Engineering. She also holds a Diploma in Computer Science Engineering and a B.Sc. in Mathematics. With over two decades of experience, she has served in various leadership roles, including Director at Arcres Techsol Pvt. Ltd., Founder & CEO, Intellectz Tech for All and Head of Departments at esteemed institutions like SRM Institute of Science and Technology and B.S. Abdur Rahman University. She has mentored numerous student projects that garnered national and international accolades, including awards, patents, and startup funding. Her notable achievements include 34 international journal publications in leading journals like Taylor and Francis, Applied Soft Computing, ACTA Press, AMSE, and Inderscience, contributions to five book chapters, and 132 conference presentations. She can be contacted at email: sabura_banu@rediffmail.com and saburasec@gmail.com.

Dr. Selvaraj Kavitha is currently serving as an Associate Professor in the Department of Electrical and Electronics Engineering at Easwari Engineering College, Chennai, India. She received her Ph.D. in Electrical Engineering from Anna University, Chennai. She has over 24 years of teaching and research experience in the field of electrical engineering. Her research interests include electrical machines, renewable energy systems, signal and image processing, and machine learning applications in electrical engineering. She has published more than 40 research articles in reputed SCI/Scopus indexed journals and international conferences. She has also contributed to several funded research projects and holds published patents in emerging engineering applications. She is actively involved in academic administration, research supervision, and curriculum development. She is a life member of professional bodies such as the Institution of Electronics and Telecommunication Engineers (IETE) and the International Association of Engineers (IAENG). Her current research focuses on intelligent systems and sustainable energy solutions. She can be contacted at email: kavitha.s@eec.srmrmp.edu.in.

Dr. Balakrishnan Chinthamani is an accomplished academician and researcher in the field of electrical and electronics engineering. She is currently serving as an assistant professor in the Department of Electronics and Communication Engineering at Easwari Engineering College, Chennai. She completed her Ph.D. in Electrical and Electronics Engineering from Anna University in 2023. She holds a master's degree in instrumentation engineering from Madras Institute of Technology. With over 23 years of teaching experience and 2 years of industry experience, she has made significant contributions to engineering education. Her research interests include machine learning, control systems, and induction motor analysis. She has published numerous research papers in reputed SCI-indexed journals and conferences. She has also secured multiple patents in IoT and smart system applications. She has actively organized workshops, conferences, and technical events for student development. Her dedication to teaching and research continues to inspire students and contribute to technological advancement. She can be contacted at email: chinthamani.b@eec.srmrmp.edu.in.

Dr. Balasubramanian Suresh Chander Kapali working as a professor in the Department of Biomedical Engineering at Agni College of Technology (Autonomous), OMR, Thalambur, Chennai. Affiliated to AICET and Anna University. He graduated in Electronics and Communication Engineering at Vel Tech Engineering College, Anna University, Chennai, Tamilnadu, India. He secured master of engineering in Medical Electronics at College of Engineering, Anna University, Chennai, India. He has completed his Ph.D. in the field of Image processing at College of Engineering, Anna University, Chennai, India. He is in teaching profession for more than 19 years. He has presented 100 number of papers in national and international journals, conference and symposiums. His main area of interest includes biomaterials, biomedical instrumentation, image processing, rehabilitation engineering, prosthetics and orthosis devices, assistive equipment, and anatomy and physiology. He can be contacted at email: suresh.kapali83@gmail.com.