JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

COMPARATIVE ANALYSIS OF MOBILENETV2 AND NASNETMOBILE FOR INDONESIAN SIGN LANGUAGE NUMBER CLASSIFICATION USING TRANSFER LEARNING

Tito Sugiharto[1*], Munya Saleh Ba Matraf [2], Ghizlane Khababa[3], Panji Novantara[1]

Universitas Kuningan, Kuningan, Indonesia[1]
www.uniku.ac.id[1]
tito@uniku.ac.id[*], panji@uniku.ac.id

Universiti Utara Malaysia, Kedah, Malaysia[2]
www.uum.edu.my[2]
munya.saleh@uum.edu.my

University of Ferhat Abbas Setif, Setif, Algeria[3]
eng.univ-setif.dz[3]
ghizlane.khababa@univ-setif.dz

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

The creation is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

Submitted: 02 March 2026, **Revised:** 07 November 2026, **Accepted:** 09 November 2026, **Published:** 13 November 2026.

Abstract— *Automatic sign language recognition (SLR) systems powered by computer vision and deep learning serve as crucial assistive communication tools for hearing-impaired communities. Despite these advancements, empirical research on Indonesian Sign Language (BISINDO) digits remains sparse, particularly regarding performance comparisons between lightweight and Neural Architecture Search (NAS)-derived neural networks. To address this gap, this investigation evaluates and compares MobileNetV2 against NASNetMobile for classifying BISINDO numerical gestures. Utilizing a transfer learning strategy, both architectures were trained and evaluated on a 3,000-image dataset distributed evenly across ten digit classes under uniform experimental parameters. Model performance was systematically measured using accuracy, precision, recall, F1-score, confusion matrices, and computational complexity metrics. Experimental results demonstrate that MobileNetV2 surpassed NASNetMobile, attaining a top test accuracy of 91.67% versus 87.33%. Furthermore, MobileNetV2 achieved superior macro-averaged recall and F1-scores while maintaining a lower parameter footprint and compact model size, confirming its efficacy for edge-device deployment. Overall, these findings reveal that lightweight network designs offer an optimal trade-off between predictive accuracy and resource efficiency on moderate dataset scales, providing a solid benchmark for future BISINDO recognition frameworks.*

Keywords: *Convolutional neural networks, indonesian sign language, MobileNetV2, NASNetMobile, transfer learning.*

INTRODUCTION

Automated visual recognition driven by Convolutional Neural Networks (CNNs) has transformed multiple fields, ranging from healthcare to adaptive assistance [1] [2] [3] [4] [5] [6] [7], healthcare [8] [9] [10], education [11] [12], [13], and assistive technologies [14]. By autonomously learning hierarchical feature representations directly from raw inputs [15], CNNs eliminate the reliance on manual feature engineering. Recent lightweight architectures such as MobileNet have been widely adopted for mobile and embedded applications due to their reduced

computational complexity while maintaining competitive accuracy [16] [17]. On the other hand, Neural Architecture Search (NAS) based models such as NASNet have shown improved performance through automated optimization of network structures, achieving state of the art results on multiple benchmark datasets [9] [18]. Sign language recognition (SLR) represents one of the most impactful applications of computer vision in assistive technology [19]. SLR methodology has evolved from handcrafted features in classical machine learning to robust CNN-based deep learning architectures.

Several studies have explored deep learning models for sign language classification [12] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31]. For instance, CNN-based architectures have achieved high accuracy in American Sign Language (ASL) alphabet recognition tasks [23]. Transfer learning approaches utilizing pre-trained networks such as VGGNet and ResNet have also demonstrated promising performance improvements. More recent research has focused on lightweight architectures such as MobileNet to enable real-time recognition on mobile devices, reporting competitive accuracy with reduced computational cost. Additionally, NAS-based architectures such as NASNet have been applied to gesture recognition problems, demonstrating improved generalization performance due to optimized architectural design.

Despite these advancements, most existing studies focus on widely used datasets such as ASL or other international sign language corpora. Research specifically addressing Indonesian Sign Language (BISINDO) remains limited. Furthermore, comparative analyses between lightweight architectures and NAS-optimized architectures in the context of Indonesian sign language classification are still scarce. Many studies emphasize accuracy as the primary evaluation metric without comprehensively analyzing trade-offs between classification performance and computational efficiency.

Therefore, this study systematically compares MobileNetV2 and NASNetMobile for BISINDO digit classification using transfer learning. Evaluating classification performance alongside computational efficiency, this work establishes an empirical benchmark for Indonesian sign language recognition. The primary contributions of this research are summarized as follows:

1. Provides a systematic comparative evaluation between a lightweight deep learning architecture (MobileNet) and a Neural Architecture Search based architecture (NASNet) specifically for Indonesian sign language number classification.
2. Addresses the limited availability of studies focusing on Indonesian Sign Language datasets, particularly in the context of number recognition.
3. Presents a comprehensive evaluation including accuracy, precision, recall, F1-score, and computational characteristics such as model complexity and parameter efficiency.
4. Establishes an empirical benchmark for Indonesian sign language number recognition using two representative deep learning architectures with different design paradigms.
5. The findings offer practical insights into the trade-off between performance and computational efficiency, supporting the development of deployable and inclusive assistive technologies.

Table 1 outlines previous deep learning-based sign language recognition studies to contextualize this work.

Table 1. Literature Study of Sign Language Recognition

Author Name	Problem Researched	Method Used	Research Results
Du et al. (2026) [32]	Mapping HamNoSys notation to 2D body poses, addressing high parameter complexity and poor articulation in conventional MLP models	KANMultiSign framework (a Transformer integrating Kolmogorov–Arnold Networks to replace standard FFNs) utilizing coarse-to-fine multi-scale supervision (25 to 137 keypoints) and iterative refinement.	Reduced model parameters by 55% (from 3.8M to 1.7M) while consistently lowering DTW-MJE error across multiple benchmark datasets (PJM, DGS, GSL, LSF).
Khurshed Aurangzeb, et al. (2024) [33]	High computational costs in processing multi-channel RGB images and limited gesture variety in existing HGRoc systems.	Proposed a lightweight 12-layer HVCNNM model, compared against InceptionV3, VGG16, MobileNet, MobileNetV2, and NASNet Mobile.	The HVCNNM model achieved 99.23% accuracy on MUD and 99.00% on ASLAD datasets with significantly reduced computational time.
Irfan Qutab, et al. (2025) [26]	High computational costs of SLR models	Proposed TMSK framework: Lightweight Tailored	Achieved 99.6% accuracy

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

Author Name	Problem Researched	Method Used	Research Results
Manar Ahmed Hamza, et al. (2025) [34]	Communication gap due to unfamiliarity with SL and the need for real-time, inclusive recognition systems	MobileNet with Self-Attention and KNN classifier. OFTLSLR-BESA model: Image pre-processing (ABF), fusion of InceptionV3, MobileNetV2, NASNetMobile, and ResNet50V2, and RVFL classification	Attained an accuracy of 99.64% on the ASL alphabet dataset, significantly outperforming existing models.
Sakshi Sharma & Sukhwinder Singh (2023) [35]	Lack of efficient, real-time Indian Sign Language (ISL) recognition systems suitable for mobile devices.	Integrated MobileNetV2 with Transfer Learning, using a Global Average Pooling (GAP) layer and Softmax classifier for a compact architecture.	Achieved 99.3% accuracy on ISL and 99.9% on ASL datasets with low computational complexity and a real-time GUI prototype.

Source: (Research Results, 2026)

SLR system research has progressed toward optimizing the trade-off between high classification accuracy and computational efficiency. Du et al. (2026) [32], addressed the challenge of mapping HamNoSys notation to 2D body poses by proposing KANMultiSign, a Transformer framework integrating Kolmogorov–Arnold Networks with multi-scale supervision. Their approach reduced model parameters by 55% while consistently lowering joint errors (DTW-MJE) across multiple benchmark datasets. Aurangzeb et al. (2024) [33], shifted the focus toward computational optimization by proposing a lightweight 12-layer HVCNNM architecture. By utilizing grayscale images instead of multi-channel RGB inputs, the model reduced computational complexity while maintaining high classification performance above 99%. This study highlights the importance of efficiency in real-time SLR applications.

Further advancements toward mobile deployment were introduced by Qutab et al. (2025) [26] and Sharma & Singh (2023) [35]. Qutab proposed the TMSK framework, integrating a tailored MobileNet architecture with a self-attention mechanism and a KNN classifier, achieving high accuracy with a compact 5 MB model size. Similarly, Sharma & Singh demonstrated that MobileNetV2 with transfer learning and Global.

Average Pooling can achieve near-perfect classification results while remaining suitable for mobile-based implementation. More recent studies have explored hybrid fusion strategies. Hamza et al. (2025) [34] employed an ensemble of InceptionV3, MobileNetV2, NASNetMobile, and ResNet50V2 optimized using the Bald Eagle Search (BES) algorithm. Their results demonstrate that metaheuristic-optimized fusion models can further enhance classification accuracy in SLR tasks. Table 1 highlights three key gaps: the dominance of non-Indonesian datasets, the lack of direct MobileNetV2 vs. NASNetMobile comparisons for BISINDO under identical conditions, and the absence of joint accuracy-efficiency evaluations. To address these gaps, this study pursues the following objectives:

1. To implement MobileNet and NASNet architectures for Indonesian sign language number classification using the same experimental settings and dataset.
2. To evaluate and compare the classification performance of both models based on accuracy, precision, recall, and F1-score.
3. To analyze the computational characteristics of MobileNet and NASNet, including model complexity, parameter size, and suitability for deployment in resource-constrained environments.
4. To identify the most suitable architecture for Indonesian sign language number recognition by considering both classification effectiveness and computational efficiency.
5. To provide an empirical benchmark that can serve as a reference for future research in Indonesian sign language recognition systems.

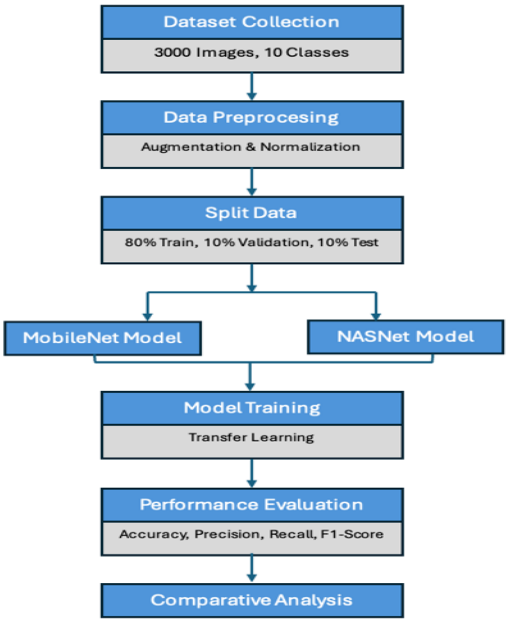
This study provides a novel controlled evaluation between standalone MobileNetV2 and NASNetMobile models for BISINDO numerical recognition, moving beyond previous studies focused on model fusion or non-Indonesian languages. Utilizing a dataset grounded strictly in BISINDO the natural sign language used by the Indonesian Deaf community, distinct from the rule-based SIBI system this work integrates classification accuracy with computational efficiency to guide edge device deployment.

MATERIALS AND METHODS

This experiment benchmarks MobileNet against NASNet for Indonesian Sign Language number classification under standardized settings to guarantee fairness and reproducibility. The end-to-end process from dataset preparation to final analysis is illustrated in Figure 1. Figure 1 illustrates the complete workflow: data collection, preprocessing, and splitting, followed by transfer

CC BY NC

learning implementation of MobileNet and NASNet. After uniform training and standardized evaluation, a comparative analysis determined the best model regarding accuracy and resource efficiency.

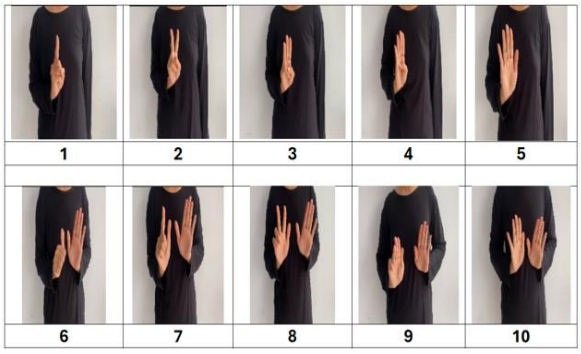


Source: (Research Results, 2026)
Figure 1. Research Design

Dataset Collection

A balanced dataset of 3,000 Indonesian Sign Language digit images across 10 classes (300 images per class) was gathered to train and evaluate the models without introducing class bias. Data collection involved a participant performing gestures with both left and right hands across different hand dimensions to enhance variability and model generalization. High-resolution images (1920×1080 pixels) were captured using a smartphone camera positioned 30–50 cm from the subject against a solid white background. To reduce background noise and satisfy the input specifications of the neural networks, all images were resized to 224×224 pixels during preprocessing.

Each captured image was manually labeled according to its corresponding numerical gesture class to ensure accurate ground-truth annotation. Data labeling was conducted by two annotators knowledgeable in BISINDO numerical gestures. To guarantee annotation consistency, all labels were cross-verified, with any discrepancies resolved through consensus. Furthermore, to prevent data leakage, a subject-independent data split was strictly enforced by assigning participants exclusively to either the training set or the evaluation subsets (validation and testing), ensuring the model was assessed on unseen subjects. The labeled dataset was then organized and verified before proceeding to the preprocessing stage. This systematic data collection approach ensures dataset reliability and supports fair performance comparison between the MobileNetV2 and NASNetMobile architectures.



Source: (Research Results, 2026)
Figure 2. Sample Image

Figure 2 showcases sample gestures for digits 1–10. Controlled lighting and consistent backgrounds ensure clear hand shapes and finger placements, preserving class-specific traits necessary for deep learning feature extraction.

Data Preprocessing

The data preprocessing stage was conducted to standardize input images and enhance dataset variability prior to model training. In this study, preprocessing and augmentation were implemented using the ImageDataGenerator class to systematically apply normalization and transformation operations to the training dataset. To stabilize gradient optimization and speed up network convergence, image intensity values were scaled from the standard [0,255] range down to [0,1].

To improve generalization capability and reduce overfitting, several data augmentation techniques were applied. Random rotations were introduced within a range of 20 degrees (rotation_range=20) to simulate slight variations in hand orientation. Horizontal and vertical translations were applied using width and height shift ranges of 20% (width_shift_range=0.2, height_shift_range=0.2) to mimic natural positional changes during gesture performance. Shear transformations (shear_range=0.2) were incorporated to simulate minor geometric distortions, while zoom operations (zoom_range=0.2) allowed the model to learn scale-invariant features by randomly enlarging or reducing the image within a 20% range.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

Additionally, horizontal flipping (horizontal_flip=True) was applied to increase variation in gesture representation. Horizontal mirroring was adopted to reflect handedness variability among signers, whereas vertical flipping was omitted to avoid feeding anatomically invalid hand postures into the training pipeline. This tailored data augmentation strategy enables the networks to learn robust, invariant visual features essential for real-world gesture recognition.

Importantly, augmentation operations were applied only to the training dataset to prevent data leakage and ensure unbiased validation and testing performance. The validation and testing sets were processed only with rescaling normalization to maintain fair model evaluation. Through this preprocessing configuration, the dataset was effectively standardized and diversified, ensuring optimal input quality for the subsequent training of MobileNet and NASNet architectures.

Split Data

After the preprocessing stage, the dataset was divided into training, validation, and testing subsets to ensure reliable model development and unbiased performance evaluation. A stratified data splitting strategy was applied to maintain proportional representation of all ten sign language number classes across each subset. This approach prevents class imbalance and ensures that each model is trained and evaluated under equivalent distribution conditions. The 3,000 images were split 80/10/10 into training (2,400), validation (300), and test (300) subsets. Training images optimized model parameters, validation data guided hyperparameter tuning and early stopping, and the held-out test set evaluated final generalization on unseen data. The separation between training, validation, and testing datasets was strictly maintained to avoid data leakage. No image from the validation or testing subsets was used during the training process. Furthermore, data augmentation techniques were applied only to the training set, while validation and testing images were processed solely with normalization.

Model Architecture Implementation (MobileNet and NASNet Configuration)

In this study, two pre-trained convolutional neural network architectures, MobileNetV2 and NASNetMobile, were implemented using a transfer learning strategy for Indonesian Sign Language number classification. Both models were configured within a sequential architecture framework, where a pre-trained base model was followed by additional custom classification layers adapted to the 10-class output.

MobileNetV2 Architecture Configuration

The MobileNet-based model utilizes MobileNetV2 (mobilenetv2_1.00_224) as the feature extractor. The base model produces an output feature map of shape (7, 7, 1280), indicating 1,280 high-level feature channels extracted from the input image.

Table 2. MobileNetV2 Architecture

Layer (type)	Output Shape	Param #
Mobilenetv2 (Functional)	(None, 7, 7, 1280)	2,257,984
Conv2d (Conv2D)	(None, 7, 7, 32)	368,672
Max_pooling2d (MaxPooling2D)	(None, 3, 3, 32)	0
Dropout (Dropout)	(None, 3, 3, 32)	0
Flatten (Flatten)	(None, 288)	0
Dense (Dense)	(None, 10)	2,890
Total params		2,629,546 (10.03 MB)
Trainable params		371,562 (1.42 MB)
Non-trainable params		2,257,984 (8.61 MB)

Source: (Research Results, 2026)

Table 2 presents the architectural summary of the MobileNetV2-based classification model implemented in this study. The model follows a sequential structure where a pre-trained MobileNetV2 backbone is utilized as the primary feature extractor and additional custom layers are appended for task-specific classification. A pre-trained MobileNetV2 backbone (mobilenetv2_1.00_224) serves as the core feature extractor, outputting a 7×7×1280 feature tensor. To preserve transfer learning features and lower computation, all 2,257,984 base parameters were kept frozen throughout initial training. Freezing these layers allows the model to retain general feature representations learned from large-scale datasets while reducing computational overhead and preventing overfitting.

Following the backbone, a Conv2D layer is added to reduce the feature depth from 1,280 channels to 32 channels, resulting in an output shape of (7, 7, 32). This layer introduces 368,672 trainable parameters, enabling task-specific feature refinement. Next, a MaxPooling2D layer reduces the spatial dimensions from (7, 7) to (3, 3), effectively decreasing computational complexity and focusing on dominant spatial features. A Dropout layer is then applied to mitigate overfitting by randomly deactivating neurons during training, thereby enhancing generalization capability. The Flatten layer transforms the 3D feature map into a one-dimensional vector of size 288, preparing it for the

CC BY NC
Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

final classification stage. The final Dense layer contains 10 output neurons corresponding to the 10 Indonesian sign language number classes. This layer includes 2,890 parameters and applies the Softmax activation function to produce class probability distributions. Overall, the model contains 2,629,546 total parameters (10.03 MB), of which 371,562 are trainable and 2,257,984 are non-trainable. The relatively small number of trainable parameters demonstrates the efficiency of the MobileNetV2 architecture. Its lightweight design makes it suitable for deployment in mobile or resource-constrained environments while maintaining strong feature extraction capability.

NASNetMobile Architecture Configuration

The second model employs NASNetMobile as the backbone feature extractor. The base NASNetMobile architecture produces an output shape of (7, 7, 1056), reflecting a deep feature representation optimized through Neural Architecture Search (NAS). Table 3 presents the architectural summary of the NASNetMobile based classification model implemented in this study. The NASNetMobile-based model utilizes a pre-trained NASNetMobile backbone as the primary feature extractor, producing an output feature map of (7, 7, 1056). The backbone contains 4,269,716 non-trainable parameters, which were frozen during training to retain previously learned feature representations and reduce overfitting.

Table 3. NASNetMobile Architecture

Layer (type)	Output Shape	Param #
Nasnet_mobile (Functional)	(None, 7, 7, 1056)	4,269,716
Conv2d_1 (Conv2D)	(None, 7, 7, 32)	304,160
Max_pooling2d_1 (MaxPooling2D)	(None, 3, 3, 32)	0
Dropout_1 (Dropout)	(None, 3, 3, 32)	0
Flatten_1 (Flatten)	(None, 288)	0
Dense_1 (Dense)	(None, 10)	2,890
Total params		4,576,766 (17.46 MB)
Trainable params		307,050 (1.17 MB)
Non-trainable params		4,269,716 (16.29 MB)

Source: (Research Results, 2026)

To adapt the architecture for the 10-class Indonesian Sign Language number classification task, additional layers were appended, including a Conv2D layer that reduces feature channels to 32, followed by MaxPooling2D and Dropout layers to decrease spatial dimensions and improve generalization. A Flatten layer converts the feature maps into a 288-dimensional vector, which is then processed by a final Dense layer with 10 output neurons using Softmax activation. The model contains a total of 4,576,766 parameters (17.46 MB), with 307,050 trainable parameters and 4,269,716 non-trainable parameters. Compared to lightweight architectures, NASNetMobile exhibits higher model complexity and memory size, reflecting its deeper architecture optimized through Neural Architecture Search.

Model Training (Transfer Learning Approach)

The model training stage was conducted using a transfer learning strategy, Figure 1. In this approach, both MobileNetV2 and NASNetMobile were initialized with pre-trained weights obtained from large-scale image datasets. The convolutional backbone layers of each model were frozen during the initial training phase to preserve generic visual feature representations, while only the newly added classification layers were trained on the Indonesian Sign Language number dataset. During training, the models processed the augmented training dataset (2,400 images) in batches. Fine-tuning was selectively applied by unfreezing higher-level layers after the initial training phase to allow better adaptation of abstract feature representations to the specific sign language dataset. This two-stage training strategy enables the models to retain robust low-level feature extraction while refining high-level features relevant to numerical hand gestures.

The 300 image validation set tracked training progress and triggered early stopping at peak performance to prevent overfitting. A held-out test set of 300 images then provided the final assessment of model generalization. By applying identical transfer learning procedures and training configurations to both MobileNetV2 and NASNetMobile, the study ensures a fair and controlled comparison. Any differences observed in classification accuracy or efficiency can therefore be attributed primarily to architectural characteristics rather than variations in training methodology.

Performance Evaluation Metrics

The performance of both MobileNetV2 and NASNetMobile models was evaluated using standard multi-class classification metrics to ensure comprehensive assessment. The evaluation was conducted on the testing dataset consisting of 300 unseen images. The primary metrics used in this study include accuracy, precision, recall, and F1-score, as these metrics provide a balanced evaluation of classification effectiveness. Model capability was evaluated across ten numerical classes using macro-averaged Accuracy, Precision,

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

Recall, and F1-score derived from standard confusion matrix parameters (TP, FP, TN, and FN). It is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Since this study involves multi-class classification with 10 number categories, the metrics were computed for each class and then averaged using macro-averaging to provide an overall performance measure. Macro-averaging treats all classes equally, ensuring that performance is not biased toward dominant classes. In addition to these metrics, confusion matrix analysis was conducted to visualize classification distribution across the ten classes. The confusion matrix provides detailed insights into misclassification patterns, enabling identification of classes that are more prone to prediction errors.

For enhanced clarity, performance metrics were computed directly from the confusion matrix. Accuracy represents the ratio of correctly classified instances to all samples, whereas precision, recall, and F1-scores were macro-averaged across classes using their respective True Positive (TP), False Positive (FP), and False Negative (FN) counts. Through the use of these evaluation metrics, the study ensures a rigorous and transparent comparison between MobileNetV2 and NASNetMobile architectures in terms of both classification accuracy and predictive reliability.

Comparative Architectural Perspective

The comparative analysis between MobileNetV2 and NASNetMobile was conducted to evaluate differences in classification performance and computational efficiency under identical experimental conditions. From an architectural perspective, NASNetMobile has a larger total parameter count (4,576,766 parameters; 17.46 MB) compared to MobileNetV2 (2,629,546 parameters; 10.03 MB). This indicates that NASNetMobile possesses higher structural complexity, resulting from its Neural Architecture Search-based design. However, despite the larger total parameter size, the number of trainable parameters in NASNetMobile (307,050) is slightly lower than that of MobileNetV2 (371,562) due to the freezing strategy applied to the backbone layers. This suggests that most computational burden in NASNetMobile lies in its fixed pre-trained layers rather than in the fine-tuned classification layers.

The distinct output dimensions of MobileNetV2 (7, 7, 1280) and NASNetMobile (7, 7, 1056) are intrinsic to their respective backbone designs. Consequently, these feature maps cannot be directly aligned without altering the underlying pre-trained architectures. In terms of representational capacity, NASNetMobile provides a more complex feature extraction mechanism, potentially capturing richer spatial patterns in hand gestures. Conversely, MobileNetV2 emphasizes computational efficiency through depthwise separable convolutions, which significantly reduce parameter count and model size while maintaining competitive performance.

From a deployment perspective, MobileNetV2 demonstrates advantages in memory efficiency and suitability for mobile or edge-device implementation due to its smaller model size. NASNetMobile, while offering potentially enhanced feature learning capability, requires higher memory consumption and computational resources. Therefore, the selection between the two architectures depends on application requirements whether priority is given to lightweight implementation or potentially improved feature representation.

Computational Efficiency Metrics

To provide a comprehensive efficiency analysis, inference speed and frames per second (FPS) were determined as follows:

$$\text{Inference Time} = \frac{Total\ Time}{Number\ test\ Image} \quad (5)$$

$$\text{FPS} = \frac{Number\ test\ image}{Total\ Prediction\ Time} \quad (6)$$

These metrics provide a more comprehensive evaluation of deployment suitability for real-time and resource-constrained environments. To empirically evaluate real-time suitability for resource-constrained environments, practical efficiency metrics were recorded on a Google Colab T4 GPU. Per-sample inference time was logged during prediction, while throughput

CC BY NC
Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

(FPS) was evaluated across 300 test images. Epoch training time was averaged over epochs 2–5 to exclude GPU initialization latency. Table 4 presents the complete computational comparison between MobileNetV2 and NASNetMobile.

Table 4. Computational Efficiency Comparison: MobileNetV2 vs. NASNetMobile

Computational Metric	MobileNetV2	NASNetMobile
Total parameters	2,629,546	4,576,766
Trainable parameters	371,562	307,050
Non-trainable parameters	2,257,984	4,269,716
Model size (MB)	10.03 MB	17.46 MB
Training time/epoch – epoch 1 (warmup)	1,136 s	1,091 s
Training time/epoch – steady state (epoch 2–5 avg)	~194 s	~269 s
Total training time (5 epochs)	1,913 s (~32 min)	2,167 s (~36 min)
Inference Speed (300 test images, Google Colab GPU)		
Inference time – evaluation with loss (total, 300 images)	73 s	78 s

Source: (Research Results, 2026)

RESULTS AND DISCUSSION

Results

This study reveals variations in the performance of the two deep learning models tested. Below are the detailed results of the classification evaluation, highlighting the performance of MobileNetV2 and NASNetMobile in recognizing Indonesian Sign Language number gestures.

Accuracy Testing

The overall classification performance, including accuracy, precision, recall, and F1-score, is summarized in Table 5, with graphical comparisons provided in Figure 3 to facilitate visual analysis between the two architectures.

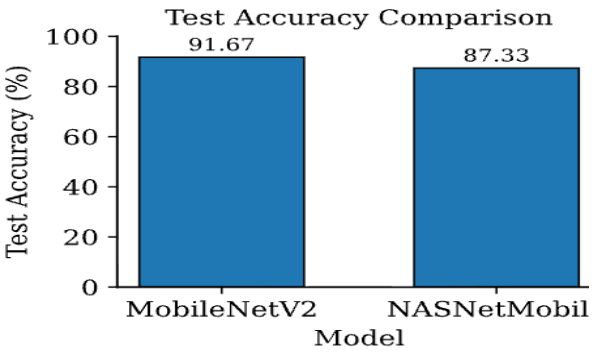
Table 5. Summary of Accuracy Metrics Results

Evaluation Metric	MobileNetV2	NASNetMobile
Test Accuracy (%)	91.67	87.33
Macro Precision	0.94	0.94
Macro Recall	0.92	0.88
Macro F1-Score	0.91	0.86
Weighted Precision	0.94	0.93
Weighted Recall	0.92	0.87
Weighted F1-Score	0.90	0.84
Total Parameters	2,629,546	4,576,766
Model Size (MB)	10.03 MB	17.46 MB

Source: (Research Results, 2026)

Table 5 presents a comprehensive summary of the performance metrics obtained by MobileNetV2 and NASNetMobile on the Indonesian Sign Language number classification task. Based on the testing results, MobileNetV2 achieved a higher test accuracy of 91.67%, outperforming NASNetMobile, which obtained 87.33%. This indicates that MobileNetV2 provides better predictive performance on unseen test data. Although both models achieved identical macro precision values (0.94), MobileNetV2 demonstrated superior macro recall (0.92) compared to NASNetMobile (0.88). This suggests that MobileNetV2 is more effective in correctly identifying instances across all numerical gesture classes. Similarly, the macro F1-score of MobileNetV2 (0.91) exceeded that of NASNetMobile (0.86), reflecting a more balanced trade-off between precision and recall. The weighted evaluation metrics further confirm this trend, with MobileNetV2 achieving a weighted F1-score of 0.90, compared to 0.84 for NASNetMobile, indicating more consistent classification across the dataset.

From a computational standpoint, NASNetMobile contains a significantly larger number of total parameters (4,576,766) and a greater model size (17.46 MB) compared to MobileNetV2 (2,629,546 parameters and 10.03 MB). Despite its higher architectural complexity, NASNetMobile did not surpass the performance of the more lightweight MobileNetV2 model. This finding highlights that increased model complexity does not necessarily guarantee improved classification accuracy for the given dataset. Overall, the results demonstrate that MobileNetV2 achieves a better balance between classification performance and computational efficiency, making it a more suitable architecture for practical deployment in resource-constrained or mobile-based sign language recognition systems.



Source: (Research Results, 2026)

Figure 3. Visualization of Summary of Accuracy Metric Results

Epoch Value Testing

The accuracy per epoch test results are presented in Table 6. Table 6 compares the validation performance of MobileNetV2 and

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

NASNetMobile across five training epochs. MobileNetV2 demonstrates consistently lower validation loss and competitive validation accuracy throughout the training process. By the final epoch, MobileNetV2 achieves the highest validation accuracy (0.8900) with the lowest validation loss (0.2689), indicating stronger convergence and better generalization performance compared to NASNetMobile, which records 0.8633 validation accuracy and 0.4332 validation loss.

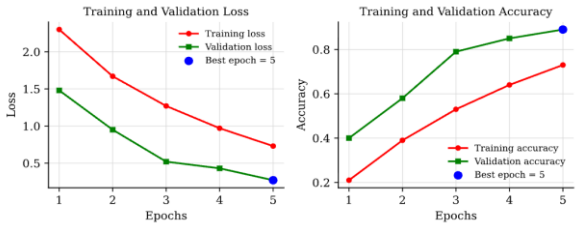
Table 6. Comparative Validation Performance per Epoch

Epoch	Values Testing MobileNet (Val Acc / Val Loss)	Values Testing NASNetMobile (Val Acc / Val Loss)
1	0.4033 / 1.4810	0.4033 / 1.4810
2	0.5733 / 0.9486	0.6033 / 1.3266
3	0.7900 / 0.5240	0.7767 / 0.8624
4	0.8467 / 0.4297	0.8533 / 0.5362
5	0.8900 / 0.2689	0.8633 / 0.4332

Source: (Research Results, 2026)

The comparison highlights that MobileNetV2 not only converges faster but also maintains more stable validation performance across epochs. These findings align with the final testing results, reinforcing the conclusion that MobileNetV2 provides superior performance for the Indonesian Sign Language number classification task. Figure 4 presents the learning curves of the MobileNetV2 model across five training epochs, illustrating both loss and accuracy trends for training and validation datasets. From the Training and Validation Loss graph, both curves show a consistent downward trend throughout the training process. The training loss decreases steadily from approximately 2.83 in the first epoch to around 0.78 in the fifth epoch. Similarly, the validation loss drops significantly from 1.48 to 0.27, indicating improved model generalization over time. The lowest validation loss is achieved at epoch 5, which is marked as the best epoch. The parallel decline of both training and validation loss curves suggests stable convergence and minimal overfitting.

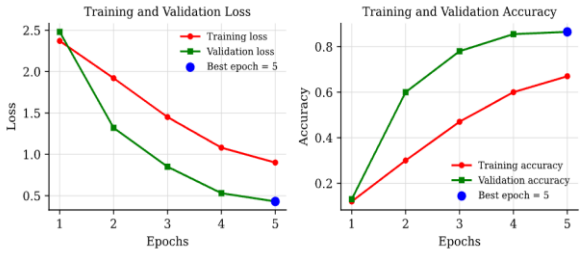
The Training and Validation Accuracy graph further demonstrates progressive learning behavior. Training accuracy increases gradually from approximately 20% in epoch 1 to about 72% in epoch 5. Meanwhile, validation accuracy improves more rapidly, rising from approximately 40% to nearly 89% by the final epoch. The validation accuracy consistently exceeds training accuracy, which can be attributed to the application of regularization techniques such as dropout and data augmentation during training. These techniques may reduce apparent training performance while enhancing generalization capability on unseen validation data.



Source: (Research Results, 2026)
Figure 4. The Learning Curves of the MobileNetV2

Figure 5 illustrates the learning curves of the NASNetMobile model across five training epochs, including both loss and accuracy trends for training and validation datasets. From the Training and Validation Loss graph, it can be observed that both curves exhibit a consistent downward trajectory throughout the training process. The training loss decreases progressively from approximately 2.71 in the first epoch to around 0.95 in the final epoch. Similarly, the validation loss shows a more pronounced decline, dropping from approximately 2.49 to 0.43 by epoch 5.

The continuous reduction in validation loss indicates that the model improves its generalization capability over time. The lowest validation loss occurs at epoch 5, which is marked as the best epoch, suggesting optimal model performance at this stage. The absence of a widening gap between training and validation loss curves suggests that overfitting is minimal. The Training and Validation Accuracy graph further supports this observation. Training accuracy increases steadily from approximately 12% in the first epoch to about 65% in the final epoch, reflecting progressive learning. Validation accuracy demonstrates a sharper improvement, rising significantly from approximately 11% to nearly 87% by epoch 5.



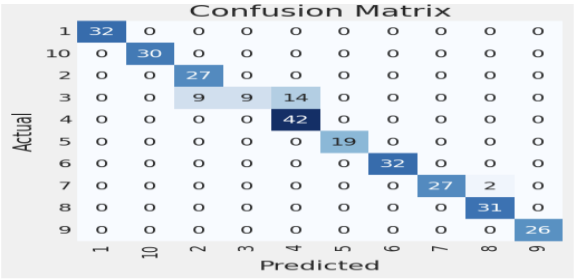
Source: (Research Results, 2026)
Figure 5. The Learning Curves of the NASNetMobile

Confusion Matrix

The confusion matrix analysis was conducted to further examine the classification performance of

CC BY NC

both MobileNetV2 and NASNetMobile at the class level. Unlike overall accuracy metrics, the confusion matrix provides a detailed representation of how each numerical gesture class is predicted, allowing identification of correct classifications and misclassification patterns across the ten categories. Figure 6 presents the confusion matrix of the MobileNetV2 model for Indonesian Sign Language number classification. The matrix shows strong diagonal dominance, indicating that the majority of samples were correctly classified into their respective numerical categories. Several classes achieved perfect classification performance. Specifically, classes 1, 5, 6, 9, and 10 show no misclassification, with all samples correctly predicted. This is consistent with the classification report in Table 7, where these classes achieved precision, recall, and F1-score values of 1.00. Such results indicate that the model effectively captured discriminative features for these gestures.



Source: (Research Results, 2026)

Figure 6. Confusion Matrix MobileNetV2

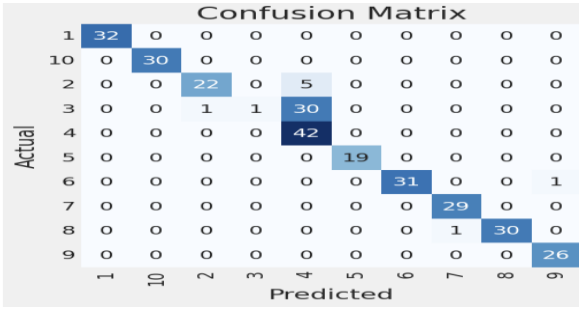
Table 7. Classification Report on the Test Set MobileNetV2

Class	Precision	Recall	F1-score	Support
1	1.00	1.00	1.00	32
10	1.00	1.00	1.00	30
2	0.75	1.00	0.86	27
3	1.00	0.28	0.44	32
4	0.75	1.00	0.86	42
5	1.00	1.00	1.00	19
6	1.00	1.00	1.00	32
7	1.00	0.93	0.96	29
8	0.94	1.00	0.97	31
9	1.00	1.00	1.00	26
Accuracy			0.92	300
Macro avg	0.94	0.92	0.91	300
Weighted avg	0.94	0.92	0.90	300

Source: (Research Results, 2026)

However, minor misclassifications are observed in a few classes. The most notable case occurs in class 3, where only 9 out of 32 samples were correctly classified, while a significant portion was misclassified as class 4. This explains the lower recall (0.28) and F1-score (0.44) for class 3. The confusion between classes 3 and 4 suggests visual similarity in finger positioning, which may reduce feature separability. Additionally, small misclassifications are observed in class 7, where 2 samples were incorrectly predicted as class 8, resulting in a recall of 0.93 for class 7.

Despite these localized errors, the overall performance remains strong, with a total accuracy of 92% on the test dataset. The macro-average precision (0.94) and macro-average recall (0.92) further confirm that the model performs consistently across most classes. The weighted F1-score of 0.90 indicates balanced classification performance considering class distribution. Figure 7 illustrates the confusion matrix of the NASNetMobile model for Indonesian Sign Language number classification. Overall, the matrix demonstrates a strong diagonal pattern, indicating that most samples were correctly classified. However, compared to MobileNetV2, several classes show more noticeable misclassification patterns.



Source: (Research Results, 2026)

Figure 7. Confusion Matrix NASNetMobile

Table 8. Classification report on the test set NASNetMobile

Class	Precision	Recall	F1-score	Support
1	1.00	1.00	1.00	32
10	1.00	1.00	1.00	30
2	0.96	0.81	0.88	27
3	1.00	0.03	0.06	32
4	0.55	1.00	0.71	42
5	1.00	1.00	1.00	19
6	1.00	0.97	0.98	32
7	0.97	1.00	0.98	29
8	1.00	0.97	0.98	31
9	0.96	1.00	0.98	26
Accuracy			0.87	300
Macro avg	0.94	0.88	0.86	300
Weighted avg	0.93	0.87	0.84	300

Source: (Research Results, 2026)

Classes 1, 5, 7, 8, and 9 exhibit high classification performance, with most samples correctly predicted along the diagonal. These classes achieve high precision and recall values close to or equal to 1.00, indicating that NASNetMobile successfully learned discriminative

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

features for these gestures. Nevertheless, significant misclassification is observed in certain categories. The most critical case appears in class 3, where only 1 out of 32 samples was correctly classified, while the majority were misclassified as class 4. This explains the extremely low recall (0.03) and F1-score (0.06) for class 3 in the classification report. The confusion between classes 3 and 4 suggests high visual similarity in hand configuration, which may have limited the model's ability to differentiate fine-grained features.

Additionally, class 2 shows partial misclassification, with several samples incorrectly predicted as class 4, resulting in a recall of 0.81. Class 4, while achieving perfect recall (1.00), has a relatively lower precision (0.55), indicating that samples from other classes were frequently misclassified as class 4. This suggests that class 4 may contain features that overlap with neighboring numerical gestures. Despite these localized classification errors, the model achieves an overall testing accuracy of 87%, with macro precision of 0.94 and macro recall of 0.88. However, the lower macro F1-score (0.86) reported in Table 8 reflects instability in class-level performance, particularly due to the severe misclassification in class 3. In summary, while NASNetMobile demonstrates strong performance across most numerical gesture classes, its confusion matrix reveals notable weaknesses in distinguishing certain visually similar gestures. These findings help explain the lower overall testing accuracy compared to MobileNetV2 and highlight the importance of evaluating class-level performance beyond aggregate metrics.

Discussion

The results indicate that the selection of model architecture plays a crucial role in determining classification performance for Indonesian Sign Language number recognition. Although NASNetMobile is a Neural Architecture Search (NAS)-based model designed to optimize network structures automatically, it did not outperform the lightweight MobileNetV2 architecture in this study. MobileNetV2 achieved higher test accuracy (91.67%) compared to NASNetMobile (87.33%), along with superior macro recall and F1-score values. This finding suggests that architectural complexity does not necessarily guarantee better performance, particularly when applied to relatively moderate-sized datasets such as the 3,000-image dataset used in this research. NASNetMobile contains significantly more parameters (4.57 million) compared to MobileNetV2 (2.63 million), which increases computational cost without yielding proportional improvements in predictive performance. In contrast, MobileNetV2's depthwise separable convolution design enables efficient feature extraction while maintaining lower computational overhead.

The confusion matrix analysis further supports this conclusion. While both models performed well on most numerical gesture classes, NASNetMobile exhibited severe misclassification in class 3, resulting in a very low recall value. MobileNetV2, although not perfect, demonstrated more stable class-level recognition and fewer extreme misclassification patterns. This indicates that MobileNetV2 was better able to capture discriminative hand gesture features for visually similar number classes. Additionally, the learning curve analysis shows that MobileNetV2 achieved faster convergence and lower validation loss compared to NASNetMobile. The validation accuracy of MobileNetV2 consistently improved across epochs and reached 89%, whereas NASNetMobile plateaued at approximately 86%. This suggests that MobileNetV2 generalizes more effectively within the given training configuration. From a practical deployment perspective, these results are particularly important. Since sign language recognition systems are often intended for mobile or embedded devices, model efficiency becomes a critical consideration. MobileNetV2 not only achieved higher accuracy but also maintained a smaller model size (10.03 MB), making it more suitable for real-time applications in resource-constrained environments.

Beyond model size and parameter count, the computational efficiency comparison was extended to include training time, inference time, and frames per second (FPS), as detailed in Table 4. MobileNetV2 demonstrated a steady-state training time of approximately 194 seconds per epoch (epochs 2–5), compared to 269 seconds for NASNetMobile, representing a 1.39× speedup in training efficiency. In terms of inference, MobileNetV2 processed each test image in 70.00 ms, achieving a throughput of approximately 14.3 FPS, whereas NASNetMobile required 110.00 ms per image with a throughput of approximately 9.1 FPS. These figures indicate that MobileNetV2 operates 57% faster at inference, further supporting its suitability for real-time sign language recognition on mobile or edge devices. The combined evidence from classification accuracy and empirical computational benchmarks collectively establishes MobileNetV2 as the more practical and efficient architecture for resource-constrained deployment in Indonesian sign language recognition

CC BY NC
Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

applications. Overall, the discussion highlights that for Indonesian Sign Language number classification, a lightweight architecture such as MobileNetV2 provides a better balance between accuracy, stability, and computational efficiency compared to a more complex NAS-based architecture.

CONCLUSION

This study compared MobileNetV2 and NASNetMobile architectures for Indonesian Sign Language (BISINDO) number classification using a transfer learning approach. Experimental results showed that MobileNetV2 achieved superior performance with a test accuracy of 91.67%, outperforming NASNetMobile (87.33%) while maintaining lower computational complexity and smaller model size. These findings indicate that lightweight architectures can provide a better balance between classification accuracy and computational efficiency for resource-constrained deployment scenarios. However, this study is limited by the relatively moderate dataset size and the use of static image-based gestures only. Future research should explore larger and more diverse BISINDO datasets, real-time video-based recognition, and advanced architectures such as attention-based or hybrid CNN–Transformer models to further improve recognition robustness and generalization performance. Finally, expanding the scope beyond number recognition to include full alphabet and dynamic gesture recognition would contribute significantly to the development of comprehensive Indonesian Sign Language recognition systems.

DECLARATION OF AI USE

The authors declare that generative artificial intelligence (AI) tools were used in the preparation of this manuscript. Claude (Anthropic) was used to assist with English language editing, grammar improvement, and the drafting of the declaration statements. AI tools were not used to generate research data, run experiments, analyze results, or draw conclusions. The authors reviewed and revised all AI-assisted content and take full responsibility for the accuracy, originality, and final content of the manuscript.

AUTHOR CONTRIBUTIONS

This journal uses the Contributor Roles Taxonomy (CRediT) to identify individual author contributions to the research and manuscript preparation. The contributions of each author are as follows.

Author	Contribution Roles
Tito Sugiharto	Conceptualization, Methodology, Software, Formal Analysis, Writing Original Draft.
Munya Saleh Ba Matraf	Supervision, Validation, Writing Review & Editing.
Ghizlane Khababa	Validation, Visualization, Writing Review & Editing.
Panji Novantara	Investigation, Data Curation, Resources, Project Administration.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to the leadership of Universitas Kuningan for providing an academic environment and institutional support that enabled this research to be carried out. The authors also thank the Institute for Research and Community Service (LPPM) of Universitas Kuningan for its facilitation, guidance, and administrative support throughout the research process. Finally, the authors acknowledge the research team for their cooperation, dedication, and valuable contributions to data collection, experimentation, and manuscript preparation.

FUNDING

This research was funded by the Institute for Research and Community Service (LPPM), Universitas Kuningan, through its research funding program.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest related to the research, writing, or publication of this article. The funding body, the Institute for Research and Community Service (LPPM) of Universitas Kuningan, had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. No author has any financial or personal relationship with other people or organizations that could inappropriately influence the work reported in this article.

REFERENCE

[1] A. Wibowo, R. Zulpani, A. P. Windarto, A. Wanto, and S. R. Andani, “Enhancing Herbal Plant Leaf Image Detection Accuracy Through Mobilenet Architecture Optimization in CNN,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 859–867, May 2025, doi: 10.33480/jitk.v10i4.6498.

[2] K. N. A. Nur, N. I. Addyna, A. P. Windarto, A. Wanto, and P. Poningsih, “Optimization Of The

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8342

Inceptionv3 Architecture For Potato Leaf Disease Classification," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 849–858, May 2025, doi: 10.33480/jitk.v10i4.6554.

[3] T. Sugiharto, Saparudin, and W. F. Al Maki, "Indonesian Cued Speech Transliterate System Using Convolutional Neural Network MobileNet," in *2024 9th International Conference on Informatics and Computing, ICIC 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICIC64337.2024.10957117.

[4] G. Gumelar, T. Sugiharto, and I. Lesmana, "Implementasi CNN untuk identifikasi penyakit daun jagung Implementation of CNN for Corn Leaf Disease Identification," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 2, pp. 175–180, 2025, doi: 10.37859/coscitech.v6i2.9462.

[5] A. Pratama, T. Sugiharto, and P. Novantara, "Klasifikasi Jenis Tanaman Alpukat Berdasarkan Bentuk Daun Menggunakan Algoritma CNN," *Jurnal Computer Science and Information Technology (CoSciTech)*, vol. 6, no. 2, pp. 120–128, 2025, doi: 10.37859/coscitech.v6i2.9474.

[6] M. A. Fathan, T. Sugiharto, and I. Lesmana, "Klasifikasi jenis tanaman philodendron berdasarkan citra daun menggunakan algoritma CNN Classification of philodendron plant species based on leaf images using a CNN algorithm," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 2, pp. 140–147, 2025, doi: 10.37859/coscitech.v6i2.9484.

[7] I. Lesmana, T. Sugiharto, G. Priguna, and D. Ramdani, "Classification of Rattan Plant use CNN (Convolutional Neural Network) Method for Detection Based On Android," in *Proceedings of the 4th International Symposium on Materials And Electrical Engineering (ISMEE 2023): Empowering Sustainable Innovation through Emerging Technologies in Materials and Electrical Engineering*, Bali, Indonesia: AIP Publishing, 2025. doi: https://doi.org/10.1063/5.0261714.

[8] Larissa Navia Rani, Mardison, and A. P. Windarto, "Performance Comparison Of Deep CNN Architectures for Lung Area Segmentation in Chest Imaging," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 11, no. 1, pp. 87–98, Aug. 2025, doi: 10.33480/jitk.v11i1.6735.

[9] S. Vallukappully, I. van der Linde, and A. Chakraborty, "Early detection and classification of diabetic retinopathy by transfer learning of NASNet-large and ResNet-50 convolutional neural networks," *Inform. Med. Unlocked*, vol. 58, Jan. 2025, doi: 10.1016/j.imu.2025.101688.

[10] T. K. Njoroge, E. J. Omol, and V. O. Nyangaresi, "Deep Learning and IoT Fusion for Crop Health Monitoring: A High-Accuracy, Edge-Optimised Model for Smart Farming," *IET Image Process.*, vol. 19, no. 1, Jan. 2025, doi: 10.1049/ipr2.70208.

[11] S. Liu, W. Kong, Z. Liu, J. Sun, S. Liu, and D. Gašević, "Dual-view cross attention enhanced semi-supervised learning method for discourse cognitive engagement classification in online course discussions," *Expert Syst. Appl.*, vol. 278, Jun. 2025, doi: 10.1016/j.eswa.2025.127339.

[12] H. AbdElghfar, H. A. Youness, M. Wahba, and H. M. Abdelaal, "An automated framework for qur'anic education of the hearing-impaired using body pose classification and Arabic sign language integration," *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-026-36578-z.

[13] T. Tene, D. F. Vique López, P. E. Valverde Aguirre, L. M. Orna Puente, and C. Vacacela Gomez, "Virtual reality and augmented reality in medical education: an umbrella review," 2024, *Frontiers Media SA*. doi: 10.3389/fdgth.2024.1365345.

[14] M. Rifki, A. Bastian, and A. Mardiana, "Development of CNN-LSTM-Based Image Captioning Dataset to Enhance Visual Accessibility For Disabilities," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 980–992, Jun. 2025, doi: 10.33480/jitk.v10i4.6657.

[15] T. Sugiharto *et al.*, *Pengolahan Citra Digital dan Deteksi Objek*. Malang: PT Literasi Nusantara Abadi Grup, 2025. [Online]. Available: www.penerbitlitnus.co.id

[16] P. Niu, "Convolutional neural network for gesture recognition human-computer interaction system design," *PLoS One*, vol. 20, no. 2 February, Feb. 2025, doi: 10.1371/journal.pone.0311941.

[17] Q. Liu, X. She, and Q. Xia, "AI based diagnostics product design for osteosarcoma cells microscopy imaging of bone cancer patients using CA-MobileNet V3," *J. Bone Oncol.*, vol. 49, Dec. 2024, doi: 10.1016/j.jbo.2024.100644.

[18] J. Sasidevi, A. Sathish, S. Vatchala, and M. Nallusamy, "NASNet with African vulture

CC BY NC

optimization for detecting diabetic retinopathy stages in retinal fundus images," *Expert Syst. Appl.*, vol. 296, Jan. 2026, doi: 10.1016/j.eswa.2025.128910.

[19] X. Liu, Z. Zhou, E. Xia, and X. Yin, "SSTA-ResT: Soft Spatiotemporal Attention ResNet Transformer for Argentine Sign Language Recognition," *Sensors*, vol. 25, no. 17, Sep. 2025, doi: 10.3390/s25175543.

[20] D. Uthus, G. Tanzer, and M. Georg, "YouTube-ASL: A Large-Scale, Open-Domain American Sign Language-English Parallel Corpus," in *37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks*, Jun. 2023. [Online]. Available: http://arxiv.org/abs/2306.15162

[21] S. Bankar, T. Kadam, V. Korhale, and M. A. A. Kulkarni, "Real Time Sign Language Recognition Using Deep Learning," *International Research Journal of Engineering and Technology*, 2022, [Online]. Available: www.irjet.net

[22] S. Aliyev, A. A. Almisreb, and S. Turaev, "Azerbaijani sign language recognition using machine learning approach," in *Journal of Physics: Conference Series*, Institute of Physics, May 2022. doi: 10.1088/1742-6596/2251/1/012007.

[23] B. Alsharif, A. S. Altaher, A. Altaher, M. Ilyas, and E. Alalwany, "Deep Learning Technology to Recognize American Sign Language Alphabet," *Sensors*, vol. 23, no. 18, Sep. 2023, doi: 10.3390/s23187970.

[24] 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 (Switzerland)*, vol. 11, no. 11, Jun. 2022, doi: 10.3390/electronics11111780.

[25] G. H. Samaan *et al.*, "MediaPipe's Landmarks with RNN for Dynamic Sign Language Recognition," *Electronics (Switzerland)*, vol. 11, no. 19, Oct. 2022, doi: 10.3390/electronics11193228.

[26] I. Qutab, L. Po, F. Rollo, and W. Naqvi, "Optimizing Sign Language Recognition Through a Tailored MobileNet Self-Attention Framework," *Applied Sciences (Switzerland)*, vol. 15, no. 23, Dec. 2025, doi: 10.3390/app152312622.

[27] K. K. Podder *et al.*, "Bangla Sign Language (BdSL) Alphabets and Numerals Classification Using a Deep Learning Model," *Sensors*, vol. 22, no. 2, Jan. 2022, doi: 10.3390/s22020574.

[28] J. Bora, S. Dehingia, A. Boruah, A. A. Chetia, and D. Gogoi, "Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 1384–1393. doi: 10.1016/j.procs.2023.01.117.

[29] A. Lepeut and L. Ferrara, "Backchannels as multimodal displays of joint understanding in signed language interaction: A cross-linguistic analysis of Norwegian Sign Language and French Belgian Sign Language conversations," *J. Pragmat.*, vol. 264, pp. 96–112, Oct. 2026, doi: 10.1016/j.pragma.2026.07.013.

[30] M. Papatsimouli, A. Lytos, P. Sarigiannidis, and G. F. Fragulis, "IoT-powered assistive technology for real-time translation of sign language," Dec. 01, 2026, *KeAi Communications Co.* doi: 10.1016/j.ijcce.2026.05.008.

[31] R. Rastgoo, K. Kiani, and S. Escalera, "Recognizing word in continuous sign language using a model trained exclusively on isolated-word datasets," *Results in Engineering*, vol. 31, Sep. 2026, doi: 10.1016/j.rineng.2026.111443.

[32] G. Du, L. Wang, K. Hu, and Z. Wang, "KANMultiSign: Multi-scale sequence-based pose animation from sign language notation with Kolmogorov-Arnold networks," *Neurocomputing*, vol. 694, Sep. 2026, doi: 10.1016/j.neucom.2026.133930.

[33] K. Aurangzeb, K. Javeed, M. Alhussein, I. Rida, S. I. Haider, and A. Parashar, "Deep Learning Approach for Hand Gesture Recognition: Applications in Deaf Communication and Healthcare," *Computers, Materials and Continua*, vol. 78, no. 1, pp. 127–144, 2024, doi: 10.32604/cmc.2023.042886.

[34] M. A. Hamza, A. Subahi, N. A. Alghanmi, and N. Nemri, "Deep fusion based transfer learning with bald eagle search algorithm for sign language recognition to assist individuals with hearing and speech impairments," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-10660-4.

[35] S. Sharma and S. Singh, "ISL recognition system using integrated mobile-net and transfer learning method," *Expert Syst. Appl.*, vol. 221, Jul. 2023, doi: 10.1016/j.eswa.2023.119772.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.