Buletin Ilmiah Sarjana Teknik Elektro
Vol. 7, No. 3, September 2025, pp. 350-361
ISSN: 2685-9572, DOI: 10.12928/biste.v7i3.13300

Improving Indonesian Sign Alphabet Recognition for Assistive Learning Robots Using Gamma-Corrected MobileNetV2

Lilis Nur Hayati [1,3], Anik Nur Handayani [1], Wahyu Sakti Gunawan Irianto [1], Rosa Andrie Asmara [2], Dolly Indra [3], Nor Salwa Damanhuri [4]
[1] Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Malang, Indonesia
[2] Information Technology Department, Politeknik Negeri Malang, Malang, Indonesia
[3] Department of Computer Science, Universitas Muslim Indonesia, Makassar, Indonesia
[4] Electrical Engineering Studies, Universiti Teknologi MARA (UiTM), Cawangan Pulau Pinang, Malaysia

ARTICLE INFORMATION

Article History:

Received 20 April 2025
Revised 14 July 2025
Accepted 23 July 2025

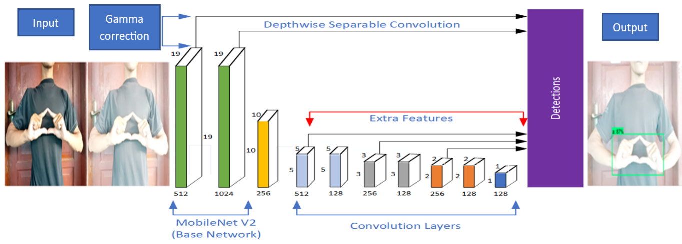
Keywords:

BISINDO;
MobileNetV2;
Gamma Correction;
SSD;
Assistive Robots

Corresponding Author:

Anik Nur Handayani,
Department of Electrical Engineering and Informatics,
Universitas Negeri Malang,
Malang, Indonesia.
Email: aniknur.ft@um.ac.id

ABSTRACT



Sign language recognition plays a critical role in promoting inclusive education, particularly for deaf children in Indonesia. However, many existing systems struggle with real-time performance and sensitivity to lighting variations, limiting their applicability in real-world settings. This study addresses these issues by optimizing a BISINDO (Bahasa Isyarat Indonesia) alphabet recognition system using the SSD MobileNetV2 architecture, enhanced with gamma correction as a luminance normalization technique. The research contribution is the integration of gamma correction preprocessing with SSD MobileNetV2, tailored for BISINDO and implemented on a low-cost assistive robot platform. This approach aims to improve robustness under diverse lighting conditions while maintaining real-time capability without the use of specialized sensors or wearables. The proposed method involves data collection, image augmentation, gamma correction (γ = 1.2, 1.5, and 2.0), and training using the SSD MobileNetV2 FPNLite 320x320 model. The dataset consists of 1,820 original images expanded to 5,096 via augmentation, with 26 BISINDO alphabet classes. The system was evaluated under indoor and outdoor conditions. Experimental results showed significant improvements with gamma correction. Indoor accuracy increased from 94.47% to 97.33%, precision from 91.30% to 95.23%, and recall from 97.87% to 99.57%. Outdoor accuracy improved from 93.80% to 97.30%, with precision rising from 90.33% to 94.73%, and recall reaching 100%. In conclusion, the proposed system offers a reliable, real-time solution for BISINDO recognition in low-resource educational environments. Future work includes the recognition of two-handed gestures and integration with natural language processing for enhanced contextual understanding.

Document Citation:

L. N. Hayati, A. N. Handayani, W. S. G. Irianto, R. A. Asmara, D. Indra, and N. S. Damanhuri, "Improving Indonesian Sign Alphabet Recognition for Assistive Learning Robots Using Gamma-Corrected MobileNetV2," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 7, no. 3, pp. 350-361, 2025, DOI: 10.12928/biste.v7i3.13300.

1. INTRODUCTION

Sign language serves as the primary mode of communication for individuals who are deaf or speech-impaired [1][2]. In the context of Indonesian language and culture, BISINDO (*Bahasa Isyarat Indonesia*) plays a vital role in promoting literacy and language acquisition among children with hearing impairments [3][4]. However, teaching BISINDO particularly at the alphabet levelremains challenging due to a limited number of qualified instructors, the lack of interactive and adaptive learning media, and the insufficient availability of affordable assistive technologies tailored to the needs of children with disabilities [5][6]. Computer vision-based sign language recognition systems have emerged as a promising solution to bridge this gap [7]–[9]. Several studies have employed deep learning architectures such as Convolutional Neural Networks (CNN) [10]–[12], You Only Look Once (YOLO) [13]–[15], and Faster Region-Based Convolutional Neural Network (Faster R-CNN) [13]–[15] for sign alphabet classification [19][20]. Although these models demonstrate high accuracy, they generally require high computational power and often struggle with real-time performance, making them less suitable for embedded systems such as mobile or child-friendly educational robots.

To address these limitations, this study proposes the use of Single Shot MultiBox Detector (SSD) [21] integrated with the lightweight MobileNetV2 [22] architecture, enhanced through gamma correction preprocessing. SSD MobileNetV2 was selected for its balance of speed and accuracy on low-resource devices, while gamma correction is used to normalize lighting variations that commonly occur in gesture images [23]. This combination is deployed in a low-cost, camera-based educational robot designed to assist BISINDO alphabet learning in real time, without relying on specialized sensors or wearable devices. The system provides multimodal feedback, including visual text display, expressive robot movements, and audio output [24]. The contribution of this research is the development of an adaptive, real-time BISINDO recognition system optimized through gamma correction and embedded into an assistive educational robot [25]. The system addresses key challenges in lighting variability and real-time performance while offering a practical, inclusive learning solution for deaf children [26][27]. Future developments will include support for two-handed gestures and integration with Natural Language Processing (NLP) to enhance interactive dialogue and contextual understanding.

2. THEORETICAL FOUNDATION OF SSD MOBILENETV2

SSD MobileNet V2 is an object detection network that combines MobileNet V2 for feature extraction and the Single Shot Multibox Detector (SSD) for object detection [21]. In Figure 1, the MobileNet-SSD architecture is depicted, where we introduce gamma-corrected image results. Subsequently, SSD MobileNet V2 is utilized for object detection and recognition, explicitly focusing on the BISINDO letters. MobileNetV2 employs depthwise separable convolutions, consisting of depthwise convolutions followed by pointwise convolutions [28]. This design reduces the number of parameters and computations compared to traditional convolutional layers while still capturing meaningful features. In the SSD MobileNet V2 architecture, the convolutional layers are a fundamental component responsible for feature extraction from input images. These layers apply convolution operations to input data, learning hierarchical features at various scales to generate feature maps.

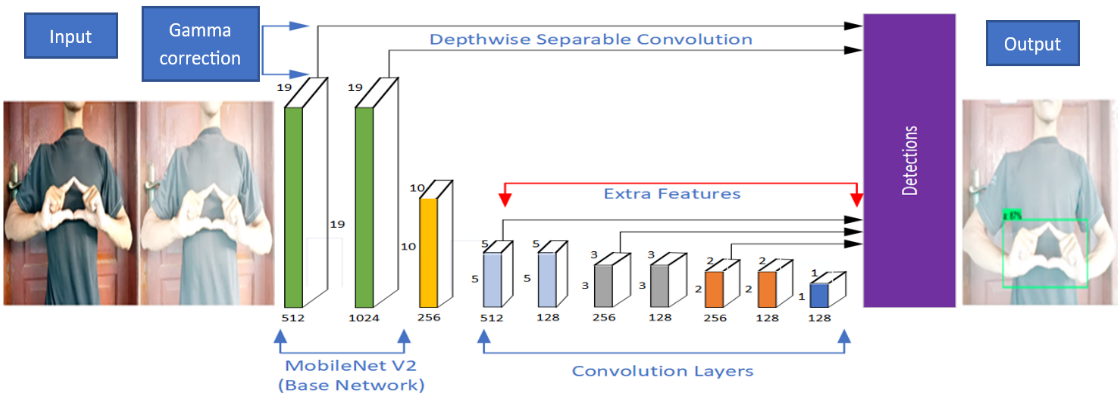


Figure 1. The architecture of MobileNet-SSD V2 incorporates gamma correction

This research employed the SSD MobileNet V2 FPNLite 320x320 model based on Table 1. This model comprises several vital elements, namely SSD (Single Shot Multibox Detector), an object detection architecture designed to detect objects in an image efficiently. FPNLite (Feature Pyramid Network Lite) is a lightweight version of the Feature Pyramid Network (FPN) designed to address scale issues in object detection. FPNLite aids the model in effectively identifying objects at various resolution levels. The model uses a 320x320 image resolution as the input. In this case, the input image has dimensions of 320x320 pixels. This resolution can influence the trade-off between the speed and accuracy of object detection.

3. METHODS

This section outlines the methodological stages of the proposed BISINDO alphabet recognition system [29][30]. The process is divided into three main phases: (1) Data Preparation, (2) Model Development, and (3) Testing Pipeline. Each phase is designed to support the development of a real-time, lightweight recognition system optimized for deployment in assistive educational robots.

3.1. Data Preparation

3.1.1. Dataset Collection

The dataset comprises 1,820 original RGB images representing 26 BISINDO alphabet letters, demonstrated by a single actor using a Realme 8i smartphone (50 MP) under varying lighting conditions (both indoor and outdoor) [30]. The camera was positioned at a fixed 70 cm distance and aligned with the chest height of the subject to ensure consistency [31].

3.1.2. Preprocessing

All images were labeled into 26 distinct classes (A–Z), then resized and cropped to 640×640 pixels using the Roboflow platform [32][33]. The dataset was split into 90% for training (1,638 images) and 10% for testing (182 images). This proportion was selected to maximize learning while retaining a small holdout set for evaluation [34].

3.1.3. Data Augmentation

To improve generalization and simulate real-world variability, four augmentation techniques were applied: brightness (±25%), exposure (±20%), saturation (±40%), and hue (±30°) [33],[35],[36]. These values were chosen to reflect common environmental lighting fluctuations observed in educational settings [37][38]. Figure references illustrate the visual impact of each technique (Figure 2 to Figure 5). Brightness is the degree of lightness or darkness of a color, influenced by the amount of reflected light. Brightness is used to achieve image variations between dark and bright conditions [24]. We employ the brightness values -25% and +25%, as illustrated in Figure 2. Exposure reflects the overall light level in a video or image. We use the exposure values -20% and + 20%, as depicted in Figure 3. Saturation refers to the intensity of a color measured on a chroma scale. It indicates the extent to which colors in an image mix with shades of gray. We use saturation values of -40% and + 40%, as illustrated in Figure 4. Hue, a component of the HSV (Hue, Saturation, Value) colour model, determines primary colours like red, yellow [39], blue, and secondary colours like orange, green, and purple. The hue values we use are -30° and 30°, as shown in Figure 5. After the augmentation process, the image data amounting to 5096 images can be formulated as follows, where the validation data is 0. The total number of images after augmentation is calculated as:

$$\begin{aligned} \text{Augmentasi Data} &= (\text{Training Data x } 3) + \text{Validation Data } + \text{Testing Data} \\ &= (1{,}638\ \times 3) + 0 + 182 \\ &= \ 5{,}096 \text{ images} \end{aligned} \tag{1}$$

This augmentation did not include a validation set due to limited data availability. Instead, model evaluation was performed using k-fold cross-validation.

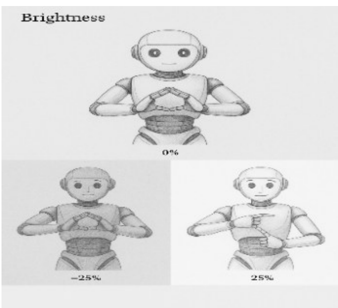


Figure 2. Brightness in The Image

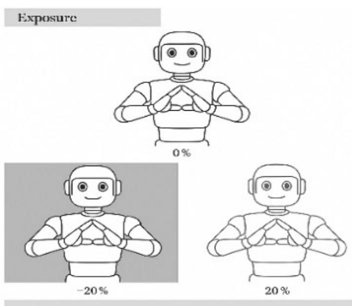


Figure 3. Exposure in The Image

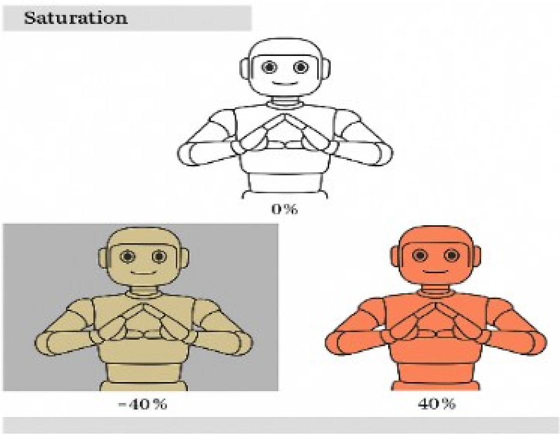


Figure 4. Saturation in The Image

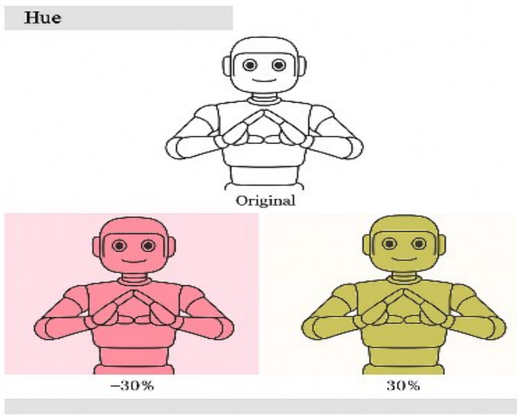


Figure 5. Hue in The Image

3.2. Model Development

3.2.1. Model Selection

The SSD MobileNetV2 FPNLite 320×320 model was selected based on its low computational complexity and real-time suitability [40][41]. As shown in Table 1, it offered the best trade-off between speed (22 ms) and mean Average Precision (22.2 mAP) among several candidates from the TensorFlow 2 Model Zoo. The lightweight architecture of MobileNetV2, which utilizes depthwise separable convolutions, enables efficient feature extraction without compromising accuracy. The FPNLite component enhances multi-scale object detection by refining feature maps at different resolutions.

Table 1. Pre-trained Tensorflow 2 Model Zoo

No	Model Name	Speed (ms)	COCO mAP
1.	SSD MobileNet V1 FPN 640x640	48	29.1
2.	SSD MobileNet V2 FPNLite 320x320	22	22.2
3.	SSD MobileNet V2 FPNLite 640x640	39	28.2
4.	SSD ResNet101 V1 FPN 640x640 (RetinaNet101)	57	35.6
5.	SSD ResNet152 V1 FPN 640x640 (RetinaNet152)	80	35.4
6.	Faster R-CNN ResNet50 V1 640x640	53	29.3
7.	EfficientDet D6 1280x1280	268	50.5

3.2.2. Model Training

Model training was performed on Google Colaboratory using TensorFlow. A batch size of 52 and 3,000 training steps were applied. The training data was converted into the TFRecord format. We trained the model with 26 class labels and monitored training loss and accuracy in real time.

3.3. Testing Pipeline

The testing pipeline comprises four main stages: Input Acquisition, Gamma Correction, Dataset Loading, and Real-Time Recognition.

3.3.1. Input Acquisition

For testing, image input was acquired via a Logitech BRIO 4K Pro webcam, directly capturing live gestures from the user. The camera was integrated into a notebook interface using OpenCV for real-time frame acquisition. High-resolution input ensures robust feature extraction by the trained model.

3.3.2. Gamma Correction

Gamma correction is used to adjust the brightness and contrast of images to align with human visual perception [42][43]. This technique involves applying a nonlinear operation to the pixel values of the image. The purpose is to correct images that are either too bright or too dark proportionally [44]. Gamma correction helps the model recognize objects under varying lighting conditions. The gamma correction function is mathematically expressed as:

$$V_{out} = A.\left(\frac{V_{in}}{A}\right)^{\gamma}$$

Vol. 7, No. 3, September 2025, pp. 350-361

where $V_{in} \in \{R, G, , B\}, A = 255$, and γ is the gamma velue (e.g., 1.2, 1.5, or 2.0). This operation enhances visibility of hand shapes under poor lighting, making the system more resilient to real-world conditions.

3.3.3. Load Trained Model

The trained model, stored as a .ckpt (checkpoint) file in Google Drive, was loaded during testing. This file contains the model weights and configuration, enabling recognition without re-training. Although referred to as a "dataset" in prior text, .ckpt is more accurately described as a model checkpoint.

3.3.4. Real-Time Recognition

Using OpenCV and Jupyter Notebook, the system performs real-time recognition of BISINDO hand gestures. Detected classes are displayed with bounding boxes and confidence scores [45]. The hardware used for testing is summarized in Table 2.

Table 2. Hardware Spesification

No	Hardware	Specification
1	Webcam	Logitech BRIO 4K Pro
2	CPU	Intel Core i7-11370H 3.3GHz (4 Cores)
3	GPU	GTX 1650 4 GB VRAM
4	VGA	NVIDIA® GeForce® GTX 1050
5	RAM	8 GB DDR4-3200 MHz
6	Storage	512 GB SSD NVMe PCIe

4. RESULT AND DISCUSSION

This section presents the results of model evaluation, indoor and outdoor testing, and practical integration into an assistive learning robot. A comprehensive discussion is provided to contextualize the findings, benchmark against prior research, and explore the implications of the proposed gamma-corrected recognition model.

4.1. Cross-Validation Performance

The system was first evaluated using 15-fold cross-validation to assess the model's generalizability. As shown in Table 3, the average precision was 64% and recall reached 70%. This discrepancy compared to subsequent testing phases indicates the baseline performance of the model without gamma correction or testing augmentation. Although moderate, these results demonstrate the model’s capacity to learn meaningful representations across folds. The relatively lower values, in contrast to high accuracy in later sections, suggest that lighting inconsistencies in the raw dataset posed challenges during training—thus justifying the use of gamma correction for further optimization.

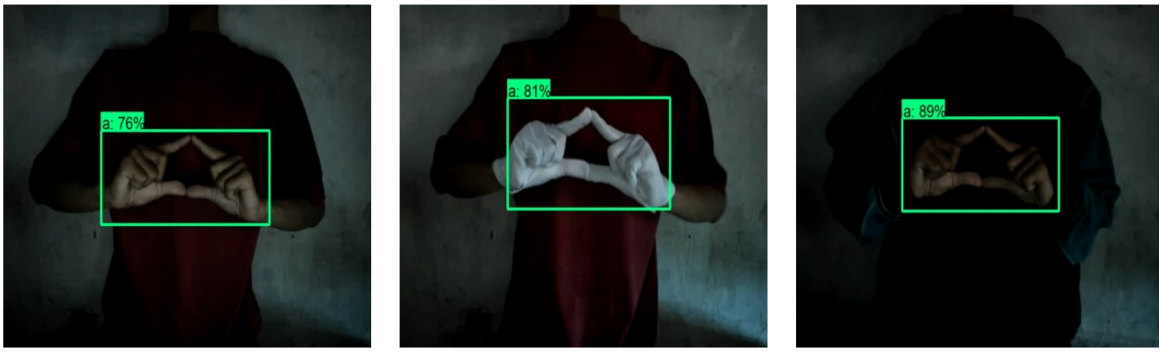
Table 3. Cross-validation

Cross-validation	Precision	Recall
Iteration 1	0.565227	0.683333
Iteration 2	0.603636	0.672727
Iteration 3	0.625000	0.625000
Iteration 4	0.666667	0.741667
Iteration 5	0.683333	0.750000
Iteration 6	0.700610	0.733333
Iteration 7	0.706944	0.766667
Iteration 8	0.670833	0.783333
Iteration 9	0.680903	0.750000
Iteration 10	0.665000	0.666667
Iteration 11	0.651389	0.650000
Iteration 12	0.577143	0.616667
Iteration 13	0.558333	0.683333
Iteration 14	0.511071	0.641667
Iteration 15	0.457639	0.575000
Average	**64%**	**70%**

4.2. Indoor Testing Results

4.2.1. Without Gamma Correction

In this testing phase, we employed two scenarios: one without gamma correction and another with gamma correction, using illumination measured with a lux light meter [46]. The testing without gamma correction is depicted in Figure 6(a) to Figure 6(c). Based on the indoor testing results, the average values indicate a relatively dark environment for both objects and backgrounds. Indoor experiments were conducted in dimly lit environments (30 lux), simulating common classroom lighting. As depicted in Figure 6 and summarized in Table 4, the model achieved an average accuracy of **94.47%**, precision of **91.30%**, and recall of **97.87%** across three test iterations. These results confirm the baseline performance of the SSD MobileNetV2 architecture in low-light indoor conditions, although slight precision dips were observed under uneven background contrast.



(a) Testing 1 (b) Testing 2 (c) Testing 3

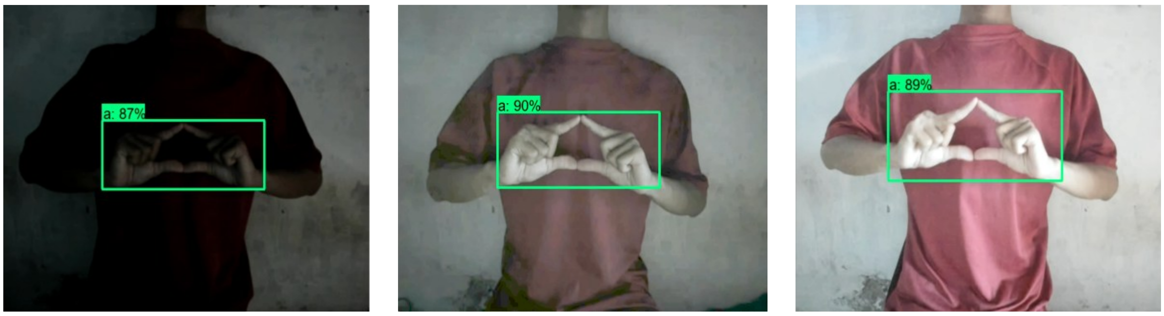
Figure 6. Indoor Testing without Gamma Correction

Table 4. Indoor Testing without Gamma Correction

No	Scenario	Accuracy	Precision	Recall
1	Testing 1	0.933	0.895	0.974
2	Testing 2	0.950	0.903	1
3	Testing 3	0.951	0.941	0.962
Average		**94.47%**	**91.30%**	**97.87%**

4.2.2. With Gamma Correction

Figure 7(a) to Figure7(c) illustrate testing using gamma correction with gamma correction values of 1, 2, and 3. The indoor testing results with excellent average values show that objects and backgrounds appear brighter compared to those not using gamma correction. Applying gamma correction (γ = 1, 2, and 3) yielded notable improvements. As shown in Figure 7 and Table 5, accuracy increased to **97.33%**, precision to **95.23%**, and recall to **99.57%**. Visual clarity of gestures improved significantly due to luminance normalization, especially in darker regions. The gamma-corrected model outperformed the baseline across all metrics. The increase in recall demonstrates better sensitivity to gesture variations. However, the impact of different gamma values warrants future ablation studies to isolate optimal gamma levels.



(a) Testing 1 (b) Testing 2 (c) Testing 3

Figure 7. Indoor Testing using Gamma Correction

Vol. 7, No. 3, September 2025, pp. 350-361

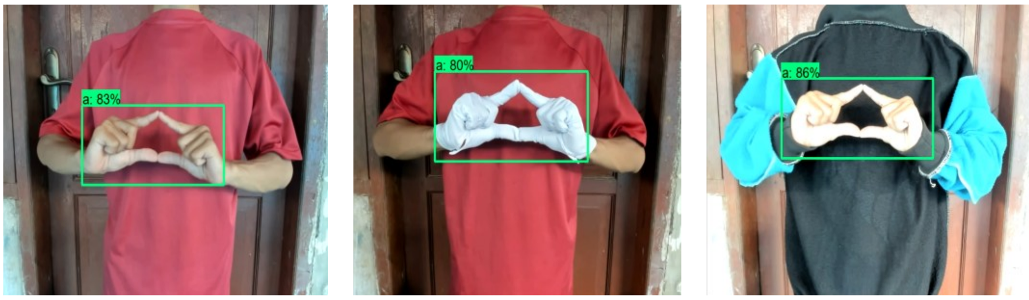
Table 5. Indoor Testing Using Gamma Correction

No	Scenario	Accuracy	Precision	Recall
1	Testing 1	0.969	0.951	0.987
2	Testing 2	0.970	0.943	1
3	Testing 3	0.981	0.963	1
	Average	**97.33%**	**95.23%**	**99.57%**

4.3. Outdoor Testing Results

4.3.1. Without Gamma Correction

In this test, we employed two scenarios: one without gamma correction and the other using gamma correction, with lighting measured using a lux light meter, as conducted in the indoor testing. The testing without gamma correction is depicted in Figure 8 (a) to Figure 8(c). Based on the indoor testing results with a good average value, the object and the background appear bright. In this experiment, during testing with gamma correction, we maintained a distance of 50 cm between the actor and the camera, with the camera height aligned with the actor's chest and lighting set to 278 lux with 3 tests each. Table 6 presents the indoor testing results using gamma correction. Table 6 shows that the average test values are excellent, with accuracy, precision, and recall values of 93.8%, 90.33%, and 97.40%, respectively.



(a) Testing 1 (b) Testing 2 (c) Testing 3

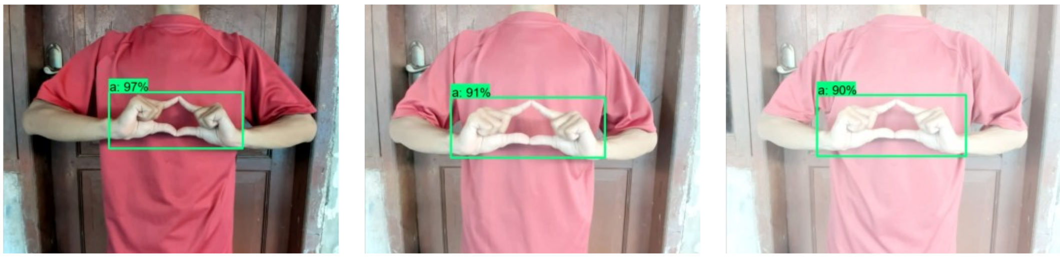
Figure 8. Outdoor Testing without Gamma Correction

Table 6. Outdoor Testing without Gamma Correction

No	Scenario	Accuracy	Precision	Recall
1	Testing 1	0.958	0.952	0,962
2	Testing 2	0.952	0.918	0,987
3	Testing 3	0.904	0.840	0,973
	Average	**93.8%**	**90.33%**	**97.40%**

4.3.2. With Gamma Correction

Figure 9 (a) to Figure(c) illustrate the testing with gamma correction, with gamma correction values of 1, 2, and 3, respectively. These gamma values are the same as those used in the earlier indoor testing. Based on the results of the indoor testing with excellent average values, both the objects and the background appear bright. In this experiment, during testing with gamma correction, we maintained a distance of 50 cm between the actor and the camera, with the camera's height aligned with the actor's chest and lighting set at 30 lux with 3 tests each. Table 7 presents the results of indoor testing using gamma correction. Table 7 shows excellent average test values, with accuracy, precision, and recall values of 97.30%, 94.73%, and 100.00%, respectively. Based on the test results in Table 6, testing with gamma correction yields better results outdoors than testing without gamma correction.



(a) Testing 1 (b) Testing 2 (c) Testing 3

Figure 9. Outdoor Testing using Gamma Correction

Table 7. Outdoor Testing with Gamma Correction

No	Scenario	Accuracy	Precision	Recall
1	Testing 1	0.981	0.963	1
2	Testing 2	0.963	0.928	1
3	Testing 3	0.975	0.951	1
Average		**97.30%**	**94.73%**	**100.00%**

4.4. Implementation Strategy and Robotic Integration

The optimized model was deployed in an assistive robot prototype designed for real-time interaction with deaf children. The model runs on a lightweight edge computing platform, requiring no high-end GPUs or cloud dependency [47]. Real-time inference operates under 100 ms per frame, enabling responsive interaction. A gamma correction module was embedded at the preprocessing stage, normalizing hand gesture frames before feeding them into the SSD MobileNetV2 classifier. This process mitigates fluctuations in classroom lighting and ensures recognition reliability. The robot provides multimodal feedback—displaying letters on-screen, offering voice outputs, and performing simple animations. This interactive loop enhances learning engagement. Data logging also enables learning progress tracking, offering pedagogical insights for educators and parents [48]. Importantly, the system avoids complex hardware like depth cameras or gloves, prioritizing accessibility. By using BISINDO, it aligns with local educational standards, reinforcing cultural relevance. The high performance across varying environments confirms its deploy ability in both school and home settings.

4.5. Comparative Discussion and Implications

Compared to previous studies in Malaysian Sign Language (99.75% accuracy with SSD-MobileNetV2) and ASL (99.91% accuracy with CNN-RGB), our system's indoor/outdoor accuracy of 97.3% with gamma correction is competitive [49]. The novelty lies in its real-time implementation with lighting normalization and resource-constrained deployability, aspects often omitted in prior works [50]. This study demonstrates the potential of vision-based systems in inclusive education [12]. However, limitations remain: only static, single-hand gestures are supported, and dynamic gesture transitions or occlusions remain untested. User experience trials are also pending, which are essential for validating real-world applicability.

5. CONCLUSIONS

This study has presented a gamma-corrected SSD MobileNetV2 model for BISINDO alphabet recognition, with a focus on supporting inclusive educational technologies through real-time assistive robotics. The experimental evaluations under both indoor and outdoor scenarios demonstrate that gamma correction significantly improves model robustness against lighting variations, leading to notable enhancements in classification accuracy, precision, and recall. Indoor testing without gamma correction yielded an average accuracy of 94.47%, precision of 91.30%, and recall of 97.87%. After applying gamma correction, performance improved to 97.33%, 95.23%, and 99.57%, respectively. Similar gains were observed in outdoor testing, with accuracy increasing from 93.80% to 97.30%, precision from 90.33% to 94.73%, and recall achieving 100% post-correction. These findings confirm that gamma correction effectively addresses luminance inconsistency, enhancing recognition reliability across diverse environments.

The integration of this lightweight model into a camera-based assistive robot illustrates the feasibility of deploying sign language recognition systems in low-resource educational settings. By eliminating reliance on wearable sensors or depth cameras, the system offers a cost-efficient and accessible learning platform. Its multimodal feedback—through visual display, voice output, and responsive animation—supports interactive and autonomous alphabet learning for deaf children. Despite these promising results, several limitations remain. The system is currently restricted to single-hand static gestures and has not yet undergone longitudinal testing in live classroom environments. Moreover, the absence of statistical validation, such as confidence intervals or significance testing, presents a potential gap in confirming the generalizability of the observed improvements. Future work will focus on expanding the system's capabilities to include dynamic and two-handed gesture recognition, as well as integrating Natural Language Processing (NLP) modules to support contextual understanding and sentence-level interpretation. Additionally, formal user studies involving children and educators will be conducted to evaluate pedagogical effectiveness, usability, and engagement. Ethical considerations regarding data privacy, especially in child-focused deployments, will also be addressed to ensure safe and inclusive AI applications in educational domains. In summary, this research contributes a practical and scalable solution for BISINDO recognition by combining computational efficiency, illumination robustness, and real-time responsiveness—laying a strong foundation for future advancements in AI-driven inclusive education.

DECLARATION

Author Contribution

All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

[1] N. Thiracitta and H. Gunawan, “SIBI sign language recognition using convolutional neural network combined with transfer learning and non-trainable parameters,” *Procedia Computer Science*, vol. 179, pp. 72-80, 2021, https://doi.org/10.1016/j.procs.2020.12.011.

[2] R. Sutjiadi, S. Sendari, H. W. Herwanto, and Y. Kristian, “Generating High-quality Synthetic Mammogram Images Using Denoising Diffusion Probabilistic Models: A Novel Approach for Augmenting Deep Learning Datasets,” *2024 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2024 - Proc.*, pp. 386–392, 2024, https://doi.org/10.1109/ICITSI65188.2024.10929446.

[3] S. Dwijayanti, Hermawati, S. I. Taqiyyah, H. Hikmarika, and B. Y. Suprapto, “Indonesia Sign Language Recognition using Convolutional Neural Network,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 415–422, 2021, https://doi.org/10.14569/IJACSA.2021.0121046.

[4] S. Subburaj and S. Murugavalli, “Survey on sign language recognition in context of vision-based and deep learning,” *Measurement: Sensors*, vol. 23, p. 100385, 2022, https://doi.org/10.1016/j.measen.2022.100385.

[5] K. G. Nalbant and Ş. Uyanik, “Computer Vision in the Metaverse,” *J. Metaverse*, vol. 1, no. 1, pp. 9–12, 2021, https://dergipark.org.tr/en/pub/jmv/issue/67581/1051377.

[6] Y. Yusmanto, H. Ar, and A. Prasetya, “Design and development of face recognition-based security system using expression game as liveness detection,” *Bull. Soc. Informatics Theory Appl.*, vol. 8, no. 2, pp. 280–294, 2024, https://doi.org/10.31763/businta.v8i2.756.

[7] A. Elhagry and R. G. Elrayes, “Egyptian Sign Language Recognition Using CNN and LSTM,” *arXiv preprint arXiv:2107.13647*, 2021, https://doi.org/10.48550/arXiv.2107.13647.

[8] B. T. Abeje, A. O. Salau, A. D. Mengistu, and N. K. Tamiru, “Ethiopian sign language recognition using deep convolutional neural network,” *Multimed. Tools Appl.*, vol. 81, no. 20, pp. 29027–29043, 2022, https://doi.org/10.1007/s11042-022-12768-5.

[9] F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O. Mahmoud, and D. S. Abdelminaam, “Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language,” *IEEE Access*, vol. 9, pp. 67548–67558, 2021, https://doi.org/10.1109/ACCESS.2021.3077386.

[10] E. Aldhahri *et al.*, “Arabic Sign Language Recognition Using Convolutional Neural Network and MobileNet,” *Arab. J. Sci. Eng.*, vol. 48, no. 2, pp. 2147–2154, 2023, https://doi.org/10.1007/s13369-022-07144-2.

[11] N. C. Basjaruddin, E. Rakhman, Y. Sudarsa, M. B. Z. Asyikin, and S. Permana, “Attendance system with face recognition, body temperature, and use of mask using multi-task cascaded convolutional neural network (MTCNN) Method,” *Green Intelligent Systems and Applications*, vol. 2, no. 2, pp. 71-83, 2024, https://doi.org/10.53623/gisa.v2i2.109.

[12] R. A. Asmara, U. D. Rosiani, M. Mentari, A. R. Syulistyo, M. N. Shoumi, and M. Astiningrum, “An Experimental Study on Deep Learning Technique Implemented on Low Specification OpenMV Cam H7 Device,” *Int. J. Informatics Vis.*, vol. 8, no. 2, pp. 1017–1029, 2024, https://doi.org/10.62527/joiv.8.2.2299.

[13] M. G. F. Odounfa, C. D. S. J. Gbemavo, S. P. G. Tahi, and R. L. Glèlè Kakaï, “Deep learning methods for enhanced stress and pest management in market garden crops: A comprehensive analysis,” *Smart Agric. Technol.*, vol. 9, p. 100521, 2024, https://doi.org/10.1016/j.atech.2024.100521.

[14] M. Chitra, H. Balasubramanian, V. L. Srinivaas and K. Jayanth, "An Adaptation and Evaluation for Object Detection of Small Objects Using YOLO8," *2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC)*, pp. 1-9, 2024, https://doi.org/10.1109/ICEC59683.2024.10837540.

[15] B. Alsharif, E. Alalwany, and M. Ilyas, “Transfer learning with YOLOV8 for real-time recognition system of American Sign Language Alphabet,” *Franklin Open*, vol. 8, p. 100165, 2024, https://doi.org/10.1016/j.fraope.2024.100165.

[16] K. Wangchuk, P. Riyamongkol, and R. Waranusast, “Bhutanese Sign Language Alphabets Recognition Using Convolutional Neural Network,” *InCIT 2020 - 5th Int. Conf. Inf. Technol.*, pp. 44–49, 2020, https://doi.org/10.1109/InCIT50588.2020.9310955.

[17] S. Aiouez, A. Hamitouche, M. Belmadoui, K. Belattar, and F. Souami, “Real-time Arabic Sign Language Recognition based on YOLOv5,” *SCITEPRESS*, vol. 17–25, pp. 17–25, 2022, https://doi.org/10.5220/0010979300003209.

[18] R. A. Alawwad, O. Bchir, and M. M. Ben Ismail, "Arabic Sign Language Recognition using Faster R-CNN," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 3, pp. 692–700, 2021, https://doi.org/10.14569/IJACSA.2021.0120380.

[19] S. T. Abd Al-Latief, S. Yussof, A. Ahmad, S. M. Khadim, and R. A. Abdulhasan, "Instant Sign Language Recognition by WAR Strategy Algorithm Based Tuned Machine Learning," *Int. J. Networked Distrib. Comput.*, vol. 12, no. 2, pp. 344–361, 2024, https://doi.org/10.1007/s44227-024-00039-8.

[20] Y. Saleh and G. F. Issa, "Arabic sign language recognition through deep neural networks fine-tuning," *Int. J. online Biomed. Eng.*, vol. 16, no. 5, pp. 71–83, 2020, https://doi.org/10.3991/ijoe.v16i05.13087.

[21] Y. C. Chiu, C. Y. Tsai, M. Da Ruan, G. Y. Shen, and T. T. Lee, "Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems," *2020 Int. Conf. Syst. Sci. Eng. ICSSE 2020*, pp. 0–4, 2020, https://doi.org/10.1109/ICSSE50014.2020.9219319.

[22] M. Kumar and R. Mann, "Masked Face Recognition using Deep Learning Model," *Proc. - 2021 3rd Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2021*, pp. 428–432, 2021, https://doi.org/10.1109/ICAC3N53548.2021.9725368.

[23] S. D. K, S. K, D. T, S. R, S. S. R. S and T. V. Dixit, "Efficient Object Detection on Low-Resource Devices Using Lightweight MobileNet-SSD," *2025 International Conference on Intelligent Systems and Computational Networks (ICISCN)*, pp. 1-6, 2025, https://doi.org/10.1109/ICISCN64258.2025.10934442.

[24] N. H. Amir, C. Kusuma, and A. Luthfi, "Refining the Performance of Neural Networks with Simple Architectures for Indonesian Sign Language System (SIBI) Letter Recognition Using Keypoint Detection," *Ilk. J. Ilm.*, vol. 17, no. 1, pp. 64–73, 2025, https://doi.org/10.33096/ilkom.v17i1.2522.64-73.

[25] S. Sharma and S. Singh, "Recognition of Indian Sign Language (ISL) Using Deep Learning Model," *Wirel. Pers. Commun.*, vol. 123, no. 1, pp. 671–692, 2022, https://doi.org/10.1007/s11277-021-09152-1.

[26] E. Martinez-Martin and F. Morillas-Espejo, "Deep Learning Techniques for Spanish Sign Language Interpretation," *Comput. Intell. Neurosci.*, vol. 2021, no. i, 2021, https://doi.org/10.1155/2021/5532580.

[27] A. Kannoth, C. Yang and M. A. Guanipa Larice, "Hand Gesture Recognition Using CNN & Publication of World's Largest ASL Database," *2021 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1-6, 2021, https://doi.org/10.1109/ISCC53001.2021.9631255.

[28] D. Liu, S. Gao, W. Chi, and D. Fan, "Pedestrian detection algorithm based on improved SSD," *Int. J. Comput. Appl. Technol.*, vol. 65, no. 1, pp. 25–35, 2021, https://doi.org/10.1504/IJCAT.2021.113643.

[29] H. Wang and X. Wang, "Chinese Sign Language Alphabet Recognition Based on Random Forest Algorithm," pp. 340–344, 2020, https://doi.org/10.1109/MetroInd4.0IoT48571.2020.9138285.

[30] D. Indra, S. Madenda, and E. P. Wibowo, "Feature Extraction of Bisindo Alphabets Using Chain Code Contour," *Int. J. Eng. Technol.*, vol. 9, no. 4, pp. 3415–3419, 2017, https://doi.org/10.21817/ijet/2017/v9i4/170904142.

[31] K. Tong, Y. Wu, and F. Zhou, "Recent advances in small object detection based on deep learning : A review," *Image Vis. Comput.*, vol. 97, p. 103910, 2020, https://doi.org/10.1016/j.imavis.2020.103910.

[32] H. W. Herwanto, A. N. Handayani, K. L. Chandrika, and A. P. Wibawa, "Zoning Feature Extraction for Handwritten Javanese Character Recognition," *ICEEIE 2019 - Int. Conf. Electr. Electron. Inf. Eng. Emerg. Innov. Technol. Sustain. Futur.*, pp. 264–268, 2019, https://doi.org/10.1109/ICEEIE47180.2019.8981462.

[33] A. Mujahid *et al.*, "Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model," *Appl. Sci.*, vol. 11, no. 9, 2021, https://doi.org/10.3390/app11094164.

[34] S. Gharge, A. Patil, S. Patel, V. Shetty and N. Mundhada, "Real-time Object Detection using Haar Cascade Classifier for Robot Cars," *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2023, pp. 64-70, 2023, https://doi.org/10.1109/ICESC57686.2023.10193401.

[35] A. Kloska *et al.*, "Influence of augmentation on the performance of the double ResNet-based model for chest X-ray classification," *Polish J. Radiol.*, vol. 88, pp. e244–e250, 2023, https://doi.org/10.5114/pjr.2023.126717.

[36] B. Li, Y. Hou, and W. Che, "Data augmentation approaches in natural language processing: A survey," *AI Open*, vol. 3, pp. 71–90, 2022, https://doi.org/10.1016/j.aiopen.2022.03.001.

[37] H. Zhao, H. Zhang, and Y. Zhao, "YOLOv7-sea: Object Detection of Maritime UAV Images based on Improved YOLOv7," *Proc. - 2023 IEEE/CVF Winter Conf. Appl. Comput. Vis. Work. WACVW 2023*, pp. 233–238, 2023, https://doi.org/10.1109/WACVW58289.2023.00029.

[38] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Inf.*, vol. 11, no. 2, pp. 1–20, 2020, https://doi.org/10.3390/info11020125.

[39] R. A. Asmara *et al.*, "Chicken meat freshness identification using colors and textures feature," *2018 Jt. 7th Int. Conf. Informatics, Electron. Vis. 2nd Int. Conf. Imaging, Vis. Pattern Recognition, ICIEV-IVPR 2018*, pp. 93–98, 2018, https://doi.org/10.1109/ICIEV.2018.8640992.

[40] A. Juneja, S. Juneja, A. Soneja, and S. Jain, "Real time object detection using CNN based single shot detector model," *J. Inf. Technol. Manag.*, vol. 13, no. 1, pp. 62–80, 2021, https://doi.org/10.22059/jitm.2021.80025.

[41] M. Zaki, "Hand Keypoint-Based CNN for SIBI Sign Language Recognition," *Int. J. Robot. Control Syst.*, vol. 5, no. 2, pp. 813–829, 2025, https://doi.org/10.31763/ijrcs.v5i2.1745.

[42] A. Asokan, D. E. Popescu, J. Anitha, and D. J. Hemanth, "Bat algorithm based non-linear contrast stretching for satellite image enhancement," *Geosci.*, vol. 10, no. 2, pp. 1–12, 2020, https://doi.org/10.3390/geosciences10020078.

[43] L. H. Zhenghao Shi, Yaning Feng, Minghua Zhao, Erhu Zhang, "Normalised gamma transformation-based contrast-limited adaptive histogram equalisation.pdf." p. 747, 2019, https://doi.org/10.1049/iet-ipr.2019.0992.

[44] A. Kumar, R. K. Jha, and N. K. Nishchal, “An improved Gamma correction model for image dehazing in a multi-exposure fusion framework,” *J. Vis. Commun. Image Represent.*, vol. 78, p. 103122, 2021, https://doi.org/10.1016/j.jvcir.2021.103122.

[45] M. Al-Faris, J. Chiverton, D. Ndzi, and A. I. Ahmed, “A review on computer vision-based methods for human action recognition,” *J. Imaging*, vol. 6, no. 6, 2020, https://doi.org/10.3390/jimaging6060046.

[46] A. M. Sarosa, A. Badriyah, R. A. Asmara, M. K. Wardani, D. F. Al Riza, and Y. Mulyani, “Performance Analysis of MobileNET on Grape Leaf Disease Detection,” *2024 Int. Conf. Adv. Inf. Sci. Dev. (ICAISD). IEEE*, vol. 64–68, 2024, https://doi.org/10.1109/ICAISD63055.2024.10894920.

[47] G R. Andrie Asmara, M. Ridwan and G. Budiprasetyo, "Haar Cascade and Convolutional Neural Network Face Detection in Client-Side for Cloud Computing Face Recognition," *2021 International Conference on Electrical and Information Technology (IEIT)*, pp. 1-5, 2021, https://doi.org/10.1109/IEIT53149.2021.9587388.

[48] I. P. Sari, “Closer Look at Image Classification for Indonesian Sign Language with Few-Shot Learning Using Matching Network Approach,” *Int. J. Informatics Vis.*, vol. 7, no. 3, pp. 638–643, 2023, https://doi.org/10.30630/joiv.7.3.1320.

[49] R. A. Asmara *et al.*, “YOLO-based object detection performance evaluation for automatic target aimbot in first-person shooter games,” *Bull. Electr. Eng. Informatics*, vol. 13, no. 4, pp. 2456–2470, 2024, https://doi.org/10.11591/eei.v13i4.6895.

[50] R. A. Asmara, B. Syahputro, D. Supriyanto, and A. N. Handayani, “Prediction of Traffic Density using YOLO Object Detection and Implemented in Raspberry Pi 3b + and Intel NCS 2,” *4th Int. Conf. Vocat. Educ. Training, ICOVET 2020*, pp. 391–395, 2020, https://doi.org/10.1109/ICOVET50258.2020.9230145.

AUTHOR BIOGRAPHY

Lilis Nur Hayati is Master's degree in Information Technology was obtained in 2005 from Universitas Gadjah Mada, and the Doctoral degree in the same field is currently being pursued at Universitas Negeri Malang. Currently, she is a lecturer in Information Systems with 9 years of teaching experience at Universitas Muslim Indonesia. Her research interests include software design, software requirements analysis, decision support systems, technopreneurship, human-computer interaction, e-business concepts, research methodology, and operating systems.
Email:lilis.nurhayati.2205349@students.um.ac.id
Google Scholar: https://scholar.google.com/citations?hl=id&user=me_9y28AAAAJ

Anik Nur Handayani is Master's degree in Electrical Engineering in 2008 from Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia, and earned her Doctoral degree in Science and Advanced Engineering from Saga University, Japan. She is currently a university lecturer at Universitas Negeri Malang, Indonesia. Her research interests include image processing, biomedical signal analysis, artificial intelligence, machine learning, deep learning, computer vision, and assistive technologies.
Email: aniknur.ft@um.ac.id
Google Scholar: https://scholar.google.com/citations?hl=en&user=nqPHjbMAAAAJ&view_op=list_works&sortby=pubdate

Wahyu Sakti Gunawan Irianto is Master’s degree in Computer Science from Universitas Indonesia, Jakarta, in 1997. He earned a Doctoral degree in Computer Science (M.Kom). He is currently a senior lecturer in the Department of Electrical Engineering at Universitas Negeri Malang, Indonesia. His research interests include computer science education, educational technology, intelligent systems, embedded and microcontroller applications, and digital systems. He has contributed to various projects, such as an interactive learning module based on Arduino and multimodal dataset research in the LUMINA project.
Email: wahyu.sakti.ft@um.ac.id
Google Scholar: https://scholar.google.com/citations?user=DAWTUlAAAAAJ&hl=en

Rosa Andrie Asmara is received his Bachelor's degree in Electronics Engineering from Universitas Brawijaya, Malang, in 2004. He obtained his Master's degree in Computer Science from Institut Teknologi Sepuluh Nopember, Surabaya, in 2009, and completed his Doctoral degree in Computer Science at Saga University, Japan, in 2013. He is currently a lecturer at Politeknik Negeri Malang, Indonesia. His research interests include machine learning, image understanding, and computer vision.
Email: rosa_andrie@polinema.ac.id
Google Scholar: https://scholar.google.co.id/citations?user=A1592kEAAAAJ&hl=en

Dolly Indra is earned his Doctoral degree in Information Technology from Universitas Gunadarma in 2017. He is currently a lecturer at the Faculty of Computer Science, Universitas Muslim Indonesia. His research interests include image processing, computer vision, microcontroller systems, and information systems.
Email: dolly.indra@umi.ac.id
Google Scholar: https://scholar.google.co.id/citations?user=94_nu_QAAAAJ&hl=en

Nor Salwa Damanhuri is received her Bachelor of Science (Hons.) in Electrical and Electronics Engineering from Universiti Tenaga Nasional (UNITEN), Malaysia, in March 2002. She completed her Master of Science in Control Systems Engineering at The University of Sheffield, United Kingdom, in September 2005, and obtained her Doctor of Philosophy (Ph.D.) in

Bioengineering from the University of Canterbury, New Zealand, in April 2015. She is currently an Associate Professor at the Centre for Electrical Engineering Studies, Universiti Teknologi MARA (UiTM), Penang Branch, Malaysia. Her research interests include biomedical engineering, digital signal processing, mathematical modeling, control systems, and solar PV system applications.
Email: norsalwa071@uitm.edu.my
Google Scholar: https://scholar.google.com/citations?user=O3DojDMAAAAJ&hl=en