Terbit *online* pada laman : http://teknosi.fti.unand.ac.id/

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |

Research Article

Comparison of FA-SSD: Performance of Feature Extractors VGG-19, MobileNetV3, and ResNet-152 for Human Body Temperature Prediction

Ridho Sholehurrohman[a,*], *Akmal Junaidi*[a], *Hamzah Hanif*[a], *Favorisen Rosyking Lumbanraja*[a], *Muhammad Reza Habibi*[b]

[a]*Department of Computer Science, Lampung University, Bandar Lampung, Indonesia*
[b]*Department of Business Statistics, Faculty of Vocational Studies, ITS, Surabaya, Indonesia*
[b]*Department of Computer Science and Electrical Engineering, Kumamoto University, Japan*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 02 Juli 2026
Revisi Akhir: 20 Agustus 2026
Diterbitkan *Online*: 28 Agustus 2026

KATA KUNCI

FA-SSD,
VGG-19,
MobileNetV3,
ResNet-152,
Body Temperature Prediction

KORESPONDENSI

E-mail: ridho.sholehurrohman@fmipa.unila.ac.id*

A B S T R A C T

Body temperature is a critical health indicator, and non-contact thermal imaging-based prediction systems are essential for early detection of infectious diseases. This study compares the performance of three FA-SSD (Feature Fusion and Spatial Attention-Based Single Shot Detector) models with different feature extractors—VGG-19, MobileNetV3, and ResNet-152—for face detection and body temperature prediction using thermal images. A subset of 520 thermal facial images from the Comprehensive Facial Thermal Dataset was used, with 80% for training, 10% for validation, and 10% for testing. Model performance was evaluated using Generalized Intersection over Union (GIoU) for detection accuracy, and Mean Absolute Error (MAE) with Mean Absolute Percentage Error (MAPE) for temperature prediction accuracy. The results showed that no single model excelled in all metrics. ResNet-152 achieved the highest average GIoU, indicating superior object detection performance. VGG-19 delivered the lowest average MAE of 0.451°C and MAPE of 1.278%, making it the best for temperature prediction. MobileNetV3 achieved the lowest minimum MAE of 0.000050°C but showed higher average errors and validation fluctuations. In conclusion, VGG-19 is recommended for clinical temperature accuracy, ResNet-152 for robust detection, and MobileNetV3 for edge deployment.

1. INTRODUCTION

Human body temperature is a vital indicator that directly reflects an individual's health and is commonly used for the early detection of various diseases, including the early symptoms of infections such as influenza and COVID-19 [1], [2]. In various public facilities, such as airports, hospitals, and offices, rapid and accurate body temperature monitoring is crucial for controlling the spread of infectious diseases [3]. However, traditional contact-based thermometer methods have limitations, including longer measurement times and physical interaction, which may increase the risk of transmission [2], [4], [5].

The development of artificial intelligence (AI) technology plays a significant role in addressing medical issues [6], [7], particularly in computer vision, which enables machines to interpret and analyze visual information [8]. One of the models used in computer vision for object detection is the Single Shot Multibox Detector (SSD), known for its speed and effectiveness in detecting objects in images and videos [9], [10]. SSD has been further developed into the Feature Fusion and Spatial Attention-Based Single Shot Detector (FA-SSD), which combines feature fusion and spatial attention to improve object detection accuracy [11]. FA-SSD enhances standard SSD by integrating multi-scale feature fusion and spatial attention mechanisms, which are particularly beneficial for capturing fine-grained thermal patterns on facial surfaces.

While FA-SSD has demonstrated strong performance in general object detection tasks using RGB images [11], [12], its application to thermal-based body temperature prediction

https://doi.org/10.25077/TEKNOSI.v12i2.2026.300-311

remains largely unexplored. To the best of our knowledge, no previous study has systematically compared the impact of different backbone architectures—specifically VGG-19, MobileNet-V3, and ResNet-152—on the dual tasks of face detection and temperature regression from thermal imagery. Furthermore, existing research on thermal face analysis has primarily focused on face detection or recognition rather than continuous temperature value prediction. This gap is significant because accurate body temperature prediction requires not only precise facial localization but also reliable extraction of thermal intensity information from the detected region.

Each feature extractor has unique characteristics and advantages that may influence performance in thermal imaging tasks. VGG-19 is known for its deep architecture [13] and its effectiveness in transfer learning due to its 19 layers [14]. MobileNetV3 excels in computational efficiency on mobile devices [15]-[18], while ResNet-152 has advantages in mitigating the vanishing gradient problem in very deep networks [19],[20],[21]. However, how these architectures affect the trade-off between detection precision and temperature regression accuracy in thermal images has not been systematically investigated.

The objective of this study is to compare the performance of the FA-SSD model with feature extractors VGG-19, MobileNetV3, and ResNet-152 in detecting faces and predicting human body temperature based on thermal images, as well as to evaluate the accuracy and prediction errors of each model using the metrics Generalized Intersection over Union (GIoU), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

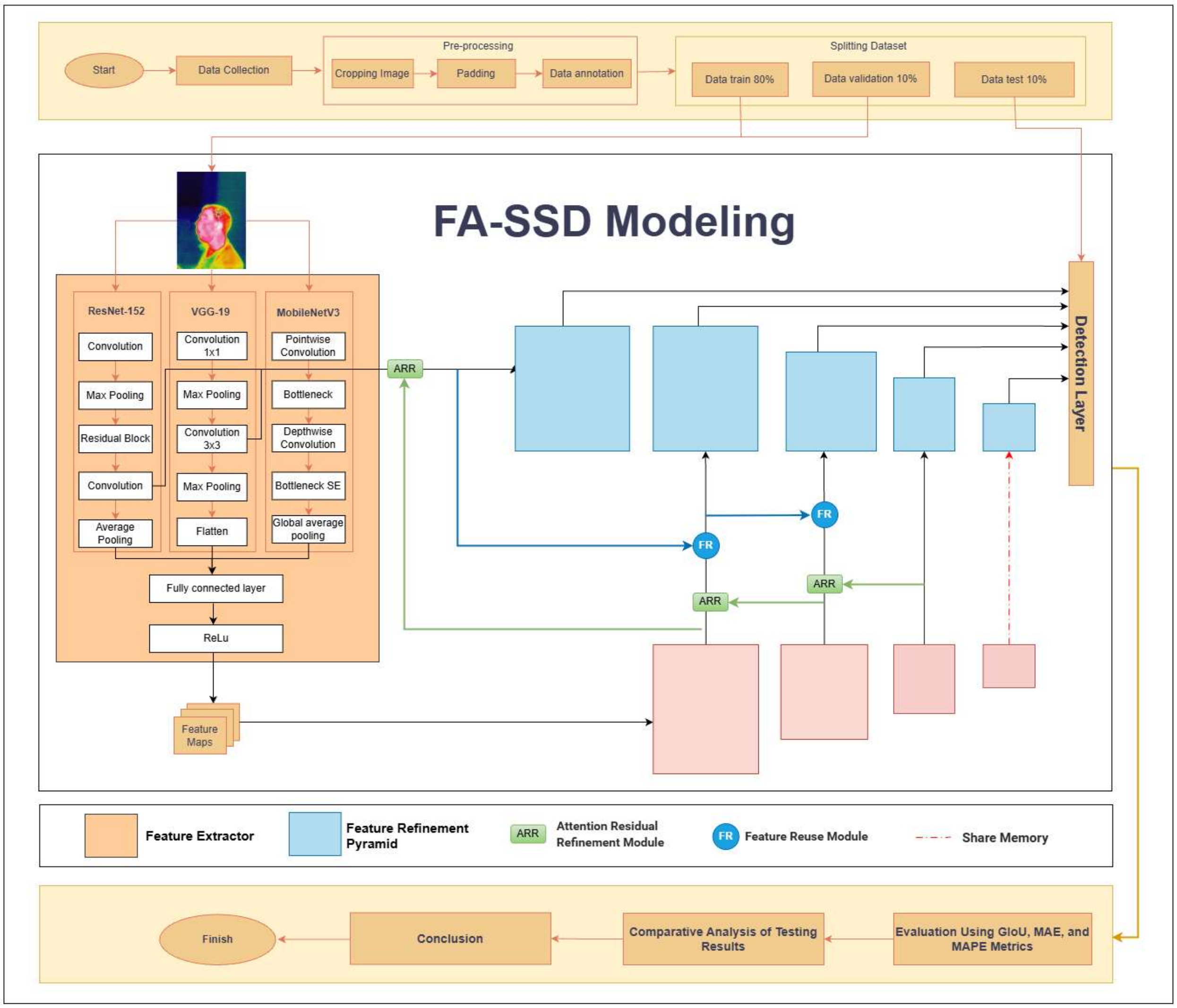


Figure 1. Research method

2. METHOD

The following describes the methodology used in this research, as shown in Figure 1. The approach encompasses the essential steps, from dataset collection to model training, to ensure the effectiveness and accuracy of human body temperature prediction based on thermal images. Each stage of the process has been carefully planned to optimize data quality and enhance the model's performance. The following subsections outline the detailed steps involved in data preprocessing, dataset splitting, and training the FA-SSD model.

Figure 1 illustrates the research methodology for developing a temperature detection model based on image data, using the FA-SSD (Feature Attention Single Shot Detector) method. In general, the research process begins with data collection, followed by data preprocessing, dataset splitting, and modeling with FA-SSD across various feature-extractor architectures. The model is then tested using the test data and evaluated using GIoU, MAE, and MAPE metrics. The evaluation results are compared to analyze the performance of the three feature extractors. Finally, a comparison of the results is made, leading to the research conclusion.

2.1. Dataset Collection

The primary material used is the Comprehensive Facial Thermal Dataset released by Abuhussein et al,. [22]. This dataset consists of 2,250 human facial thermal images captured using the UNI-T UTi165A camera, which is used for thermal-based face detection and recognition. The data collection process was conducted directly using the UNI-T UTi165A thermal camera, which is specifically designed for face detection, face recognition, and emotion expression analysis. The selected dataset provides a diverse representation of facial expressions and environmental variables, making it suitable for predicting body temperature under varying conditions. In this study, a subset of the dataset was used, containing 520 images with a rainbow color palette, which includes variations in emotional expressions and environmental conditions, as well as a temperature range from 0°C to 50°C. The selected subset includes only images with clear facial visibility and minimal occlusion to ensure reliable annotation and temperature extraction. The resolution of the thermal images produced is 240×320 pixels. This dataset is publicly available through Mendeley Data at https://data.mendeley.com/datasets/8885sc9p4z/1. The dataset has been anonymized and is publicly available for research purposes. No additional ethical approval was required as this study uses only pre-existing, de-identified thermal images.

2.2. Preprocessing Data

In this study, several preprocessing steps were performed to prepare the dataset for analysis and ensure high-quality data for model training. The first step was to crop the thermal images to remove unnecessary elements, such as digital temperature panels and other information located at the top and right sides. The original image size of 240 × 320 pixels was cropped by 38 pixels from the top and 43 pixels from the right, resulting in a final image size of 197 × 282 pixels. This cropping process is essential for eliminating visual noise from measurement instruments, ensuring the model focuses solely on the relevant target object.

After cropping, padding was applied to standardize the image size. Proportional padding was added to the left, right, top, and bottom, bringing the final image size to 512 × 512 pixels. The target size of 512 × 512 pixels was selected to meet the input requirements of the FA-SSD backbone networks while preserving sufficient spatial resolution for face detection and temperature extraction. This ensures that the proportions and details of the main object are preserved without distortion, while centering the important areas of the image. No data augmentation was applied to maintain consistency with real-world deployment scenarios.

The next step involved annotating the images to delineate the target areas for model training. LabelImg, a GUI-based annotation tool, was employed to manually label the facial regions with bounding boxes and the label "person." The annotations were saved in the Pascal VOC (XML) format, which is fully compatible with the FA-SSD architecture implemented in this study. A total of 520 images were meticulously annotated, with each annotation file stored in XML format, ensuring that filenames corresponded to their respective image files. This annotation process is essential to ensure the model is trained to focus precisely on the facial region, which is crucial for accurate body temperature prediction.

Figure 2. Illustration of the Data Annotation Process with Facial Bounding Box

2.3. Model Training and Hyperparameter Tuning

Training was conducted using Visual Studio Code, with a virtual environment managed through Anaconda. The entire model development and training process leveraged CUDA acceleration on the NVIDIA GeForce GTX 1660 Ti GPU with Max-Q Design. A random seed of 42 was set for dataset splitting and model initialization to ensure reproducibility. The model used in this study is the Feature Fusion and Spatial Attention-Based Single Shot Detector (FA-SSD), integrated with three feature extractors: VGG-19, MobileNetV3, and ResNet-152.

The dataset was randomly divided into three parts: 80% for training, 10% for validation, and 10% for testing. This division allows for efficient model training while ensuring that sufficient data is reserved for both validation and testing. The 80% training data ensures the model has enough examples to learn effectively, while the 10% validation data helps fine-tune the model and prevent overfitting. The remaining 10% testing data is used for final performance evaluation, providing an unbiased assessment of the model's generalization ability.

The model was optimized using a combined loss function: GIoU loss for bounding box regression and Mean Squared Error (MSE) loss for temperature prediction. To optimize model performance during training, hyperparameter tuning was carried out, including adjustments to the number of epochs, batch size, learning rate, dropout rate, and optimizer type. The optimal combination of hyperparameters was determined using EarlyStopping, which halts training automatically if validation performance does not improve after a set number of epochs (patience = 10), and ReduceLROnPlateau, which adaptively adjusts the learning rate with a reduction factor of 0.5 and patience of 3 when the model's performance stagnates. These strategies were essential for preventing overfitting and accelerating the model's convergence.

 https://doi.org/10.25077/TEKNOSI.v12i2.2026.300-311

The hyperparameters used for all three models are summarized in Table 1. These settings were consistently applied across all models for fair comparison.

Table 1. Hyperparameter Tuning for FA-SSD Models

Parameter	Value
Epoch	40
Batch size	16
Optimizer	Adam
Learning rate	0.0001
EarlyStopping	Patience = 10
ReduceLROnPlateau	Reduction factor = 0.5, Patience = 3
Dropout	0.5

2.4. *Model Evaluation*

The performance of the trained models was evaluated using several widely accepted metrics, including Generalized Intersection over Union (GIoU), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics were chosen to assess both the accuracy of object detection and the precision of body temperature predictions.

2.4.1. *Generalized Intersection over Union (GIoU)*

Generalized Intersection over Union (GIoU) is an extension of the Intersection over Union (IoU), designed to overcome the limitations of IoU when the two bounding boxes do not overlap. While IoU focuses only on the overlapping area between two bounding boxes, GIoU considers the space outside the boxes, providing additional information about the relative positioning and distance between the two objects [23],[24],[25]. This makes GIoU a more comprehensive metric for evaluating object detection performance, especially when objects are not overlapping. The formula for GIoU is defined as [23]:

$$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|} \quad (1)$$

where $\boldsymbol{IoU}$ is the traditional Intersection over Union $\frac{|A \cap B|}{|A \cup B|}$, $\boldsymbol{A}$ is the ground truth bounding box, then $\boldsymbol{B}$ is the predicted bounding box, and $\boldsymbol{C}$ is the smallest enclosing bounding box that can cover both $\boldsymbol{A}$ and $\boldsymbol{B}$ [23].

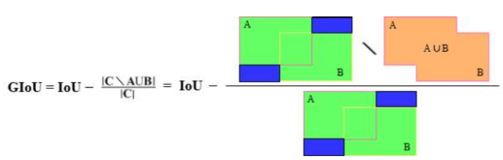


Figure 2. Illustration of GIoU Calculation

Figure 3 illustrates the GIoU calculation, in which C denotes the smallest enclosing box. This adjustment enables GIoU to penalize inaccurate predictions, even when there is no overlap between the predicted box and the ground truth box, thereby measuring how far the predicted box is from the correct position

2.4.2. *Mean Absolute Error (MAE)*

Mean Absolute Error (MAE) is an evaluation metric used to assess how well a model's predictions align with the actual data or observations. MAE is calculated as the average of the absolute errors between predicted and actual values [26]. A lower MAE value indicates better model performance, as it suggests that the difference between predictions and actual values is smaller. The formula for MAE is [27]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i' - y_i| \quad (2)$$

where $\boldsymbol{y_i'}$ is the predicted value, then $\boldsymbol{y_i}$ is the actual value, and $\boldsymbol{n}$ is the number of data points.

2.4.3. *Mean Absolute Percentage Error (MAPE)*

Mean Absolute Percentage Error (MAPE) is a statistical metric used to evaluate the accuracy of predictions in forecasting models. It calculates the average absolute percentage error between predicted values and actual values [28]. MAPE is particularly useful when comparing model performance across different datasets or scales. The formula for MAPE is [26]:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i'}{Y_i}\right| \times 100 \quad (3)$$

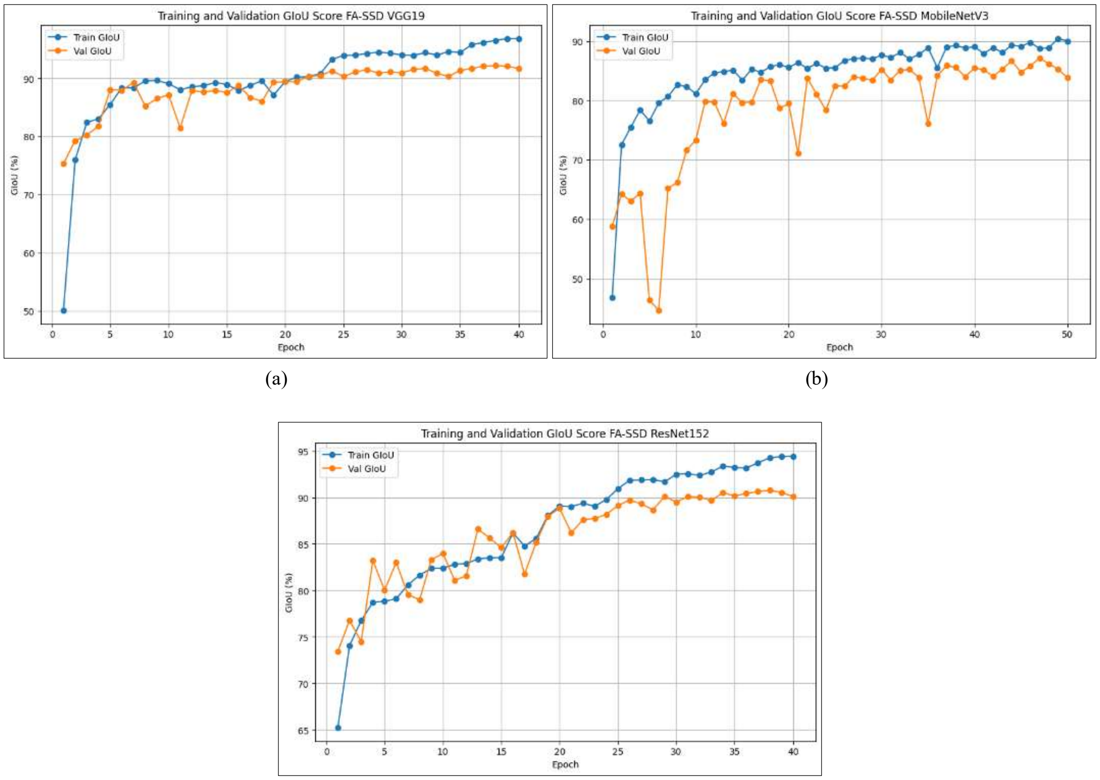
where $\boldsymbol{Y_i}$ is the actual value for the $\boldsymbol{i}$-$\boldsymbol{th}$ sample, then $\boldsymbol{Y_i'}$ is the predicted value for the $\boldsymbol{i}$-$\boldsymbol{th}$ sample, and $\boldsymbol{n}$ is the total number of data points.

3. RESULT AND DISCUSSION

This section presents and discusses the findings of our experiments, evaluating the performance of the proposed FA-SSD models with feature extractors VGG-19, MobileNetV3, and ResNet-152 using the metrics described in the previous section. The results are analyzed to compare the effectiveness of each backbone architecture and to identify the model that provides the best overall detection performance.

3.1. *Training Performance Analysis*

The performance of the models was evaluated during training using Generalized Intersection over Union (GIoU), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

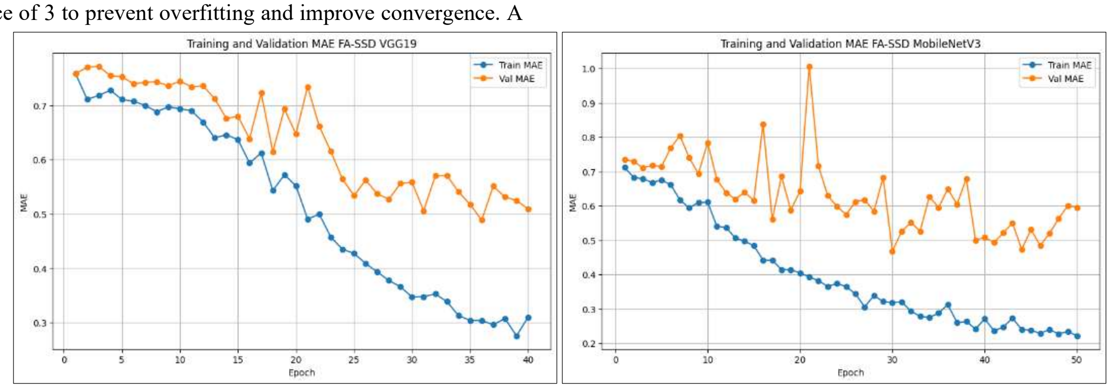


(a) (b)

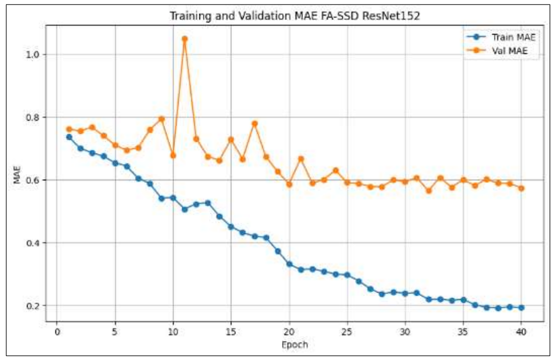
(c)

Figure 4. GIoU Training Graphs of FA-SSD: (a) VGG-19, (b) MobileNetV3, and (c) ResNet-152

Figure 4 shows the GIoU training graphs for FA-SSD: (a) VGG-19, (b) MobileNetV3, and (c) ResNet-152. In the early epochs, GIoU increased sharply for all models. VGG-19 reached nearly 97% by the end of training, while MobileNetV3 stabilized between 85% and 90%, and ResNet-152 achieved a steady range of 92% to 95%. The training process was conducted for 40 epochs with a batch size of 16, using the Adam optimizer with a learning rate of 0.0001. EarlyStopping was applied with a patience value of 10, and ReduceLROnPlateau with a reduction factor of 0.5 and patience of 3 to prevent overfitting and improve convergence. A dropout rate of 0.5 was used to enhance generalization, particularly for VGG-19. For validation, VGG-19 maintained a stable GIoU of 90-92%, while MobileNetV3 showed more fluctuations (80-87%), and ResNet-152 remained stable at around 90%. These results suggest that VGG-19 achieved the best balance between training and validation, while MobileNetV3 showed greater validation instability, possibly due to its smaller model size and different optimization behavior.



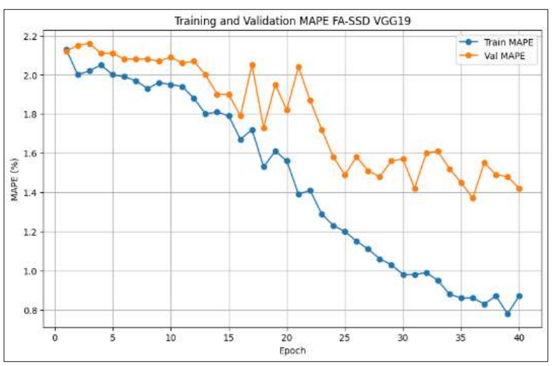
(a) (b)



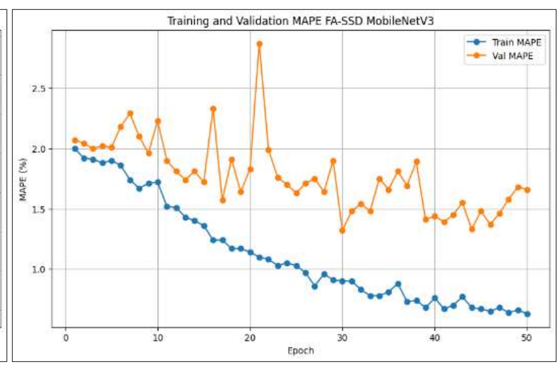
(c)

Figure 5. MAE Training Graphs of FA-SSD: (a) VGG-19, (b) MobileNetV3, and (c) ResNet-152.

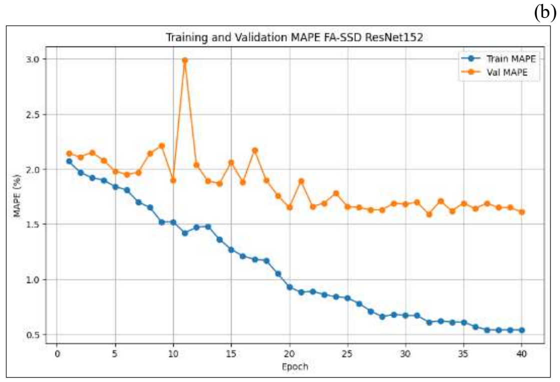
Figure 5 illustrates the MAE training graphs for the models: (a) VGG-19, (b) MobileNetV3, and (c) ResNet-152. The training MAE consistently decreased across all models, indicating that the models improved at predicting body temperature over time. VGG-19 showed the most consistent decrease, reflecting strong performance in temperature prediction. The training hyperparameters—such as a batch size of 16, a learning rate of 0.0001, and a dropout rate of 0.5—helped prevent overfitting and enabled efficient learning. Similarly, ResNet-152 showed a gradual decrease in MAE, indicating improvement in prediction accuracy. However, MobileNetV3 showed some fluctuations in the validation MAE, suggesting instability, possibly due to its lightweight architecture, which is more sensitive to variations in thermal features.



(a)



(b)



(c)

Figure 6. MAPE Training Graphs of FA-SSD: (a) VGG-19, (b) MobileNetV3, and (c) ResNet-152.

Figure 6 presents the MAPE values for the models: (a) VGG-19, (b) MobileNetV3, and (c) ResNet-152. The training MAPE for all models decreased consistently, indicating that the models' relative prediction error diminished over time. VGG-19 had the lowest MAPE, indicating superior prediction accuracy. The use of the Adam optimizer, learning rate of 0.0001, and dropout rate of 0.5 contributed to stable learning and superior performance. ResNet-152 showed consistent improvement in MAPE, with only slight fluctuations, indicating good model performance. However, MobileNetV3 exhibited larger spikes in validation MAPE, reflecting instability in its predictions. This fluctuation could be attributed to the smaller batch size of 16 and the dropout rate of 0.5, which may have caused the model to struggle with generalizing to the validation data.

3.2. *FA-SSD Evaluation*

Model evaluation was conducted on two main aspects: object detection (bounding box) and body temperature prediction. The metric used for object detection evaluation is the Generalized Intersection over Union (GIoU). At the same time, temperature prediction accuracy was assessed using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Prediction results were automatically collected for all test images and analyzed descriptively to compare the performance of the three feature extractors tested. This series of research steps aims to identify the optimal FA-SSD model with a feature extractor for object detection and body temperature prediction from thermal images.

3.2.1. Object Detection Evaluation

In this stage, model performance was evaluated in terms of object detection by measuring the Generalized Intersection over Union (GIoU). The GIoU metric is used to assess the similarity between the model's predicted bounding box and the ground truth in the test data. The higher the GIoU value, the better the model is at precisely detecting the object's location. In the object detection process, the model initializes a predicted bounding box as an initial estimate of the object's position in the image. This bounding box is then adjusted to align more closely with the actual position (ground truth). GIoU is used to measure how well the predicted bounding box matches the ground truth by comparing the intersection area of the two regions to their union area. In this evaluation, the average GIoU values for each model—FA-SSD VGG-19, FA-SSD MobileNetV3, and FA-SSD ResNet-152—were analyzed. Figure 7 below presents a barplot visualization of the average performance of the three models.

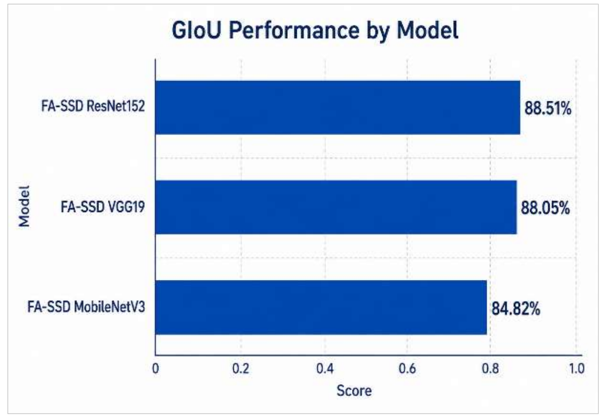


Figure 7. Comparison of Average GIoU Performance Model

Based on Figure 7, FA-SSD with the ResNet-152 extractor demonstrated superior performance in detecting facial objects on the dataset used in this study.

3.2.2. Temperature Prediction Evaluation

The model's performance in predicting temperature was evaluated by comparing its predictions with the actual temperature in the test data. To calculate the model's performance in temperature prediction, the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics were used. The MAE value indicates the average absolute difference between the predicted and actual temperatures. Meanwhile, MAPE measures the average percentage error in predictions.
The evaluation results on the test data provide an overview of how accurately the model predicts temperature. Table 2 below presents the statistical results of prediction differences for each model used in this study.

Table 2. Temperature Prediction Results of Different Models

Model		FA-SSD VGG-19	FA-SSD MobileNetV3	FA-SSD ResNet-152
MAE (°C)	Min	0.009430	0.000050	0.017208
	Max	2.348637	2.154587	2.218922
	Average	0.451361	0.472088	0.545047
MAPE (%)	Min	0.027333	0.000142	0.050464
	Max	6.364869	6.489719	6.683499
	Average	1.277950	1.343460	1.546691

Based on Table 2, the performance results for the three FA-SSD models are presented using the MAE and MAPE metrics. The FA-SSD VGG-19 model showed the best overall temperature prediction performance. This model achieved the lowest average error, with an MAE of 0.451361°C and a MAPE of 1.277950%. The minimum MAE achieved was 0.009430, and the maximum was 2.348637. These results indicate that FA-SSD VGG-19 achieves the highest temperature prediction accuracy among the three models.

Next, the FA-SSD MobileNetV3 model ranked second. This model recorded an average MAE of 0.472088°C and an average MAPE of 1.343460%. Interestingly, this model achieved the lowest MAE of 0.000050°C, although its overall average error was slightly higher than that of VGG-19.

Meanwhile, the FA-SSD ResNet-152 model showed the lowest temperature prediction accuracy in this comparison. This model produced an average MAE of 0.545047°C and an average MAPE of 1.546691%. Based on these results, it can be concluded that FA-SSD ResNet-152 has the largest average temperature prediction error among the three models tested. Overall, no single model excelled in all metrics, so model selection should be tailored to the application requirements, whether prioritizing object detection localization precision or body temperature estimation accuracy.

3.3. *FA-SSD Testing Results*

After conducting the quantitative evaluation of the three models in Section 3.2, this section presents visual examples of prediction results from each model on the test data. Figure 8 shows sample predictions from FA-SSD VGG-19, FA-SSD MobileNetV3, and

https://doi.org/10.25077/TEKNOSI.v12i2.2026.300-311

FA-SSD ResNet-152 along with the metric values generated for each prediction. It should be noted that the values shown in Figure 8 are the results of single prediction samples, not overall average values.

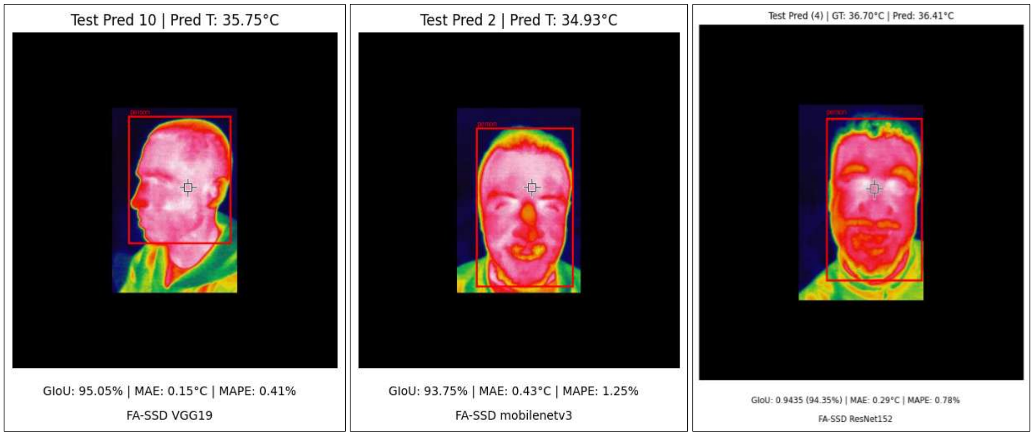


Figure 8. FA-SSD Testing Results: VGG-19, MobileNetV3, and ResNet-152.

In Figure 8, the FA-SSD VGG-19 model demonstrated excellent performance on a single test sample. This model achieved a GIoU value of 95.05%, indicating precise object detection on that sample. For temperature prediction, the model predicted 35.75°C with an MAE of 0.15°C and a MAPE of 0.41%. These results are consistent with the evaluation in Section 3.2, which showed that FA-SSD VGG-19 has the lowest average MAE and MAPE among the three models.

Next, the FA-SSD MobileNetV3 model achieved a GIoU score of 93.75% on the test data. The predicted temperature was 34.93°C with an MAE of 0.43°C and a MAPE of 1.25%. These results align with the findings in Section 3.2, which indicate that MobileNetV3 has an average error slightly higher than that of VGG-19.

Furthermore, the FA-SSD ResNet-152 model demonstrated solid performance. With a Ground Truth temperature of 36.70°C, the model predicted 36.41°C, yielding a GIoU of 94.35%, an MAE of 0.29°C, and a MAPE of 0.78%. Although the GIoU value in this sample (94.35%) is slightly lower than that of VGG-19 (95.05%), it should be recalled that, based on the evaluation in Section 3.2, the ResNet-152 model has the highest average GIoU among the three models for overall object detection.

Overall, the visual examples in Figure 8 reinforce the evaluation findings in Section 3.2. FA-SSD VGG-19 excels in temperature prediction (lowest average MAE and MAPE), while FA-SSD ResNet-152 excels in object detection (highest average GIoU). A-SSD MobileNetV3 offers a balance with lower computational complexity, as reflected in its competitive average MAE and MAPE despite the lightest architecture

3.4. Analysis and Discussion

The experimental results reveal an interesting trade-off between object detection accuracy and temperature prediction precision across the three feature extractors. While ResNet-152 achieves the highest average GIoU score, it does not yield the most accurate temperature predictions. Conversely, VGG-19 delivers the lowest MAE and MAPE despite having a slightly lower average GIoU than ResNet-152. This finding suggests that precise bounding box localization does not necessarily translate into better temperature estimation, likely because temperature prediction depends more on the consistency of pixel intensities within the detected region than on the exact boundary alignment. The superior detection performance of ResNet-152 can be attributed to its deep residual architecture, which enables better feature extraction and gradient flow during training. However, the same architecture may introduce effects that slightly degrade the model's ability to capture fine-grained temperature variations, potentially due to the loss of high-frequency thermal information through deep residual connections. In contrast, VGG-19, with its more straightforward architecture, appears to preserve better the thermal intensity information needed for accurate temperature regression.

MobileNetV3 presents an interesting case. Despite having the lightest architecture and showing more validation fluctuations during training, it achieved the lowest minimum MAE among all models. This indicates that MobileNetV3 achieves excellent predictions on certain samples but struggles to maintain consistency across the entire dataset. This behavior is typical for lightweight models that trade off some generalization capacity for computational efficiency.

From a practical perspective, the choice of model should be guided by the specific application requirements. For fever screening in clinical settings where temperature accuracy is paramount, VGG-19 is the most appropriate choice despite its higher computational cost. For applications that require robust facial localization before temperature extraction, such as multi-person screening in crowded environments, ResNet-152 offers higher detection reliability. For deployment on edge devices with limited computational resources, MobileNetV3 provides a viable balance between accuracy and efficiency.

For future work, ensemble strategies that combine ResNet-152 for detection with VGG-19 for temperature prediction could be explored to leverage the strengths of both architectures. Additionally, testing on more diverse thermal datasets collected under varying environmental conditions would help validate the generalizability of these findings. Furthermore, optimization techniques such as quantization or pruning could be applied to

MobileNetV3 to enhance its deployment feasibility on edge devices while maintaining prediction accuracy.

4. CONCLUSION

This study evaluated the performance of three FA-SSD models with different feature extractors—VGG-19, MobileNetV3, and ResNet-152—for object detection and body temperature prediction from thermal images, using the GIoU, MAE, and MAPE metrics. The experimental results showed that no single model excelled in all evaluation metrics. FA-SSD ResNet-152 achieved the highest average GIoU, indicating superior object detection performance due to its deep residual architecture. However, this model produced the highest average temperature prediction errors with an MAE of 0.545°C and an MAPE of 1.547%. In contrast, FA-SSD VGG-19 delivered the most accurate temperature predictions, achieving the lowest average MAE of 0.451°C and MAPE of 1.278%, despite having a slightly lower average GIoU than ResNet-152. FA-SSD MobileNetV3, as the lightest architecture, showed competitive performance with an average MAE of 0.472°C and MAPE of 1.343%, and notably achieved the lowest minimum MAE of 0.000050°C among all models.

From a practical perspective, the choice of model should be guided by specific application requirements. For clinical fever screening where temperature accuracy is critical, FA-SSD VGG-19 is the most suitable choice. For applications that require robust facial localization before temperature extraction, such as multi-person screening in crowded environments, FA-SSD ResNet-152 offers higher detection reliability. For deployment on resource-constrained devices such as mobile or embedded systems, FA-SSD MobileNetV3 strikes a viable balance between accuracy and computational efficiency.

Several limitations of this study should be acknowledged, including the characteristics of the dataset and the controlled testing environments. For future work, ensemble strategies that combine ResNet-152 for detection with VGG-19 for temperature prediction could be explored to leverage the strengths of both architectures. Additionally, testing on larger and more diverse thermal datasets and applying optimization techniques such as quantization or pruning to MobileNetV3 would further enhance its deployment feasibility on edge devices.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to all individuals and institutions that have contributed to this research. Special thanks to the Department of Computer Science at Universitas Lampung for providing resources, facilities, and support throughout the study. We also acknowledge Abuhussein et al. (2024) for releasing the Comprehensive Facial Thermal Dataset, which provided the invaluable thermal images for our analysis. Additionally, we appreciate the guidance and feedback from our colleagues, whose insights significantly enhanced the quality of this work.

REFERENCES

[1] N. I. D. Zaman, Y. W. Hau, M. C. Leong, et al., "A review on the significance of body temperature interpretation for early infectious disease diagnosis," Artificial Intelligence Review, vol. 56, pp. 15449–15494, 2023, doi: 10.1007/s10462-023-10528-x..

[2] S. Mirjalali, S. Peng, Z. Fang, C.-H. Wang, S. Wu, "Wearable sensors for remote health monitoring: Potential applications for early diagnosis of Covid-19," Advanced Materials Technologies, vol. 7, no. 1, p. 2100545, 2022, doi: 10.1002/admt.202100545.

[3] X. Rong, L. Yang, H. Chu, M. Fan, "Effect of delay in diagnosis on transmission of COVID-19," Mathematical Biosciences and Engineering, vol. 17, no. 3, pp. 2725–2740, 2020, doi: 10.3934/mbe.2020149.

[4] L. Evangelisti, E. De Cristo, S. Monteleone, C. Guattari, P. Gori, I. Pini, T. de Rubeis, D. Ambrosini, "A New Customized Measurement System for a Non-Contact, Enhanced Thermometric Method," Energies, vol. 18, p. 1537, 2025, doi: 10.3390/en18061537.

[5] K. C. Lee, Y. P. Lin, Y. L. Tzeng, et al., "The impact of age on comparative diagnostic accuracy of temporal artery thermometers and non-contact infrared thermometers for fever detection: a systematic review and meta-analysis," BMC Infectious Diseases, vol. 25, p. 31, 2025, doi: 10.1186/s12879-024-10332-0.

[6] R. Hirani, et al., "Artificial Intelligence and Healthcare: A Journey through History, Present Innovations, and Future Possibilities," Life, vol. 14, no. 557, pp. 1–25, 2024, doi:10.3390/life14050557.

[7] A. Al Kuwaiti, et al., "A Review of the Role of Artificial Intelligence in Healthcare," Journal of Personalized Medicine, vol. 13, p. 951, 2023, doi:10.3390/jpm13060951.

[8] A. A. Khan, A. A. Laghari, S. A. Awan, "Machine Learning in Computer Vision: A Review," EAI Endorsed Transactions on Scalable Information Systems, vol. 8, p. e4, 2021, doi:10.4108/eai.21-4-2021.169418.

[9] Juhartini, D. Arwidiyarti, and Desmiwati, "Single Shot Multibox Detector (SSD) in Object Detection: A Review", IJACI, vol. 1, no. 2, pp. 118–127, Jul. 2025, doi:10.71129/ijaci.v1i2.pp118-127.

[10] W. Liu, et al., "SSD : Single Shot MultiBox Detector," in Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 1–17, doi:10.1007/978-3-319-46448-0_2.

[11] J.-S. Lim, M. Astrid, H.-J. Yoon, S.-I. Lee, "Small Object Detection using Context and Attention," 2019, pp. 4321–4330, doi: 10.48550/arXiv.1912.06319.

[12] D. Jiang, B. Sun, S. Su, Z. Zuo, P. Wu, X. Tan, "FASSD: A Feature Fusion and Spatial Attention-Based Single Shot Detector for Small Object Detection," Electronics, vol. 9, no. 1536, pp. 1–20, 2020, doi:10.3390/electronics9091536.

[13] G. Meena, K. K. Mohbey, A. Indian, S. Kumar, "Sentiment Analysis from Images using VGG-19 based Transfer Learning Approach," Procedia Computer Science, vol. 204, pp. 411–418, 2022, doi:10.1016/j.procs.2022.08.050.

[14] M. Z. Khaliki, M. S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN," Scientific Reports, vol. 14, p. 2664, 2024, doi:10.1038/s41598-024-52823-9.

[15] T. Shahriar, “Comparative analysis of lightweight deep learning models for memory-constrained devices,” 2025, doi:10.48550/arXiv.2505.03303.

[16] S. G. Mohammad, A. E. A. Ali, M. M. Saeid, "Optimization MobileNetV3-small Model for Diabetic Foot Ulcers Classification," International Journal of Intelligent Engineering & Systems, vol. 18, no. 9, 2025, doi: 0.22266/ijies2025.1031.51.

[17] D. Saha, M. P. Mangukia, A. Manickavasagan, "Real-Time Deployment of MobileNetV3 Model in Edge Computing Devices Using RGB Color Images for Varietal Classification of Chickpea," Applied Sciences, vol. 13, p. 7804, 2023, doi:10.3390/app13137804.

[18] H. Wang, V. Bhaskara, A. Levinshtein, et al., "Efficient Super-Resolution Using MobileNetV3," in Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2020, vol. 12537, doi:10.1007/978-3-030-67070-2_5.

[19] S. Athisayamani, R. S. Antonyswamy, V. Sarveshwaran, M. Almeshari, Y. Alzamil, V. Ravi, “Feature Extraction Using a Residual Deep Convolutional Neural Network (ResNet-152) and Optimized Feature Dimension Reduction for MRI Brain Tumor Classification,” Diagnostics, vol. 13, no. 4, 2023, doi:10.3390/diagnostics13040668.

[20] M. K. Panda, B. N. Subudhi, T. Veerakumar, V. Jakhetiya, “Modified ResNet-152 network with hybrid pyramidal pooling for local change detection,” IEEE Trans. Artif. Intell., vol. 5, no. 4, pp. 1599–1612, Apr. 2024, doi:10.1109/TAI.2023.3299903.

[21] S. Basodi, C. Ji, H. Zhang, Y. Pan, "Gradient Amplification: An Efficient Way to Train Deep Neural Networks," Big Data Mining and Analytics, vol. 3, no. 3, pp. 196–207, 2020, doi:10.26599/BDMA.2020.9020004.

[22] A. Abuhussein, M. F. Darwish, A. Darwish, and A. E. Hassanien, "Comprehensive Facial Thermal Dataset," Mendeley Data, V1, 2024. doi: 10.17632/8885sc9p4z.1.

[23] H. Rezatofighi, et al., "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 658-666, doi:10.1109/CVPR.2019.00075.

[24] Q. Tang, H. Qiang, Y. Tian, X. Feng, W. Hao, M. Xie, “ATDIOU: Arctangent Differential Loss Function for Bounding Box Regression,” Sensors, vol. 26, p. 1545, 2026, doi:10.3390/s26051545.

[25] B. Han, Q. Tang, J. Song, Z. Wang, Y. Yang, “GAOC: A Gaussian Adaptive Ochiai Loss for Bounding Box Regression,” Sensors, vol. 26, p. 368, 2026, doi:10.3390/s26020368.

[26] D. Chicco, M. J. Warrens, G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.” PeerJ Computer Science, vol. 7, p.e623, 2021, doi:10.7717/peerj-cs.623.

[27] S. M. Robeson, and C. J. Willmott, “Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," PLoS ONE, vol. 18, no. 2, 2023, doi:10.1371/journal.pone.0279774.

[28] K. Warneke, S. D. Siegel, J. Afonso, et al., “What the mean absolute percentage error (MAPE) should adopt from Bland–Altman analyses,” German Journal of Exercise and Sport Research, 2025, doi:10.1007/s12662-025-01084-3.

BIBLIOGRAPHY

Ridho Sholehurrohman
The author holds a Master of Mathematics degree from Sepuluh Nopember Institute of Technology and a Master of Informatics Engineering degree from IBI Darmajaya. He is currently an active lecturer in the Department of Computer Science at the University of Lampung. In addition to his teaching responsibilities, he is passionate about independent research and has produced numerous scholarly publications. His research interests include data handling, object detection, and classification. Moreover, he actively contributes to community service activities facilitated by the university. For communication purposes, the author can be reached via email at ridho.sholehurrohman@fmipa.unila.ac.id.

Akmal Junaidi
The author is an Associate Professor at the Department of Computer Science, Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia. He earned an M.Sc. from the University of Twente, the Netherlands, in 2003 and completed his Dr. rer. nat. at the Technische Universität Dortmund, Germany, in 2016. His research interests include expert systems, machine learning, computer science, and image processing. He can be contacted via email at akmal.junaidi@fmipa.unila.ac.id.

Hamzah Hanif
The author completed primary education at SD N 1 Blitarejo (2015), junior high school at SMP N 1 Gadingrejo (2018), and senior high school at SMA N 1 Gadingrejo (2021). The author earned a Bachelor of Computer Science (S.Kom) from the University of Lampung. During undergraduate studies, the author served as a teaching assistant for Fundamentals of Programming (2022/2023) and Distributed Data Processing (2023/2024).

Favorisen Rosyking Lumbanraja
The author is an Associate Professor at the Department of Computer Science, Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia. He obtained his M.Si. degree from Institut Pertanian Bogor, Indonesia, in 2011, and completed his Ph.D. at Kanazawa University, Japan, in 2017. His research interests include bioinformatics, expert systems, machine learning, and computer science. He can be contacted via email at favorisen.lumbanraja@fmipa.unila.ac.id.

Mochammad Reza Habibi
The author graduated from Sepuluh Nopember Institute of Technology (ITS), Surabaya, where he obtained a Master's degree in Computational Mathematics. He is currently a Ph.D. student in the Department of Computer Science and Electrical Engineering at Kumamoto University, Japan. He also serves as a lecturer in the Department of Business Statistics at ITS. His research interests focus on machine learning and its applications. In addition, he actively participates in university-based community service activities. For correspondence, the author can be reached at reza.habibi@its.ac.id .