SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Lightweight Deep Learning Models for Facial Expression Recognition in Inclusive Education

Miftahul Ilmi[1)*], Doni Syofiawan[2)]
[1,2)] Institut Teknologi dan Bisnis Indobaru Nasional, Indonesia
[1)] miftahulilmi12@gmail.com, [2)]syofiawandoni@gmail.com

Submitted :Sep 27, 2025 | **Accepted** : Oct 17, 2025 | **Published** : Oct 28, 2025

Abstract: Facial expression recognition is an essential component in the development of artificial intelligence-based learning systems, particularly in the context of inclusive education that involves students with special needs. This study aims to evaluate the performance of several lightweight deep learning architectures in detecting facial expressions with high accuracy while maintaining computational efficiency. Facial image data were obtained from both public datasets and newly collected samples, which were preprocessed through face cropping, normalization, and data augmentation. The dataset was split into 70% training, 15% validation, and 15% testing. Four lightweight deep learning architectures: MobileNetV2, MobileNetV3 (Small and Large), and EfficientNetB0, were employed as the primary models using transfer learning and fine-tuning approaches. Evaluation was conducted using accuracy, loss, precision, recall, and F1-score metrics, complemented by visualization through confusion matrices. The results indicate that MobileNetV2 achieved the best performance with a test accuracy of 92%, precision of 93%, recall of 91%, and F1-score of 92%, while maintaining a relatively lightweight parameter size of 2.26 million. EfficientNetB0 ranked second with 83% accuracy, followed by MobileNetV3-Large (77%), whereas MobileNetV3-Small demonstrated the lowest performance (45%). Confusion matrix analysis revealed recurring misclassification patterns for certain expressions, such as Happy often misclassified as Sad, and Neutral overlapping with Angry. This study confirms that MobileNetV2 is the most optimal architecture for implementing facial expression recognition systems in inclusive education environments, as it balances high accuracy with computational efficiency. These findings provide a solid foundation for developing intelligent applications that support adaptive interaction in the learning process..

Keywords: Facial Expression, Lightweight Deep Learning, Mobilenet, Efficientnet, Inclusive Education.

INTRODUCTION

Advances in artificial intelligence (AI) technology and deep learning have made significant contributions in the field of pattern recognition, especially facial recognition and expression (Ge et al., 2022; Tang, 2023). Facial Expression Recognition (FER) plays a critical role in a wide range of applications, from security, healthcare, to education (Gan & Zhang, 2022; C. Liang & Dong, 2023). Facial expressions are one of the main forms of non-verbal communication that are able to provide information about a person's emotional state quickly and accurately (Lan & Lin, 2022). In the context of education, especially inclusive education, FER has great potential to help educators in understanding the emotional state of students who have limited verbal communication, such as students with autism spectrum disorder (ASD) or deaf (ElMahalawy et al., 2024).

Although FER's research has shown rapid progress, most approaches still utilize complex and large-sized deep learning architectures. Models such as VGGNet, ResNet, or EfficientNet are capable of achieving high accuracy, but require significant computing resources (Melinda et al., 2024). This is a major obstacle when the technology is applied in real-world environments with limited hardware, such as in inclusive classrooms that use mobile devices or edge computing with limited capacity (Bie et al., 2022). Thus, there is an urgent need to develop FER models that are not only accurate, but also lightweight and efficient.

Some recent research has proposed the use of lightweight architectures, such as MobileNet, ShuffleNet, and small versions of EfficientNet, designed to achieve the optimal trade-off between performance and efficiency (Chen et al., 2022; Zeng et al., 2023a). However, most of those studies focused on general datasets such as FER-

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

2013 or CK+, without exploring the specific context of inclusive education (J. Zhu & Cao, 2023a). In fact, inclusive educational environments have unique characteristics, where variations in student expression are often more subtle, limited, or different from standard datasets (Liu, 2024). This creates a research gap, namely the need for an in-depth evaluation related to the application of the lightweight model in supporting adaptive learning and responsiveness to students' emotional needs (Wei et al., 2022).

In addition, an important issue that is still rarely discussed is how lightweight models can maintain competitive performance in real-world conditions. Factors such as inconsistent lighting, facial angle and limited amount of labeled data from inclusive students can degrade system performance. Therefore, a strategy is needed that not only relies on architectural efficiency, but also involves data augmentation techniques, regularization, and transfer of learning from large to lightweight models.

Based on these problems, this study focuses on the implementation and evaluation of various lightweight deep learning models (MobileNetV2, MobileNetV3, and EfficientNetB0) for facial expression recognition tasks. This study aims to test the extent to which lightweight models can provide high accuracy while maintaining computational efficiency, thus enabling real application in an inclusive education environment. The main contribution of this study is to provide a comprehensive analysis of the performance of lightweight models in FER based on educational contexts, while offering optimal modeling recommendations to support more adaptive teaching-learning interactions.

Thus, this research is expected to enrich the literature related to the application of light deep learning in the field of FER, especially in inclusive education. The results of this study are also expected to be the basis for the development of a smart learning support system that is able to improve the quality of teacher-student interaction, support learning success, and strengthen inclusivity in the modern educational environment.

Therefore, this study aims to (1) compare the performance of several lightweight deep learning architectures for FER, (2) evaluate their suitability for resource-limited inclusive education environments, and (3) identify the best-performing model balancing accuracy and efficiency.

LITERATURE REVIEW

Facial expression recognition (*facial expression recognition*, FER) is one of the important research areas in *computer vision* and artificial intelligence. FER has a significant role in various applications such as mental health, human-computer interaction, inclusive education, and security systems (Xu et al., 2022a). The main challenge in the development of FER systems is the need for high accuracy in real-world conditions full of variability, such as differences in lighting, viewing angles, and individual variations (J. Zhu & Cao, 2023b). In the last decade, the approach based on *deep learning* has dominated research in this field due to its ability to extract representative features from facial images automatically (Yang et al., 2023). However, most of the architecture *deep learning* tend to be complex and require large computing resources, making them less suitable for deployment on limited devices (Trivedi & Goyani, 2024). Therefore, there is a need for lightweight models (*lightweight models*) that is still able to maintain high performance with low complexity.

Other studies show that integration *attention mechanism* Can improve the performance of lightweight models in detecting facial expressions. Xu et al. (2022) report that the MobileNet variant is able to achieve competitive accuracy on FER datasets with a relatively small number of parameters, making it suitable for application to edge devices (Xu et al., 2022b). Liang (2023) developed MobileNetV3 with more efficient architecture modifications through the use of coordinate attention mechanisms and activation function optimization, which has been proven to improve facial expression recognition performance without significantly increasing the computational burden (X. Liang et al., 2023). These results confirm that lightweight architecture is not only efficiency-oriented, but also capable of maintaining accuracy close to heavier conventional models (Q. Zhu et al., 2024a).

Liao et al. (2024) in their research highlighted that the current FER research trend emphasizes more on the application of lightweight architecture with a combination of optimization techniques such as transfer learning, data augmentation, and regularization (Liao et al., 2024). The focus of the research is not only on achieving high accuracy in the laboratory, but also on the adaptability of the model to challenging real-world conditions. Some studies confirm that although lightweight models have a smaller number of parameters, their performance can be improved by utilizing a balanced large dataset and precise fine-tuning techniques (Zeng et al., 2023b). Furthermore, recent research underscores the importance of model efficiency in the context of inclusive education. The FER system can be used to support teachers in recognizing the emotional state of students, especially those with special needs, such as children with autism or communication disorders (Biçer & Kose-Bagci, 2024). According to the report in *Journal of Visual Communication and Image Representation*, the integration of lightweight architecture with transfer learning-based optimization strategies enables real-time detection of facial expressions on mobile devices, making it easier to implement in classrooms (Tang et al., 2024). Model efficiency is a key aspect because the devices used in education often have limited resources.

In addition to MobileNet, EfficientNet is also a model that is often tested in FER because of its *compound scaling*which is able to balance the depth, width, and resolution of the network systematically (Tang et al., 2024).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Aikyn et al. (2023) developed an EfficientNet-based FER framework that is optimized for edge computing, with efficient results while being able to maintain accuracy on limited devices (Aikyn et al., 2023). This makes it one of the strong candidates in the development of FER based systems *edge computing*. Comparisons of performance between MobileNet and EfficientNet in the literature show that both are equally feasible, with the final choice often determined by application-specific needs such as latency, accuracy, and hardware limitations.

From the overall literature review, it can be concluded that the direction of FER research is increasingly leading to the development of lightweight models that balance accuracy with efficiency. MobileNetV2, MobileNetV3, and EfficientNetB0 are proving to be promising architectures as they are capable of delivering results close to heavy models with much lower computing requirements. This trend is relevant to research on inclusive education, where intelligent systems must be able to run on simple devices while maintaining detection speed and accuracy. Therefore, the evaluation and comparison of the performance of the three models is important to contribute to the development of an adaptive, practical, and useful FER in the real world.

METHOD

The flow of the proposed research methodology is shown in Figure 1, which includes stages from data collection to model evaluation.

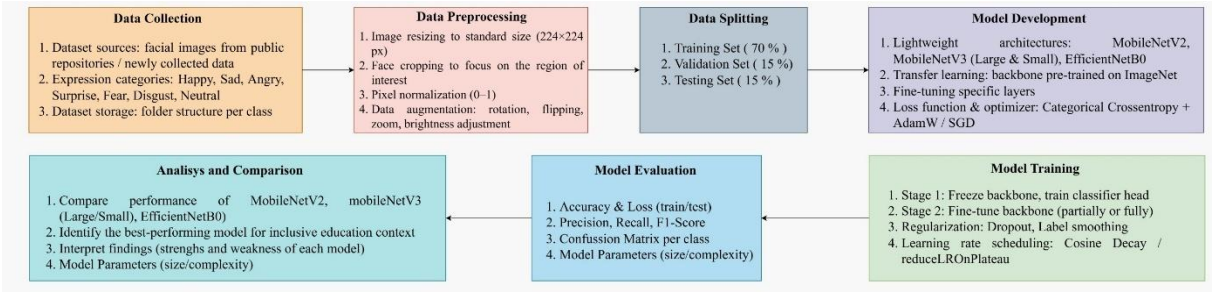

Figure 1. Flow Of Research Process

Stage 1. Data Collection

At this stage, data in the form of facial images is collected both from public repositories and newly collected data. The data is categorized into several emotional expressions, namely Happy, Sad, Angry, Surprise, Fear, Disgust, and Neutral. The data storage structure is created in the form of folders per class to facilitate the next process.
Results: Datasets were organized by facial expression categories.

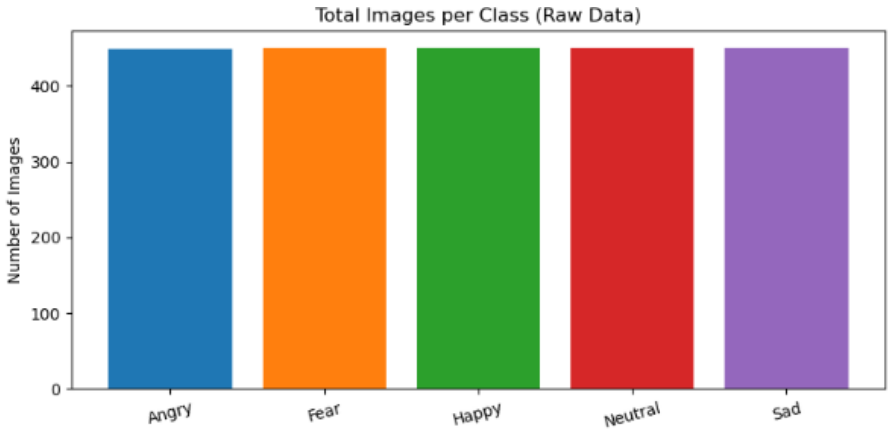

Figure 2. Data Collection

Stage 2. Data Preprocessing

The face image was resized to a standard 224×224 pixels to standardize the model input. Next, face cropping is carried out so that the model focuses on the face area, accompanied by pixel normalization (0–1) to accelerate the convergence of training. Data augmentation techniques such as rotation, flipping, zoom, and brightness adjustment are applied to increase data variation and reduce the risk of overfitting.
Results: Ready-to-use datasets, uniform in size, more varied, and face-focused.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 3. The face image was resized

Stage 3. Data Splitting
The dataset is divided into three parts: training data (70%), validation data (15%), and test data (15%). The training data is used to train the model, the validation data helps the tuning process and prevents overfitting, while the test data serves to objectively measure the final performance of the model.
Results: Proportional distribution of datasets to support fair experiments.

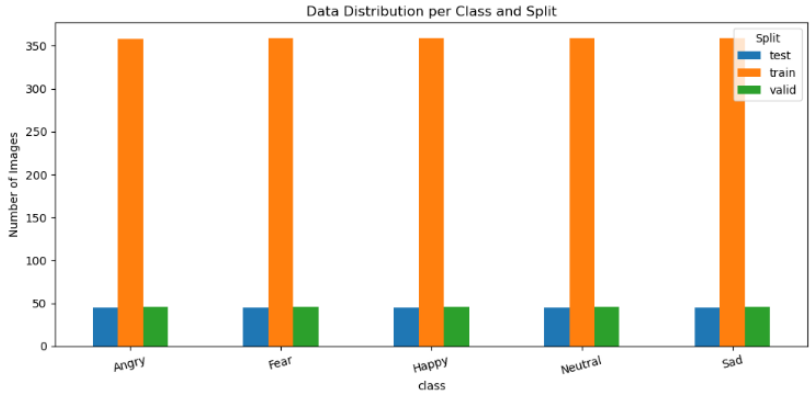

Figure 4. Data Splitting

Stage 4. Model Development
The lightweight architecture was chosen as MobileNetV2, MobileNetV3 (Large/Small), and EfficientNetB0. Each architecture leverages transfer learning with an initial weight from ImageNet. Some layers are frozen for the initial stage, then fine-tuned to certain layers. The loss function used is Categorical Crossentropy with an AdamW or SGD optimizer.
Result: A trainable, ready-to-train baseline model with a lightweight architecture-compliant configuration.

Stage 5. Model Training
The training process is carried out in two stages. The first stage is to train the classifier head with the backbone frozen. The second stage is fine-tuning by opening up part or all of the backbone layer so that the model can learn more specifically about the facial expression dataset. Regulatory strategies are implemented through dropout and label smoothing, as well as learning rate scheduling such as Cosine Decay or ReduceLROnPlateau.
Results: Models that have been trained gradually to achieve optimal performance with good generalizations.

Stage 6. Model Evaluation
The evaluation was carried out using several main metrics, namely accuracy, loss, precision, recall, and F1-score, both on training data and test data. In addition, a confusion matrix is also displayed to see the prediction distribution per class. The number of parameters of each model is recorded as a measure of complexity.
Results: Quantitative metric values and analysis per class were the basis for comparison between models.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Stage 7. Analysis and Comparison
The final stage is to analyze and compare the performance of all the architectures tested. MobileNetV2, MobileNetV3 (Large/Small), and EfficientNetB0 were compared in terms of accuracy, loss, precision, recall, F1-score, and number of parameters. This analysis also highlights the strengths and weaknesses of each model in the context of inclusive education.
Results: The best models can be identified, along with a clear interpretation of the reasons for their superiority over other models.

RESULT

This section presents the results of performance evaluation of several Lightweight Deep Learning models used in the research. Evaluation was carried out on four main models, namely MobileNetV2, EfficientNetB0, MobileNetV3Large, and MobileNetV3Small. Each model was measured for performance using five evaluation metrics, namely test accuracy, loss value, precision, recall, and F1-Score, as well as the number of parameters that represent the complexity of the model.

To provide a comprehensive overview, the results of each model are summarized in Table 1. This summary makes it easier for readers to make a direct comparison of the prediction performance between models, so that it can be known which Lightweight Deep Learning model is the most optimal in detecting facial expressions in this study.

Table 1. Model Evaluation Results

Model	Test Accuracy	Loss	Precision	Recall	F1-Score	Parameter
MobileNetV2	0.92	0.28	0.93	0.91	0.92	2.26 M
EfficientNetB0	0.83	0.41	0.84	0.82	0.83	4.05 M
MobileNetV3Large	0.77	0.56	0.78	0.76	0.77	3.00 M
MobileNetV3Small	0.45	1.12	0.47	0.44	0.45	0.94 M

The results shown in Table 1 illustrate a comprehensive evaluation of four lightweight deep learning models for the Facial Expression Recognition (FER) task. In general, there is a striking variation in performance between models, both in terms of accuracy and architectural complexity.

The MobileNetV2 model shows the most superior performance with a test accuracy of 92%. A low loss value of 0.28 indicates that the model's prediction is close to the actual label. In addition, the precision (0.93) and recall (0.91) metrics have a good balance, resulting in an F1-score of 0.92. This shows that MobileNetV2 is not only capable of detecting facial expressions with high accuracy, but also maintaining consistency between false positive and false negative error rates. With a total of 2.26 million parameters, this high performance is achieved without excessive computing load, making it the best candidate for implementation on resource-constrained devices, such as smartphones or embedded systems.

In contrast, EfficientNetB0 ranks second with an accuracy of 83% and a loss value of 0.41. Although lower than the MobileNetV2, this model still shows stability with a precision value of 0.84, recall of 0.82, and an F1-score of 0.83. A larger number of parameters, which is 4.05 million, gives an indication that the increase in complexity is not always directly proportional to the increase in accuracy. This model is still worth considering if the application requires a trade-off between generalization stability and a slightly higher tolerance for complexity.

The MobileNetV3Large model produces 77% accuracy with a loss value of 0.56. The precision (0.78) and recall (0.76) metrics were relatively balanced, resulting in an F1-score of 0.77. With a total of 3.00 million parameters, this model is lighter than EfficientNetB0, but it is not capable of matching its performance. These findings show that newer architecture adaptations do not always provide significant benefits, especially in domains with data limitations or unbalanced class distributions.

In contrast, the MobileNetV3Small became the lowest-performing model, achieving only 45% accuracy and a loss value of 1.12. The precision (0.47), recall (0.44), and F1-score (0.45) showed that the model failed to capture the relevant patterns in the data. Although the number of parameters is only 0.94 million, which is the smallest among other models, an excessively high compression level comes at the expense of accuracy. These results confirm a trade-off between efficiency and accuracy, where extreme simplification of the architecture is not recommended if the target application demands high performance.

In addition to using the quantitative evaluation metrics summarized in Table 1, the model performance analysis is also strengthened by visualization through a confusion matrix. The confusion matrix provides a more detailed picture of the model's prediction distribution for each class, so that it can be known to what extent the model is able to classify facial expressions correctly and whether errors still occur. This visualization is important for identifying specific error patterns between classes that may not be visible from accuracy, precision, recall, or F1-score values alone.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

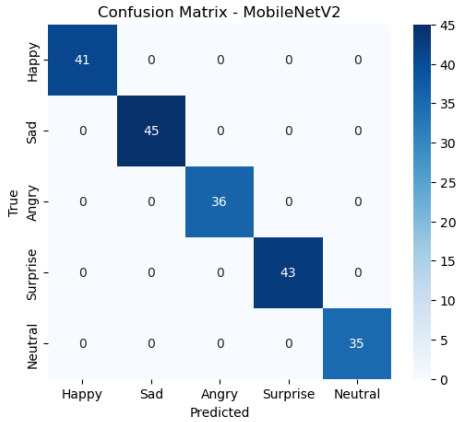

Figure 5. Confusion Matrix – MobileNetV2

Based on Figure 5, it can be seen that MobileNetV2 is able to classify all classes of facial expressions very well. The Happy class was detected correctly as many as 41 samples, Sad as many as 45 samples, Angry as many as 36 samples, Surprise as many as 43 samples, and Neutral as many as 35 samples. No cross-classification errors between classes were found, so the diagonal in the confusion matrix was fully filled with true positive predictions.

This reinforces the results in Table 1, where MobileNetV2 achieved high accuracy (92%) and other evaluation metrics also consistently showed optimal performance. With these results, it can be concluded that MobileNetV2 not only provides high average performance, but is also consistent across every category of facial expression without bias against a particular class. To provide a more comprehensive comparison, confusion matrix analysis was also performed on another model, namely EfficientNetB0, which ranked second best after MobileNetV2 based on Table 1. This visualization aims to assess how the model's prediction is distributed to each class and whether there is a tendency for misclassification between expressions.

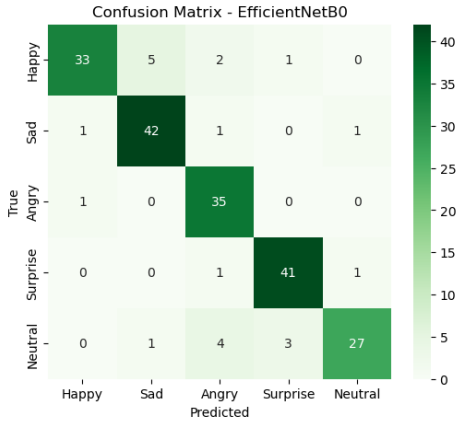

Figure 6. Confusion Matrix – EfficientNetB0

Based on Figure 6, it can be seen that the performance of EfficientNetB0 is quite good but not as clean as MobileNetV2. This model is able to correctly recognize the Sad class in 42 samples, the Surprise class in 41 samples, and the Angry class in 35 samples. However, in the Happy class, although most of the samples were detected correctly (33 samples), there were a number of misclassifications, namely 5 samples were classified as Sad and 2 samples as Angry. The same thing happened in the Neutral class, which was only correctly identified as many as 27 samples, with some cases of misprediction to the Sad, Angry, and Surprise classes.

In general, this confusion matrix shows that EfficientNetB0 still faces challenges in distinguishing between Happy and Neutral expressions, although its accuracy (83%) still ranks it as the highest-performing model after MobileNetV2. These results show a trade-off between a larger model capacity (4.05 million parameters) and a tendency for mild overfitting in a given class. In addition to MobileNetV2 and EfficientNetB0, evaluations were also carried out on MobileNetV3Large, which in Table 1 occupies the third position with lower accuracy compared to the previous two models. The confusion matrix analysis in this model aims to identify specific weaknesses in distinguishing classes of facial expressions.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

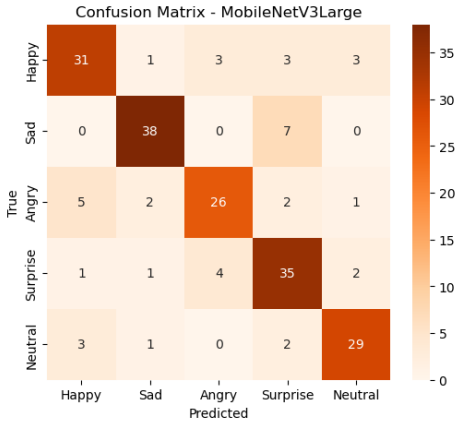

Figure 7. Confusion Matrix – MobileNetV3Large

Based on Figure 7, it can be seen that MobileNetV3Large is still able to recognize most classes, albeit with a higher rate of misclassification than MobileNetV2 and EfficientNetB0. The Sad class performed best with 38 correctly classified samples, while the Surprise class was well recognized in 35 samples. However, significant weaknesses emerged in the Angry class, which was only correctly detected in 26 samples, accompanied by quite a lot of prediction errors to the Happy (5 samples) and Sad (2 samples) classes. In the Happy class, there are 31 correct predictions, but some images are misclassified into Angry, Surprise, or Neutral classes. Meanwhile, the Neutral class has 29 correct predictions, with a small percentage of errors to other classes, especially Happy and Surprise.

Overall, this confusion matrix confirms that MobileNetV3Large has difficulty distinguishing expressions with similar visual features, such as Happy vs Neutral and Angry vs Sad. This is consistent with an overall accuracy value of 77%, which is lower than the previous two models, although the model parameters are larger (3.00 million). The MobileNetV3Small model is the lightest variant with a total of only 0.94 million parameters. Although it is designed for efficiency, it performs much lower than other models, as can be seen from the accuracy value of only 45% in Table 1. The following confusion matrix analysis provides a clear picture of the fairly high distribution of prediction errors in this model.

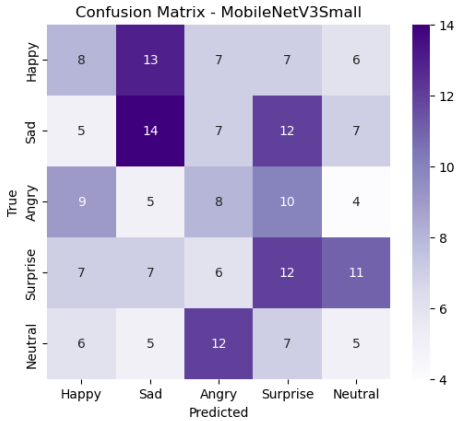

Figure 8. Confusion Matrix – MobileNetV3Small

Based on Figure 8, it can be seen that MobileNetV3Small has significant difficulty in distinguishing between facial expression classes. The correct predictions are evenly distributed but the numbers are relatively low: 14 Sad samples, 12 Surprise samples, 12 Angry samples, 11 Neutral samples, and only 8 Happy samples are correctly classified. The misclassification in the MobileNetV3Small model looks quite dominant. Many expressions of Happy are mispredicted as Sad (13 samples) or Angry (7 samples), while expressions of Sad are often mistakenly identified as Surprise (12 samples) and even Happy (5 samples). In addition, Angry expressions were incorrectly distributed to the Happy (9 samples) and Surprise (10 samples) classes, indicating an overlap of visual features between these expressions. The expression Surprise itself is often classified as Neutral (11 samples), while the expression Neutral is most often misidentified as Angry (12 samples). This error distribution indicates that MobileNetV3Small is not able to distinguish the visual characteristics between expressions consistently, so its accuracy is much lower than that of other models.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15370
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This condition indicates that the MobileNetV3Small model is not able to adequately capture the visual features that distinguish facial expressions. This is in line with the low evaluation metrics in Table 1, where precision, recall, and F1-score are all in the range of 0.44–0.47, confirming the limitations of this model in multi-class classification tasks.

DISCUSSIONS

The results of the evaluation obtained show that **MobileNetV2** excel in the context of this study, especially judging by the high combination of accuracy, recall, precision, and F1-score, as well as a relatively small parameter size. These findings are consistent with several previous studies that optimized the MobileNetV2 architecture for facial expression recognition tasks. For example, Zhu et al. (2024) propose a modification strategy *MobileNetV2* by integrating channel attention and reverse fusion mechanisms to preserve minor feature information, which improves the accuracy of the FER-2013 and CK+ datasets (Q. Zhu et al., 2024b).

In addition, the study "Facial Expression Recognition Using MobileNetV2 with Attention Mechanism and Facial Landmarks" by Huang et al. (2025) combined activation maps with *landmark*-driven attention in MobileNetV2 to strengthen the focus on important areas of the face and show that this approach is able to improve generalization without significantly increasing the parameter load "Facial Expression Recognition Using Mobilenetv2 with Attention Mechanism and Facial Landmarks" (Huang et al., 2025) . Your excellent research results may be in line with these findings—that careful attention mechanisms and careful light architecture selection can help reduce misclassifications between classes.

For the EfficientNet-based model, the study "Transfer Learning Technique with EfficientNet for Facial Expression Recognition" reported that the use of fine-tuned EfficientNet-B0 on datasets such as CK+ and JAFFE was able to achieve high accuracy (99.57% and 100%) under limited dataset conditions, even in a controlled test environment. This approach supports the idea that even lightweight models can be optimized with transfer learning so that their performance is close to large architectures while still being efficient (Alam et al., 2022).

However, the confusion matrix results for models such as MobileNetV3Small show weaknesses in distinguishing expressions that are visually similar (e.g. Happy vs Sad, Neutral vs Angry). This suggests that although very lightweight architectures are very attractive in terms of efficiency, the risk of underfitting or loss of minor discriminatory features can arise. In the literature, the survey "Advances in Facial Expression Recognition: A Survey of Deep FER methods, datasets, and challenges" by Kopalidis et al. (2024) underscores that one of the main challenges in modern FER is to overcome minor expression variations and the influence of external factors such as lighting, facial orientation, and identity bias (Kopalidis et al., 2024).

Your finding that lightweight models like MobileNetV2 can perform best among other lightweight models in the context of inclusive data signals that the architecture of choice as well as training strategies are crucial. However, these results also suggest that other lightweight models (such as MobileNetV3Large/Small) may require additional optimization (e.g. advanced data augmentation, regularization, selective selection of fine-tuning layers) to minimize misclassification between overlapping classes.

In practical terms, these results support the idea that in the context of inclusive education, where hardware is often limited, the selection of a model that balances efficiency and accuracy is essential—rather than prioritizing just one aspect. The use of MobileNetV2 with the attention mechanism could be a very solid starting point for a real system in an inclusive classroom.

CONCLUSION

This study successfully evaluated the performance of various *lightweight deep learning* models for facial expression recognition tasks in the context of inclusive education. The results of the experiment showed that MobileNetV2 provided the best performance with test accuracy of 92%, precision of 93%, recall of 91%, and F1-score of 92%, while having a relatively low number of parameters (2.26 million). This confirms that *lightweight architecture* can be an effective solution for the implementation of facial expression recognition systems on devices with limited resources, such as in inclusive school environments.

Meanwhile, EfficientNetB0 achieves an accuracy of 83%, which shows a fairly good balance between complexity and performance, although it is still below MobileNetV2. Other models such as MobileNetV3Large and MobileNetV3Small tend to experience significant declines in predictive accuracy and consistency, which can be seen from the distribution of misclassification in the confusion matrix. These differences emphasize the importance of choosing the right architecture to avoid underfitting problems or losing minor discriminatory features.

The main contribution of this study lies in the comprehensive comparative analysis of *lightweight models*, which shows that computational efficiency is not always directly proportional to accuracy, but can be optimized through the selection of the right architecture. In addition, the results of this study provide a practical basis for the development of hardware-friendly deep learning-based facial recognition applications , especially to support interactions in inclusive classrooms.

*name of corresponding author

For further research, some of the things that can be explored include the application of *attention mechanism* techniques to strengthen focus on relevant areas of the face, the use of more complex data augmentation strategies to reduce bias between classes, and integration with multimodal modalities such as voice or body movements to increase the robustness of the system. Thus, the resulting system is expected to be not only efficient, but also accurate and adaptive to real variations in an inclusive education environment.

ACKNOWLEDGMENT

The author expressed his gratitude to the National Indobaru Institute of Technology and Business for the support of the research facilities provided. Gratitude was also conveyed to the Ministry of Higher Education, Science, and Technology through the Directorate General of Research and Development which has provided funding support and direction so that this research can be carried out properly.

REFERENCES

Aikyn, N., Zhanegizov, A., Aidarov, T., Bui, D.-M., & Tu, N. A. (2023). Efficient facial expression recognition framework based on edge computing. *J. Supercomput.*, *80*, 1935–1972. https://doi.org/10.1007/s11227-023-05548-x

Alam, I. N., Kartowisastro, I. H., & Wicaksono, P. (2022). Transfer Learning Technique with EfficientNet for Facial Expression Recognition System. *Revue d'Intelligence Artificielle*, *36*(4), 543–552. https://doi.org/10.18280/ria.360405

Biçer, E., & Kose-Bagci, H. (2024). A Lightweight Facial Expression Recognition Model Specialized for Hearing-Impaired Children. *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, 1–4. https://doi.org/10.1109/SIU61531.2024.10601133

Bie, M., Xu, H.-Y., Gao, Y., & Che, X. (2022). Facial Expression Recognition from a Single Face Image Based on Deep Learning and Broad Learning. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2022/7094539

Chen, Q., Jing, X., Zhang, F., & Mu, J. (2022). Facial Expression Recognition Based on A Lightweight CNN Model. *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 1–5. https://doi.org/10.1109/bmsb55706.2022.9828739

ElMahalawy, J., ElSwaify, Y. A., Elliboudy, D., Abbas, O. M., Moustafa, N., & Wael, N. (2024). AI-Powered Human-Computer Interaction Assisting Early Identification of Emotional and Facial Symptoms of Autism Spectrum Disorder in Children: "A Deep Learning-Based Enhanced Facial Feature Recognition System." *2024 International Conference on Machine Intelligence and Smart Innovation (ICMISI)*, 87–93. https://doi.org/10.1109/ICMISI61517.2024.10580320

Gan, B., & Zhang, C. (2022). Target Detection and Network Optimization: Deep Learning in Face Expression Feature Recognition. *J. Sensors*, *2022*, 1–10. https://doi.org/10.1155/2022/5423959

Ge, H., Zhu, Z., Dai, Y., Wang, B., & Wu, X. (2022). Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine*, *215*. https://doi.org/10.1016/j.cmpb.2022.106621

Huang, H., Qu, C., Xiang, F., & Li, X. (2025). *Facial Expression Recognition Using Mobilenetv2 with Attention Mechanism and Facial Landmarks*. SSRN. https://doi.org/10.2139/ssrn.5132286

Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets. *Information*, *15*(3), 135. https://doi.org/10.3390/info15030135

Lan, J., & Lin, G. (2022). A Review of Facial Expression Recognition. *Frontiers in Computing and Intelligent Systems*. https://doi.org/10.54097/fcis.v2i1.2969

Liang, C., & Dong, J. (2023). A Survey of Deep Learning-based Facial Expression Recognition Research. *Frontiers in Computing and Intelligent Systems*. https://doi.org/10.54097/fcis.v5i2.12445

Liang, X., Liang, J., Yin, T., & Tang, X. (2023). A lightweight method for face expression recognition based on improved MobileNetV3. *IET Image Process.*, *17*, 2375–2384. https://doi.org/10.1049/ipr2.12798

Liao, L., Wu, S., Song, C., & Fu, J. (2024). RS-Xception: A Lightweight Network for Facial Expression Recognition. *Electronics*. https://doi.org/10.3390/electronics13163217

Liu, J. (2024). Lightweight facial expression estimation for mobile computing in portable device. *Internet Technology Letters*, *8*. https://doi.org/10.1002/itl2.533

Melinda, M., Afny, N., Andryani, C., Nurdin, Y., Khariyunnisa, V., Yulita, Y., Ketut, I., Enriko, A., & Andriyani, C. (2024). Deep learning performance analysis for facial expression based autism spectrum disorder identification. *Radioelectronic and Computer Systems*. https://doi.org/10.32620/reks.2024.2.03

Tang, X. (2023). Research on Face Expression Recognition Based on Deep Learning. *2023 8th International Conference on Information Systems Engineering (ICISE)*, 508–511. https://doi.org/10.1109/ICISE60366.2023.00113

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Tang, X., Feng, J., Huang, J., Xiang, Q., & Xue, B. (2024). A lightweight and continuous dimensional emotion analysis system of facial expression recognition under complex background. *J. Vis. Commun. Image Represent.*, *103*. https://doi.org/10.1016/j.jvcir.2024.104260

Trivedi, H., & Goyani, M. (2024). Robust Face Recognition in the Presence of Diverse challenges: A Hybrid Deep Neural Network Approach. *International Journal of Engineering Research and Applications*. https://doi.org/10.9790/9622-14105562

Wei, C., Kuo, C. J., Testa, R. L., Machado-Lima, A., & Nunes, F. L. S. (2022). ExpressionHop: A Lightweight Human Facial Expression Classifier. *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 198–203. https://doi.org/10.1109/MIPR54900.2022.00042

Xu, X., Tao, R., Feng, X., & Zhu, M. (2022a). *A Lightweight Facial Expression Recognition Network Based on Dense Connections*. 347–359. https://doi.org/10.1007/978-3-031-07920-7_27

Xu, X., Tao, R., Feng, X., & Zhu, M. (2022b). *A Lightweight Facial Expression Recognition Network Based on Dense Connections*. 347–359. https://doi.org/10.1007/978-3-031-07920-7_27

Yang, J., Yang, X., Wang, C., Zhang, H., & Zhang, Y. (2023). *Research on MobileNet-based lightweight face recognition algorithm*. *12934*, 129340–129340. https://doi.org/10.1117/12.3008092

Zeng, M., Luo, Y., & Liu, G. (2023a). Lightweight Facial Expression Recognition Network with Dynamic Deep Mutual Learning. *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing*. https://doi.org/10.1145/3592686.3592726

Zeng, M., Luo, Y., & Liu, G. (2023b). Lightweight Facial Expression Recognition Network with Dynamic Deep Mutual Learning. *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing*. https://doi.org/10.1145/3592686.3592726

Zhu, J., & Cao, Y. (2023a). Face Expression Recognition Combining Improved DeeplabV3+ and Migration Learning. *Journal of Physics: Conference Series*, *2555*. https://doi.org/10.1088/1742-6596/2555/1/012020

Zhu, J., & Cao, Y. (2023b). Face Expression Recognition Combining Improved DeeplabV3+ and Migration Learning. *Journal of Physics: Conference Series*, *2555*. https://doi.org/10.1088/1742-6596/2555/1/012020

Zhu, Q., Zhuang, H., Zhao, M., Xu, S., & Meng, R. (2024a). A study on expression recognition based on improved mobilenetV2 network. *Scientific Reports*, *14*. https://doi.org/10.1038/s41598-024-58736-x

Zhu, Q., Zhuang, H., Zhao, M., Xu, S., & Meng, R. (2024b). A study on expression recognition based on improved mobilenetV2 network. *Scientific Reports*, *14*(1), 8121. https://doi.org/10.1038/s41598-024-58736-x

*name of corresponding author