Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16510
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Analysis of Image Augmentation Effects on MobileNetV2 for Gastroesophageal Reflux Disease Endoscopy Image Classification

Dinda Chesar Putri Ramadhani[1)], Risqy Siwi Pradini[2)*], M. Syauqi Haris[3)]
[1,2,3)]Fakultas Sains dan Teknologi, Institut Teknologi Sains dan Kesehatan RS.DR. Soepraoen Kesdam V/BRW, Indonesia
[1)]dindachesar2@gmail.com, [2)]risqypradini@itsk-soepraoen.ac.id, [3)]haris@itsk-soepraoen.ac.id

Submitted :Jul 3, 2026 | **Accepted** : Jul 17, 2026 | **Published** : Jul 30, 2026

Abstract: Gastroesophageal Reflux Disease (GERD) is a common digestive tract disorder, and endoscopic image-based diagnosis still faces challenges due to variations in lighting, image capture angles, and limited data that can cause deep learning models to overfit and reduce generalization capabilities. While previous studies frequently prioritize overall accuracy, this study distinguishes itself by focusing on increasing recall as a critical indicator of clinical sensitivity for positive case detection. This study aims to analyze the effect of applying image augmentation on the performance of GERD endoscopic image classification using a lightweight MobileNetV2 architecture, offering a practical solution for resource-constrained clinical settings. The research methods include data collection, class mapping, 80:20 data split, and standardized preprocessing stages. Two training scenarios were compared: without augmentation and with strategic on-the-fly augmentation applied through a transfer learning approach. Evaluation was carried out using accuracy, precision, recall, F1-score, and ROC-AUC. The results showed that augmentation improved model performance, with accuracy increasing from 76.72% to 78.06%, recall increasing from 79.33% to 85.27%, F1-score increasing from 77.86% to 80.04%, and AUC increasing from 0.8317 to 0.8481. These findings indicate that the augmentation strategy effectively improves the generalization of the lightweight model, significantly reducing missed diagnoses to support early detection of GERD. In addition, the results demonstrate that applying augmentation techniques can help the model learn more robust visual features from limited medical image datasets. This approach also contributes to reducing overfitting and improving model reliability in practical clinical image analysis scenarios.

Keywords: GERD; Image Augmentation; Medical Image Classification; MobileNetV2; Recall

INTRODUCTION

Endoscopy serves as a key diagnostic technique for indentifying disorders of thr upper digestive tract, particularly Gastroesophageal Reflux Disease (GERD) (Siboni et al. 2025). GERD occurs when stomach acid increases and irritates the esophageal mucosa, this condition can cause symptoms such as heartburn to serious complications such as esophagitis or esophageal cancer (Shaqran et al. 2023). Although endoscopy is effective, the quality of endoscopic images is often affected by lighting and the angle of the image, which can lead to inaccurate diagnoses (Niu et al. 2025). In addition, the manual image capture process also takes a long time (Yen et al. 2023). This condition encourages the need for an automatic classification system that is stable, accurate, and has high sensitivity in detecting positive cases.

Deep Learning technology enables computer systems to learn to recognize complex visual patterns and features from medical images, which makes it possible to assist in the automated diagnosis process (Tsuneki 2022). Several studies, such as GerdNet-RF conducted by Yen et al., using a combination of transfer learning and machine learning, have achieved accuracy of up to 92.5% (Yen et al. 2022). These results indicate that deep learning has great potential to improve endoscopic image-based diagnosis.

In the context of medical diagnosis, evaluating model performance depends not only on accuracy, but also on the model's ability to correctly detect all positive cases (Hicks et al. 2022). The recall (sensitivity) metric shows

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16510
e-ISSN : 2541-2019
p-ISSN : 2541-044X

how well the model is able to identify patients who truly have GERD without missing any false negatives (Abdulahi, Ogundokun, and Adenike 2024). In the detection of diseases such as GERD, errors in the form of false negatives can have more serious consequences than false positives because they can cause delays in clinical treatment.

However, due to the limited amount of data and high visual variation, deep learning modeling faces challenges (Kebaili, Lapuyade-Lahorgue, and Ruan 2023). As a result, deep learning models experience overfitting and have difficulty generalizing to new data (Pallikkavaliyaveetil and Chandrasekaran 2026). To address this issue, the variation of artificial data is expanded through image augmentation techniques without re-acquisition. Image augmentation techniques are the process of artificially expanding training data through transformations such as rotation, translation, changes in brightness, contrast, and zoom (Park et al. 2023). The application of this technique helps deep learning models learn more diverse visual patterns, thereby increasing robustness and generalization capabilities, especially in endoscopic images (Khalifa, Loey, and Mirjalili 2022).

Several previous studies have shown that image augmentation techniques can improve model performance on endoscopic images. A Generative Adversarial Networks (GAN)-based study produced synthetic augmentations that improved model generalization and image variation of precancerous gastric lesions (Magalhães, Neto, and Cunha 2023). In addition, Malik et al., showed that deep learning models are capable of achieving high performance on endoscopic images, with the VGG-19 + CNN model achieving an accuracy of up to 99.45% (Malik et al. 2024). These results indicate that image augmentation contributes to enhancing the accuracy and stability of deep learning models used in endoscopic image analysis.

However, no research has specifically evaluated the effect of augmentation techniques on the performance of the MobileNetV2 architecture on GERD endoscopic images, particularly on limited datasets with varying lighting and shooting angles. This has led to a lack of research in selecting the optimal augmentation strategy. While deeper CNN architectures such as VGG or ResNet often deliver high accuracy, they suffer from massive parameter sizes and heavy computational burdens, making them less practical for real-time clinical applications. Therefore, this study selects the MobileNetV2 model as the primary architecture because it offers a significantly lower parameter count and higher computational efficiency while still delivering strong, competitive performance in medical image classification (Ali Hasan, Siwi Pradini, and Anshori 2025). MobileNetV2 achieves this through its depthwise separable convolution and inverted residual block mechanisms, which allow a substantial reduction in parameters without compromising feature extraction capabilities (Zhou et al. 2024). These advantages make MobileNetV2 highly capable of maintaining an optimal balance between high accuracy and low computational cost, making it highly suitable for limited yet visually complex medical datasets (Zhou et al. 2024), (Fang et al. 2025).

Thus, this study aims to analyze the comparative performance of the MobileNetV2 model before and after the application of image augmentation techniques. The main focus of the evaluation is directed at the effect of augmentation on increasing recall and the generalization ability of the model in classifying GERD endoscopic images. Therefore, this study is expected to provide two main contributions: first, demonstrating an image augmentation strategy in improving recall and generalization ability of the MobileNetV2 model on endoscopy datasets, and second, providing an accurate MobileNetV2-based implementable approach to support automated diagnostic systems in GERD detection.

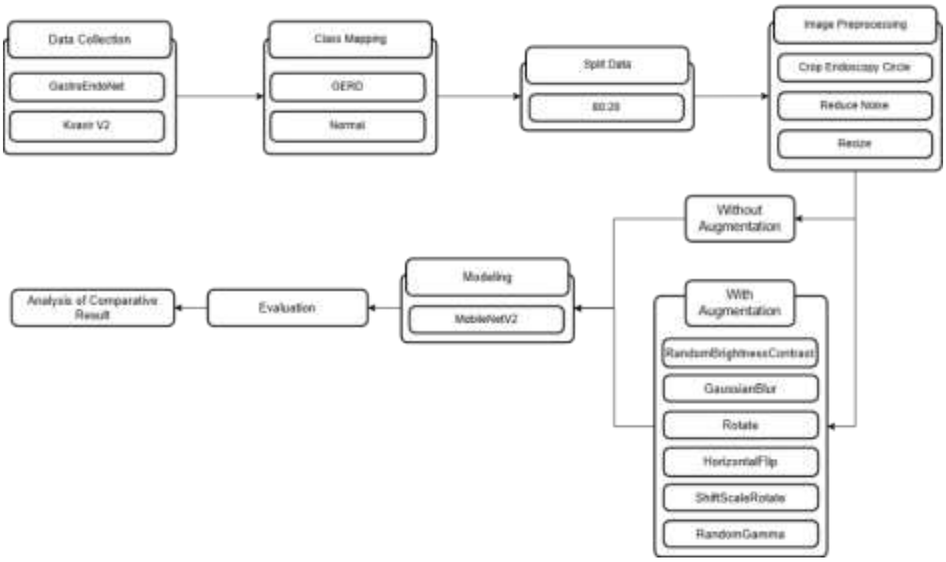

Fig. 1 Research Flow

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16510
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

The research flow consists of several main stages: data collection, class mapping, data splitting, preprocessing, augmentation, modeling using MobileNetV2, evaluation, and comparative analysis. This flow separates two training scenarios: without and with augmentation, so that the effect of augmentation on model performance can be analyzed, as shown in Fig. 1.

Data Collection

This study used two endoscopic image datasets, namely GastroEndoNet and Kvasir V2, both of which contain images of the esophageal mucosa (Kowshir et al. 2025), (Pogorelov et al. 2017). GastroEndoNet consists of 974 images for the GERD class and 1,103 for the Normal GERD class, while Kvasir V2 provides 1,000 images for the Esophagitis and Normal Z-Line classes. Using these two datasets allows the study to obtain a wider variety of endoscopic images for GERD classification.

Class Mapping

At this stage, the class mapping process is carried out to standardize the labels of the two datasets to suit the needs of the two-class classification, as shown in Fig. 2.

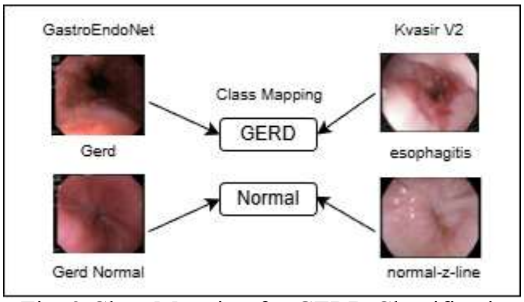

Fig. 2 Class Mapping for GERD Classification

Fig. 2 explains the binary class mapping process from two source datasets. The Esophagitis class in Kvasir V2 and the Gerd class in GastroEndoNet are combined into the GERD class, as both present esophageal mucosal inflammation with endoscopic patterns such as erythema, erosion, or constriction (Shaqran et al. 2023). Meanwhile, the Normal Z-Line class on Kvasir V2 and Normal Gerd on GastroEndoNet were combined into the Normal class because both showed esophageal mucosa without signs of inflammation or damage (Pogorelov et al. 2017). The merger forms two final classes, namely GERD and Normal shown in Fig. 3.

(a) (b)

Fig. 3 Example of Endoscopic Images (a) GERD class and (b) Normal class

After the class mapping and merging process, a total of 1,974 GERD images and 2,103 Normal images were obtained, thus forming a larger and more balanced combined dataset for model training.

Split Data

The combined dataset, which has undergone a class mapping process, is divided into two subsets, with 80% as the training set and 20% as the test set. The division is done randomly while maintaining the class proportions in each subset. This division aims to provide the model with sufficient data for learning, while also providing test data that the model has never seen before to measure its generalization ability. The 80:20 proportion was chosen as a common and optimal strategy in deep learning research on datasets with a limited number of images, maximizing the data available for deep learning model learning (Bichri, Chergui, and Hain 2024).

Image Preprocessing

The Image Preprocessing stage is applied to all training and test data to standardize image quality before entering the model. The first step is Crop Endoscopy Circle with circle fitting to form a circular binary mask, so that only the endoscopy area becomes the Region of Interest (ROI), while the dark border is removed (Budd et al. 2023). After that, use Bilateral Filter to

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16510
e-ISSN : 2541-2019
p-ISSN : 2541-044X

reduce noise without removing important edge structures (Lv et al. 2022). All images were then resized to the 224 × 224 pixels standard input of MobileNetV2. This step ensures that the endoscopic images are cleaner, more uniform, and ready for the feature extraction process, as shown in Fig. 4.

Fig. 4 Example of Image Preprocessing Application

Image Augmentation

The Image Augmentation stage is carried out to artificially add image variations so that the model is more robust and does not overfit easily (Kebaili et al. 2023). Augmentation is only applied to the training data in the augmented scenario, while in the unaugmented scenario, the data is used as is after preprocessing. The augmentation process is performed on-the-fly via a custom data generator, so transformations are applied dynamically to each batch (Oza et al. 2022). Augmentation was implemented using the Albumentations library, which was chosen for its high performance, good flexibility, and for providing appropriate transformations for medical images without altering important anatomical structures (Saputra et al. 2026).

The augmentation techniques used include RandomBrightnessContrast, RandomGamma, GaussianBlur, Rotate, ShiftScaleRotate, and HorizontalFlip. The augmentation parameters used are RandomBrightnessContrast (p=0.4), RandomGamma (p=0.3), GaussianBlur (p=0.3), Rotate (limit=20°, p=0.4), ShiftScaleRotate (shift_limit=0.07, scale_limit=0.10, rotate_limit=10°, p=0.4), and HorizontalFlip (p=0.4). These parameters were chosen to simulate variations in lighting and angle of endoscopic image acquisition without changing the main anatomical structures.

Modeling Using MobileNetV2

The MobileNetV2 architecture is built on the concept of inverted residuals and linear bottlenecks, which are designed to produce a lightweight model that is still capable of extracting important features from medical images (Zhu et al. 2024). Each main block consists of three stages, namely 1x1 expansion convolution with ReLU6 activation, 3x3 depthwise convolution, and 1x1 linear projection convolution as a bottleneck (Sandler et al. 2018). This structure is depicted in Fig. 5, which is adapted from the MobileNetV2 architecture diagram on (Zhu et al. 2024) and is used to clarify the feature processing flow in blocks with stride 1 and stride 2.

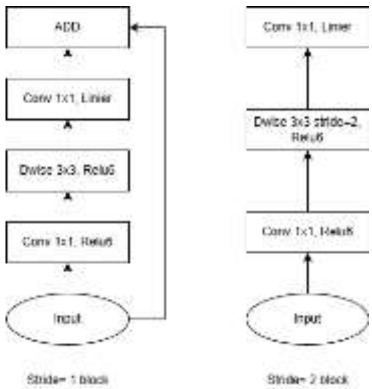

Fig. 5 MobileNetV2 Architecture

The characteristics of MobileNetV2 are suitable for complex endoscopic image analysis with limited data. The model was initialized using Transfer Learning with ImageNet pre-trained weights and received 224 × 224 pixel input. To optimize the learning process, most of the MobileNetV2 backbone was frozen to maintain common features, while the final layers and classifier head were unlocked for training. The additional classifier layer included Global Average Pooling, followed by Batch Normalization and a Dropout layer (0.3) to serve as a regularization strategy for minimizing overfitting. The output layer was Dense with sigmoid activation for binary classification between GERD and Normal classes. The model was configured with the Adam optimizer at a learning rate of 1 × 10^(-4) and utilized Binary Crossentropy as the loss function. The training process employed batch size of 32 sample per batch over 30 epochs. To prevent overfitting, Early Stopping (patience=6) was used based on validation accuracy, and ReduceLROnPlateau (factor=0.3, patience=4) to gradually adjust the learning

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16510 p-ISSN : 2541-044X

rate if validation performance did not improve. Then it is trained on data that has gone through the preprocessing and augmentation stages so that the classification performance is more stable and robust.

Evaluation and Analysis of Comparative Result

Model evaluation was conducted using test data to assess its performance in differentiating GERD and normal images. The evaluation process was conducted in two different scenarios: without image augmentation and with image augmentation. Model performance was assessed using several evaluation metrics, including accuracy, precision, recall, F1-score, confusion matrix, and ROC curve and AUC. These metrics were applied to obtain a comprehensive assessment of the model, particularly in terms of sensitivity and its capability to distinguish between the two classes (Kocak et al. 2025). The ROC curve illustrates how the True Positive Rate and False Positive Rate cary across different threshold values, while AUC quantifies the model's capability to distinguish between GERD and Normal classes (Müller, Soto-Rey, and Kramer 2022).

a. Accuracy

Accuracy represents the ratio of correctly predicted instances to the total number of samples.

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

b. Precision

Precision is used to measure the level of accuracy of true positive predictions.

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

c. Recall

Recall or sensitivity measures how well the model is able to recognize the positive class.

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

d. F1-Score

F1-Score represents the harmonic mean of precision and recall.

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

In the context of GERD detection, recall (sensitivity) is a key metric because false negatives can lead to missed inflammatory conditions, potentially delaying diagnosis and medical treatment. The evaluation results from both scenarios were then compared to determine the effect of image augmentation on improving recall (sensitivity) and the generalization ability of the MobileNetV2 model.

RESULT

Image Augmentation

Image augmentation was applied on-the-fly only to the training data to increase data variation and prevent overfitting. The test data was not augmented to maintain the objectivity of the model evaluation. The transformations applied included lighting, rotation, translation, scaling, blurring, and horizontal flipping. The augmentation parameters were designed to maintain the anatomical structure of the esophagus so as not to change the clinical meaning of the image.

Medical image augmentation is known to improve model generalization by enriching the variation of data distribution without changing class labels, especially in datasets with high variation in lighting and shooting angles. This is visually shown in Fig. 6, Augmentation produces variations in light intensity and object orientation that resemble real conditions in endoscopic procedures, such as changes in lighting, camera angle movements and changes in image orientation that are common in endoscopic procedures.

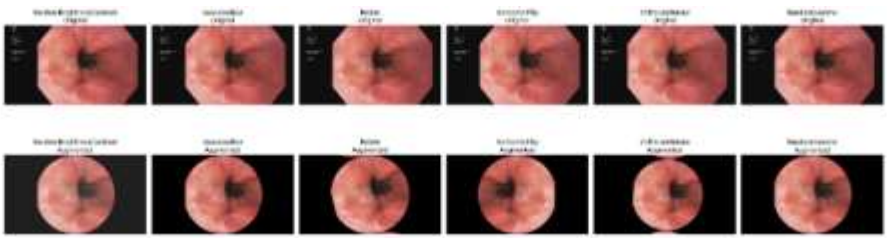

Fig. 6 Example of Image Augmentation Application

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16510
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Model Performance Without Augmentation

The MobileNetV2 model was first trained without applying image augmentation. The Adam optimizer was applied during model training, with the learning rate, batch size, and training epochs selected through experimentation. The training process showed a significant increase in training accuracy compared to validation accuracy, indicating a tendency for overfitting to the training data, as shown in Fig. 7.

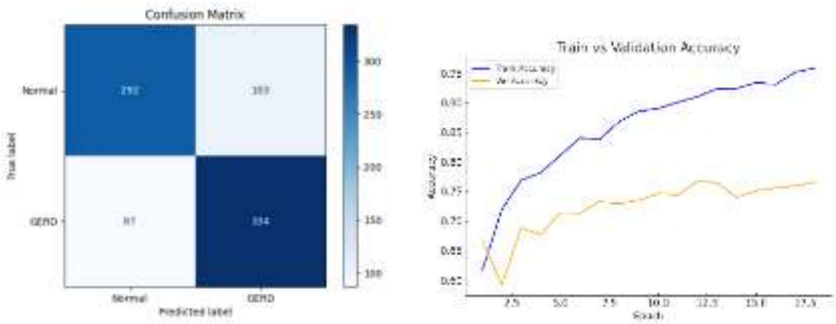

Fig. 7 Train vs Validation Accuracy Graph and Confusion Matrix Without Image Augmentation

The confusion matrix on Fig. 7, the model successfully classified 334 GERD images correctly, while 87 GERD images were still misclassified as Normal (false negatives). The model successfully recognized 292 Normal images, while 103 were incorrectly labeled as GERD (false positives). Quantitatively, the model without augmentation produced an accuracy of 76.72%, a precision of 76.43%, a recall of 79.33%, and an F1-score of 77.86%. The recall value of 79.33% indicates that the model is able to recognize most GERD cases, but there are still a number of false negatives that have the potential to impact the delay in clinical detection. In the context of medical diagnosis, false negative errors are riskier than false positives because they can cause cases of disease to go undetected.

Model Performance With Augmentation

The model was then retrained using the scenario with image augmentation. The training curve shows a smaller gap between training and validation, indicating an improvement in the model's generalization ability, as shown in Fig. 8.

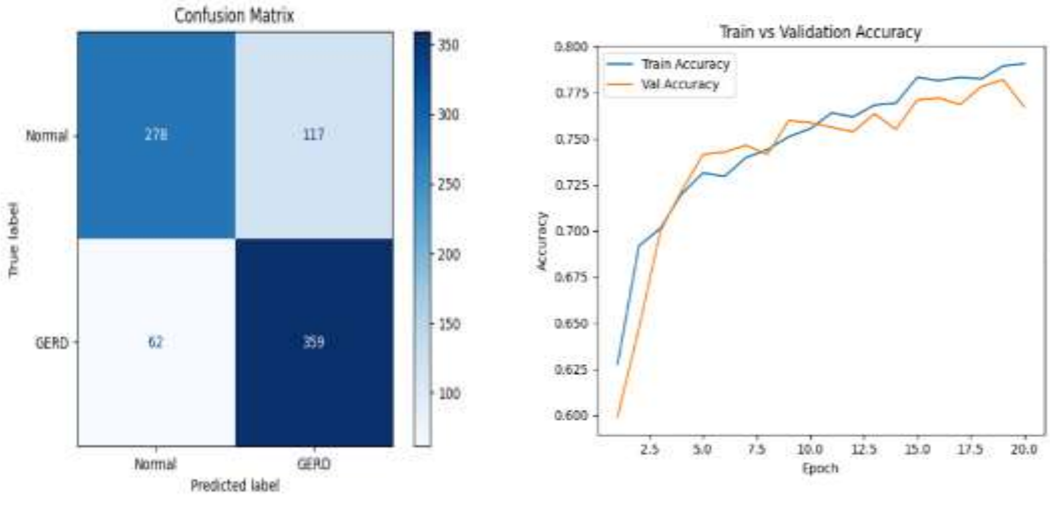

Fig. 8 Train vs Validation Accuracy Graph and Confusion Matrix with Image Augmentation

The confusion matrix is shown in Fig. 8, this scenario shows an increase in the number of correct predictions in the GERD class to 359 images, while the number of false negatives decreased to 62 images. The model accurately recognized 278 Normal images were correctly classified and 117 images were incorrectly predicted as GERD. In general, the model trained with image augmentation obtained an accuracy of 78.06%, a precision of 75.42%, a recall of 85.27%, and an F1-score of 80.04%. The increase in recall from 79.33% to 85.27% indicates that the application of image augmentation increases the model's sensitivity in detecting GERD cases and significantly reduces the number of false negatives.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16510 p-ISSN : 2541-044X

ROC and AUC Curve

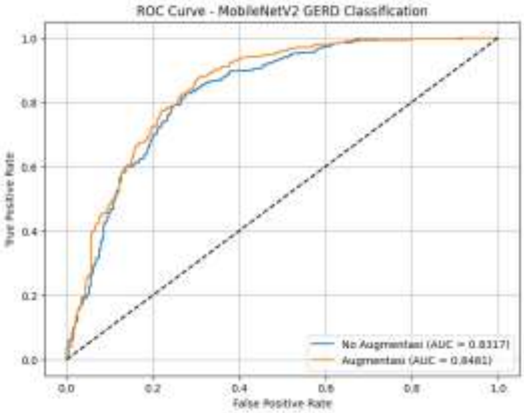

Fig. 9 ROC and AUC Curve Graph

Based on the ROC Curve graph shown in Fig. 9, the figure shows a comparison of the performance of the MobileNetV2 model in two scenarios, namely without augmentation and with augmentation. The ROC curve illustrates how the True Positive Rate (TPR/sensitivity) varies against the False Positive Rate (FPR) across different classification thresholds. Points closer to the upper-left corner indicate better classification performance in separating GERD and non-GERD classes. The graph shows that the curve with augmentation (orange line) is consistently slightly above the curve without augmentation (blue line), indicating that the model with augmentation has better discrimination ability across almost the entire threshold range.

The AUC (Area Under the Curve) values support these findings. The unaugmented model achieved an AUC of 0.8317, while the augmented model achieved an AUC of 0.8481. This improvement indicates that the augmented model has a higher probability of providing a higher prediction score for GERD images compared to normal images. Generally, an AUC value above 0.8 is categorized as good classification performance, so both models are in the good category, but the augmented scenario demonstrates more optimal performance.

Thus, based on ROC and AUC analyses, the application of image augmentation has been shown to improve the generalization ability and discriminative power of MobileNetV2 in classifying GERD endoscopic images. This is consistent with the previously analyzed recall improvement results, thus strengthening the conclusion that augmentation positively contributes to the overall model performance.

Comparative Analysis with Previous Research

The comparison of both scenarios indicates that the use of image augmentation improves the model results, particularly in terms of recall, which is the primary focus of this study. An increase of 5.94% in recall, along with a reduction in false negatives from 87 to 62 images, indicates improved effectiveness of the model in detecting GERD cases. While the increase in accuracy is relatively moderate, the increase in sensitivity has more important clinical implications because it reduces the risk of falsely detecting GERD cases. In medical classification systems, recall or sensitivity is often considered a more critical metric than accuracy because it directly relates to the system's ability to identify patients who truly have the disease.

Furthermore, the more stable training curve in the augmentation scenario indicates that the additional image variations helped the model learn esophageal inflammation patterns more comprehensively. These results confirm that the applied augmentation strategy is able to enhance the generalization ability of MobileNetV2 and improve the GERD detection performance on datasets with varying lighting and image capture angles. To place these results in the context of previous research, a comparison was made with several studies related to gastrointestinal endoscopy image classification as shown in Table 1. Previous studies generally used more complex or ensemble-based architectures, such as a combination of ResNet and Inception, which achieved accuracies of up to 87%–94%. However, the model proposed in this study uses the lightweight MobileNetV2 architecture with a structured augmentation approach.

To understand the broader significance of the findings of this study, the results obtained were then compared with several previous studies on endoscopic image classification.

*name of corresponding author

Table 1. Summary of GERD Classification Model Performance and Previous Research

Author & Year	Augmentation	Model	Accuracy	Recall
Akshaj Singh Bisht et al, 2025 (Bisht 2025)	Not mentioned	DeepGastroNet (Modified SqueezeNet)	77%	Gerd: 0.68 Gerd Normal: 0.82
Abdul Kadar Masum et al., 2025 (Kadar, Masum, et al. 2025)	Data augmentation (Not detailed)	ResNet-50 Inception-V3	73% 77%	Gerd: 0.65 Gerd Normal: 0.78 Gerd: 0.80 Gerd Normal: 0.61
Abdul Kadar Masum et al., 2025 (Kadar, Bitto, et al. 2025)	Flip, Rotation, Skew, Zoom	Ensemble	87%	Gerd: 0.84 Gerd Normal: 0.84
Muhammad John et al., 2025 (Abbas et al. 2025)	Horizontal flipping, rotation, brightness, contrast	GID-Xpert	75%	Gerd: 0.66 Gerd Normal:0.83
Our Research	RandomBrightnessContrast, RandomGamma, GaussianBlur, Rotate, ShiftScaleRotate, dan HorizontalFlip	MobileNetV2	78%	Gerd: 0.85 Gerd Normal: 0.70

Based on the comparison results on Table 1, It can be concluded that the application of image augmentation contributes to improving classification performance, especially in the recall metric for the GERD class. Studies without augmentation show a relatively lower recall for GERD, such as in Akshaj Singh Bisht et al. (Bisht 2025) with a value of 0.68, while studies that apply more varied augmentation tend to produce higher recall, such as Abdul Kadar Masum et al. (Kadar, Bitto, et al. 2025) with a recall of 0.84 in the ensemble model. In this context, this study, which uses the MobileNetV2 architecture with a combination of RandomBrightnessContrast, RandomGamma, GaussianBlur, Rotate, ShiftScaleRotate, and HorizontalFlip augmentations, shows a GERD recall of 0.85 with an accuracy of 78%, which is the highest recall value compared to other studies in the table. Although the accuracy is not the highest overall, the better recall achievement indicates that the MobileNetV2 model used is able to detect positive cases of GERD more sensitively.

In medical classification, increased recall has more important clinical implications because it is directly related to reducing misdetection of disease cases, so this study occupies a strong position in terms of detection sensitivity compared to previous studies. These findings indicate that this study successfully fills the gap in previous research related to the evaluation of image augmentation strategies in the MobileNetV2 architecture for Gastroesophageal Reflux Disease classification on limited endoscopy datasets.

DISCUSSIONS

The results showed that the application of image augmentation had a positive impact on the classification performance of MobileNetV2, particularly on increasing the recall value of the GERD class. An increase in recall demonstrates enhanced sensitivity of the model in recognizing disease cases and lowering the number of false negatives, which is essential in medical imaging diagnostics. The contribution of the performance improvement is influenced by a combination of six augmentation techniques, namely RandomBrightnessContrast and RandomGamma which simulate lighting variations, GaussianBlur which increases robustness to noise, and Rotate, ShiftScaleRotate, and HorizontalFlip which represent variations in orientation and angle of image capture. The combination of these techniques produces realistic visual variations without changing the main anatomical structure, thus helping the model learn esophageal inflammation patterns more comprehensively and improving generalization capabilities.

Furthermore, the use of a combined dataset expands the image diversity, making the model less dependent on the specific characteristics of a single data source, while augmentation helps address data limitations and variations in endoscopic image conditions such as lighting changes, reflections, and camera movement. This explains why the scenario with augmentation demonstrates more stable performance. Improved recall has important clinical implications as it reduces the risk of GERD cases going undetected and potentially supports early detection. Thus, this study confirms that a structured augmentation strategy can improve the sensitivity of lightweight models like MobileNetV2, while also addressing a research need related to augmentation evaluation in endoscopic image-based GERD classification on limited datasets.

CONCLUSION

The experimental results show that the application of image augmentation can improve model performance compared to without augmentation, indicated by an increase in accuracy from 76.72% to 78.06%, recall from 79.33% to 85.27%, F1-score from 77.86% to 80.04%, and AUC from 0.8317 to 0.8481. This increase in recall indicates that the model is more sensitive in detecting GERD cases and is able to reduce false negatives, which are important factors in medical image diagnosis systems, while also improving training stability and generalization ability to new data. Overall, the combination of structured augmentation strategies has proven effective in improving the sensitivity and discriminative ability of lightweight models such as MobileNetV2 on limited endoscopy datasets with high image condition variations. For further development, research can focus on comparing MobileNetV2 with other CNN architectures (such as EfficientNet, DenseNet, or ConvNeXt), testing various targeted augmentation strategies, and employing cross-validation techniques. Additionally, future studies should include validation on multi-center datasets or clinical data from different hospitals to improve robustness and ensure generalization to real clinical cases.

REFERENCES

Abbas, Muhammad John, Hend Alshaya, Wided Bouchelligua, Nehal Hassan, and Inzamam Mashood Nasir. 2025. "Hierarchical Multi-Stage Attention and Dynamic Expert Routing for Explainable Gastrointestinal Disease Diagnosis." *Diagnostics* 15(21):1–26. doi: 10.3390/diagnostics15212714.

Abdulahi, Abdulrahman Tosho, Roseline Oluwaseun Ogundokun, and Ajiboye Raimot Adenike. 2024. "PulmoNet : A Novel Deep Learning Based Pulmonary Diseases Detection Model." *BMC Medical Imaging* 1–19. doi: 10.1186/s12880-024-01227-2.

Ali Hasan, Hildan Adisqi, Risqy Siwi Pradini, and Mochammad Anshori. 2025. "Klasifikasi Spesies Jamur Beracun Agaricus Xanthodermus Dan Amanita Muscaria Menggunakan Transfer Learning Dengan Arsitektur MobileNetV2." *Jurnal Pendidikan Dan Teknologi Indonesia* 5(9):2443–51. doi: 10.52436/1.jpti.785.

Bichri, Houda, Adil Chergui, and Mustapha Hain. 2024. "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets." *International Journal of Advanced Computer Science and Applications* 15(2):331–39. doi: 10.14569/IJACSA.2024.0150235.

Bisht, Akshaj Singh. 2025. "DeepGastroNet : A Deep Learning Network for Improved Gastroesophageal Reflux Disease and Gastrointestinal Polyps Classification Using Endoscopic Images." *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* 1412–17. doi: 10.1109/ICAAIC64647.2025.11331233.

Budd, Charlie, Luis C. Garcia-Peraza Herrera, Martin Huber, Sebastien Ourselin, and Tom Vercauteren. 2023. "Rapid and Robust Endoscopic Content Area Estimation: A Lean GPU-Based Pipeline and Curated Benchmark Dataset." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* 11(4):1215–24. doi: 10.1080/21681163.2022.2156393.

Fang, Yu-jen, Kun-hua Lee, Riya Karmakar, Arvind Mukundan, Yaswanth Nagisetti, Chien-wei Huang, and Hsiang-chen Wang. 2025. "Transforming Endoscopic Image Classification with Spectrum-Aided Vision for Early and Accurate Cancer Identification." *Diagnostics* 15(21):1–23. doi: 10.3390/diagnostics15212732.

Hicks, Steven A., Inga Strümke, Vajira Thambawita, Malek Hammou, Michael A. Riegler, Pål Halvorsen, and Sravanthi Parasa. 2022. "On Evaluation Metrics for Medical Applications of Artificial Intelligence." *Scientific Reports* 12(1):1–9. doi: 10.1038/s41598-022-09954-8.

Kadar, Abdul, Abu Kowshir Bitto, Shafiqul Islam Talukder, Fokrul Islam, Mohammed Shamsul Alam, and Khandaker Mohammad. 2025. "An Explainable AI Based Deep Ensemble Transformer Framework for Gastrointestinal Disease Prediction from Endoscopic Images." *EAI Endorsed Transactioons* 4:1–9. doi: 10.4108/airo.9795.

Kadar, Abdul, Muhammad Masum, Fokrul Islam Khan, Ruhul Rabbi, Arif Hossen, Khandaker Mohammad, Mohi Uddin, and Abu Kowshir Bitto. 2025. "Enhancing Endoscopic Diagnosis of GERD and Polyps : A Comparative Study of Deep Learning Models for Multi-Class Classification." *2025 8th International Conference on Control, Robotics and Informatics (ICCRI)* 16–20. doi: 10.1109/ICCRI67201.2025.11262505.

Kebaili, Aghiles, Jérôme Lapuyade-Lahorgue, and Su Ruan. 2023. "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review." *Journal of Imaging* 9(4):81. doi: 10.3390/jimaging9040081.

Khalifa, Nour Eldeen, Mohamed Loey, and Seyedali Mirjalili. 2022. "A Comprehensive Survey of Recent Trends in Deep Learning for Digital Images Augmentation." *Artificial Intelligence Review* 55(3):2351–77. doi: 10.1007/s10462-021-10066-4.

Kocak, Burak, Michail E. Klontzas, Arnaldo Stanzione, Aymen Meddeb, Nathaniel Mercaldo, Oliver Díaz, and Renato Cuocolo. 2025. "Evaluation Metrics in Medical Imaging AI : Fundamentals , Pitfalls , Misapplications , and Recommendations." *European Journal of Radiology Artificial Intelligence* 3(June).

*name of corresponding author

doi: 10.1016/j.ejrai.2025.100030.

Kowshir, Abu, Hasan Imam, Kamrul Hassan, Aka Das, and Syed Minhaz. 2025. "GastroEndoNet : Comprehensive Endoscopy Image Dataset for GERD and Polyp Detection." *Data in Brief* 60:111572. doi: 10.1016/j.dib.2025.111572.

Lv, Hui, Pengfei Shan, Hongfang Shi, and Li Zhao. 2022. "An Adaptive Bilateral Filtering Method Based on Improved Convolution Kernel Used for Infrared Image Enhancement." *Signal, Image and Video Processing* 16(8):2231–37. doi: 10.1007/s11760-022-02188-1.

Magalhães, Bruno, Alexandre Neto, and António Cunha. 2023. "Generative Adversarial Networks for Augmenting Endoscopy Image Datasets of Stomach Precancerous Lesions: A Review." *IEEE Access* 11(November):136292–307. doi: 10.1109/ACCESS.2023.3338545.

Malik, Hassaan, Ahmad Naeem, Abolghasem Sadeghi-niaraki Rizwan, Ali Naqvi, Seung-won Lee, and Seung-won Lee. 2024. "Multi-Classification Deep Learning Models for Detection of Ulcerative Colitis , Polyps , and Dyed-Lifted Polyps Using Wireless Capsule Endoscopy Images." *Complex & Intelligent Systems* 10(2):2477–97. doi: 10.1007/s40747-023-01271-5.

Müller, Dominik, Iñaki Soto-Rey, and Frank Kramer. 2022. "Towards a Guideline for Evaluation Metrics in Medical Image Segmentation." *BMC Research Notes* 15(1):1–8. doi: 10.1186/s13104-022-06096-y.

Niu, Ke, Zeyun Liu, Xue Feng, Heng Li, and Kaize Shi. 2025. "Endoscopic Depth Estimation Based on Deep Learning: A Survey." *ArXiv* 3613:0–3.

Oza, Parita, Paawan Sharma, Samir Patel, Festus Adedoyin, and Alessandro Bruno. 2022. "Image Augmentation Techniques for Mammogram Analysis." *Journal of Imaging* 8(5):1–22. doi: 10.3390/jimaging8050141.

Pallikkavaliyaveetil, Nazreen, and Sriram Chandrasekaran. 2026. "Small Data , Big Challenges : Machine- and Deep-Learning Strategies for Data-Limited Drug Discovery." *Advanced Drug Delivery Reviews* 229(September 2025):115762. doi: 10.1016/j.addr.2025.115762.

Park, Hyun Cheol, In Pyo Hong, Sahadev Poudel, and Chang Choi. 2023. "Data Augmentation Based on Generative Adversarial Networks for Endoscopic Image Classification." *IEEE Access* 11(April):49216–25. doi: 10.1109/ACCESS.2023.3275173.

Pogorelov, Konstantin, Thomas De Lange, Dag Johansen, Peter Thelin Schmidt, Michael Riegler, Mathias Lux, and Pål Halvorsen. 2017. "Kvasir : A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection." *Association for Computing Machinery* 164–169. doi: 10.1145/3083187.3083212.

Sandler, Mark, Andrew Howard, Menglong Zhu, and Andrey Zhmoginov. 2018. "MobileNetV2: Inverted Residuals and Linear Bottlenecks." *ArXiv* 4510–20. doi: https://doi.org/10.48550/arXiv.1801.04381.

Saputra, Kurniawan Aji, Farrikh Al Zami, Defri Kurniawan, and Muhammad Naufal. 2026. "Multi-Disease Retinal Classification Using EfficientNet-B3 and Targeted Albumentations : A Benchmark on Kaggle Retinal Fundus Images Dataset." *Sinkron : Jurnal Dan Penelitian Teknik Informatika* 10(1):232–41. doi: 10.33395/sinkron.v10i1.15530.

Shaqran, Tariq M., Mustafa M. Ismaeel, Aljoharh Abdulaziz Alnuaman, Fatimah A. Al Ahmad, Ghadeer A. Albalawi, Jori N. Almubarak, Rakan S. AlHarbi, Rayan S. Alaqidi, Yaqin A. AlAli, Khaled S. Alfawaz, and Abdulrhman A. Daghriri. 2023. "Epidemiology, Causes, and Management of Gastro-Esophageal Reflux Disease: A Systematic Review." *Cureus* 15(10). doi: 10.7759/cureus.47420.

Siboni, S., M. Sozzi, P. Visaggi, R. De Maron, and E. Asti. 2025. "New Tools for the Grading and Scoring of Gastro-Esophageal Reflux Disease." *Revista de Gastroenterología de México (English Edition)* 90(3):412–27. doi: 10.1016/j.rgmxen.2025.09.005.

Tsuneki, Masayuki. 2022. "Deep Learning Models in Medical Image Analysis." *Journal of Oral Biosciences* 64(3):312–20. doi: 10.1016/j.job.2022.03.003.

Yen, Hsu Heng, Hui Yu Tsai, Chi Chih Wang, Ming Chang Tsai, and Ming Hseng Tseng. 2022. "An Improved Endoscopic Automatic Classification Model for Gastroesophageal Reflux Disease Using Deep Learning Integrated Machine Learning." *Diagnostics* 12(11):1–15. doi: 10.3390/diagnostics12112827.

Yen, Timothy, Blake Jones, Jeannine M. Espinoza, Sarguni Singh, Jonathan Pell, Anna Duloy, Sachin Wani, Frank I. Scott, and Swati G. Patel. 2023. "Optimizing Endoscopy Procedure Documentation Improves Guideline - Adherent Care in Upper Gastrointestinal Bleeding American Society of Anesthesiologists." *Digestive Diseases and Sciences* 68(6):2264–75. doi: 10.1007/s10620-023-07823-6.

Zhou, Guoping, Qiyu He, Xiaoli Liu, Xinghua Kai, Weikang Cao, Junning Ding, Bufeng Zhuang, Shuhua Xu, and Myo Thwin. 2024. "Optimizing MobileNetV2 for Improved Accuracy in Early Gastric Cancer Detection Based on Dynamic Pelican Optimizer." *Heliyon* 10(16):e35854. doi: 10.1016/j.heliyon.2024.e35854.

Zhu, Qiming, Hongwei Zhuang, Mi Zhao, Shuangchao Xu, and Rui Meng. 2024. "A Study on Expression Recognition Based on Improved MobilenetV2 Network." *Scientific Reports* 14(1):1–11. doi: 10.1038/s41598-024-58736-x.