International Journal of Advances in Data and Information Systems
Vol. 7, No. 1, April 2026, pp. 62~75
ISSN: 2721-3056, DOI: 10.59395/ijadis.v7i1.1482

Performance Evaluation of AdamW, RMSProp, and Nadam Optimizers on EfficientNetB2 Model for Image Data Classification

Fanita Damayanti[1], Sugiyarto Surono[2], Aris Thobirin[3]
[1,2,3] Mathematics, Universitas Ahmad Dahlan, Indonesia

Article Info

Article history:
Received Dec 03, 2025
Revised Dec 30, 2025
Accepted Jan 05, 2026

Keywords:
EfficientNetB2
Optimizer
RMSprop
Histopathological Image
Classification

ABSTRACT

This study examines the effect of different optimization algorithms on the performance of the EfficientNetB2 model in classifying lung and colon histopathology images. Three commonly used optimizers AdamW, RMSprop, and Nadam were analyzed to compare their influence on convergence trends, classification accuracy, and overall learning consistency. Using a five-class dataset covering benign and malignant tissue samples, the experimental results show that all three optimizers are able to deliver reliable predictions, although with varying performance characteristics. RMSprop emerges as the most effective optimizer, achieving the highest accuracy across all evaluation stages, with 99.05% during training, 99.16% on validation, and 98.72% on testing, along with the lowest loss values. This indicates that RMSprop facilitates faster and more stable convergence compared to the other two methods. AdamW also demonstrates strong predictive performance but shows limitations when distinguishing cancer types with closely similar morphological structures. Nadam attains high accuracy in early stages yet exhibits lower initial stability than RMSprop. Overall, pairing EfficientNetB2 with RMSprop provides the most optimal configuration for this classification task. These results offer valuable insights for designing better training strategies and strengthening the effectiveness of medical imaging based computer aided diagnostic systems.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Fanita Damayanti,
Mathematics,
Universitas Ahmad Dahlan,
Jl. Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul, Special Region of Yogyakarta.
Email: 2200015040@webmail.uad.ac.id

1. INTRODUCTION

Technological developments in deep learning have greatly advanced image-based classification systems, particularly in enhancing accuracy and efficiency in pattern recognition tasks [1]. Among many architectural models, Convolutional Neural Networks (CNN) remain essential because of their strong capability in extracting spatial and visual features that support classification performance [2]. Despite their advantages, traditional CNN architectures still encounter limitations related to computational cost and the need for extensive training data. These challenges have encouraged researchers to design lighter and more efficient CNN based models capable of achieving better performance in identifying image patterns [3].

To overcome these challenges, a number of new architectures have been proposed to strike a more balanced compromise between accuracy and computational cost, with EfficientNet being one of the most notable examples [4]. EfficientNet is recognized for producing high performance while maintaining a relatively low number of parameters [5]. The model achieves its efficiency through a compound scaling approach that expands its overall capacity in a balanced and

proportional way. By scaling these dimensions together, the model can enhance performance without significantly increasing computational load [6], [7]. The EfficientNet family includes variants from B0, B1, B2, B3, B4, B5, up to B7, each providing a different degree of model complexity. Mid-range variants like B2–B4 offer solid accuracy while keeping computational requirements manageable, which makes them a good fit for medical imaging applications [8].

EfficientNet-B2 was selected for this study because it provides a well-balanced combination of computational efficiency and classification accuracy. This version extends EfficientNet-B0 by applying compound scaling that expands the model's capacity and allows it to handle inputs of 260 × 260 pixels more effectively. These adjustments allow the model to capture finer spatial details while remaining computationally efficient [9]. Although the EfficientNet family has been widely researched, most studies focus primarily on model performance without examining how optimization algorithms influence the training process [10]. In fact, the choice of optimizer significantly affects convergence behavior, training stability, and the model's generalization capability [11]. Various optimization approaches have been developed, including momentum-based techniques, adaptive optimizers, and different learning rate strategies [12].

Modern optimizer algorithms commonly applied to deep learning models are AdamW [13], RMSProp [14], and Nadam (Nesterov-accelerated Adaptive Moment Estimation) [15], each of which is designed to speed up the convergence process and improve stability in neural networks. The algorithm often compares its performance in the CNN architecture to determine the optimizer's most efficient strategy for training accuracy and time. AdamW is a development of Adam with the concept of decoupled weight decay, which is the process of regularizing weight from the main optimizer. This approach makes parameter updates more stable and improves the generalization of the model in gradient updates [16]. RMSProp, on the other hand, adjusts how quickly each weight is updated by relying on an exponentially smoothed history of past gradient magnitudes, helping the model converge more rapidly and efficiently [17]. Nadam combines the principle of Adaptive Moment Estimation (Adam) with the predictive momentum of the Nesterov Accelerated Gradient (NAG), resulting in a more stable weight update, as well as accelerating convergence in deep learning model training [18].

Earlier studies have shown that the EfficientNet architecture and its variants perform well in classifying lung and colon histopathology images, with EfficientNet-B2 achieving the highest reported accuracy of 97.24%. However, this study only focused on the influence of the model without considering the role of algorithm optimizers in the training process. Therefore, the comparison between AdamW, RMSProp, and Nadam is an important topic to research in modern architectures such as EfficientNet-B2 [19].

Building on this foundation, the present study investigates how various optimizers influence the performance of EfficientNet-B2 when classifying tissue specimens from respiratory and intestinal regions using data sourced from the LC25000 collection. Specimens taken from both respiratory and intestinal tissues were examined using data obtained from the LC25000 dataset. Three commonly applied optimization algorithms AdamW, RMSProp, and Nadam were examined by analyzing their impact on convergence patterns, classification performance, the consistency of the learning process, and how well the network delivers reliable results when handling previously unseen images. The outcomes of this research are expected to support the formulation of training methods that strengthen both the reliability and overall performance of deep learning systems in the interpretation of histopathological image data.

2. RESEARCH METHOD

The study is carried out through a sequence of systematic stages, starting from the data input phase and followed by preprocessing, dataset division, and augmentation procedures. The modeling process encompasses the development of a Baseline CNN and the implementation of an EfficientNetB2 network optimized using AdamW, RMSprop, and Nadam. Each component of the workflow is subsequently assessed through an evaluation phase. The complete methodological process is illustrated in the flowchart provided in Figure 1.

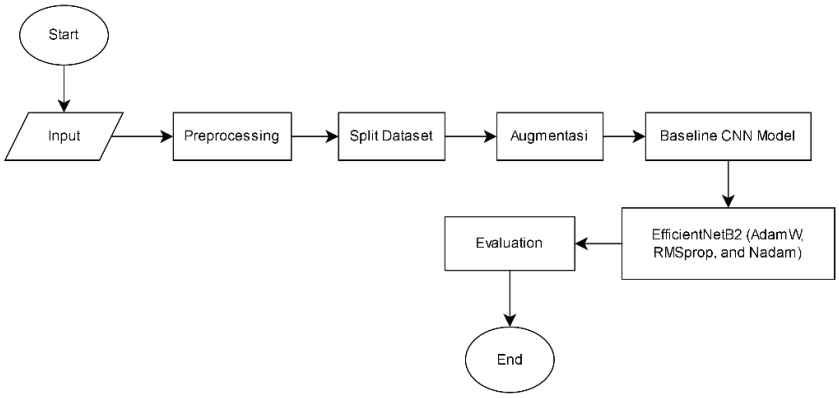
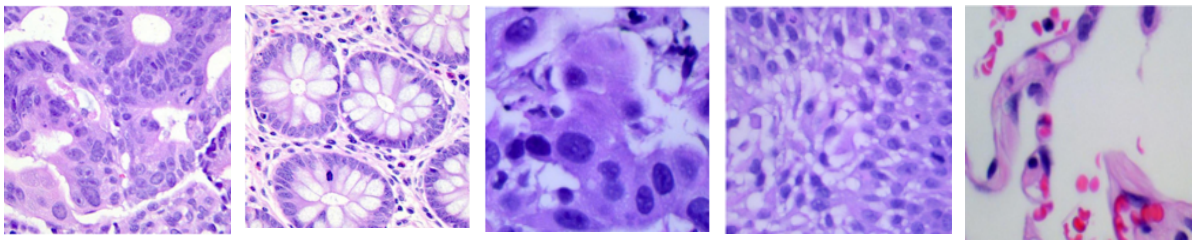


Figure 1. Research Flowchart

2.1. Dataset

The image collection used in the research consists of two categories of histopathological samples that represent conditions affecting both respiratory and intestinal tissues. Obtained from Kaggle, the dataset contains 25,000 images in total. These images are divided uniformly into five classes, each consisting of 5,000 entries. Illustrative examples from each class are provided in Figure 2.



(a) (b) (c) (d) (e)

Figure 2. Lung and Colon Class Datasets
(a) Colon Adenocarcinoma, (b) Colon Benign Tissue, (c) Lung Adenocarcinoma, (d) Lung Squamous Cell Carcinoma, (e) Lung Benign Tissue

The LC25000 dataset is a publicly available histopathological image dataset that was constructed through extensive augmentation from a smaller set of original image patches. As a result, some images in the dataset exhibit a high degree of visual similarity. This characteristic has been highlighted in recent studies as a potential source of data leakage when augmented images derived from the same original source appear across different data splits. Therefore, an appropriate data splitting strategy is required to ensure a fair and unbiased model evaluation.

2.2. Preprocessing and Split Dataset

Preprocessing is the initial process in the research method that aims to prepare raw image data into clean, structured, and ready-to-use image data in the model training process [20]. Each image was initially resized to 260 × 260 pixels to align with the EfficientNetB2 architecture and to lower computational demands while preserving essential morphological features. Data split is the stage of dividing the dataset into several parts, namely training, validation, and testing so that the model can be trained and adjusted properly [21]. The data collection was later partitioned into three segments, allocating 80% for model training, 10% for validation purposes, and the final 10% for performance testing, as summarized in Table 1.

To minimize the risk of data leakage caused by correlated augmented images, the dataset splitting process was performed prior to the augmentation stage. Under this strategy, all augmentation procedures were applied exclusively to the training data. The validation and testing sets consisted only of non-augmented images, allowing the evaluation results to better reflect the

model's generalization capability. Although this approach cannot completely eliminate inherent similarities within the LC25000 dataset, it significantly reduces cross-split correlation and aligns with best practices in medical image classification studies.

Table 1. Result from the Split Dataset

Class	Training	Validation	Testing	Total
Colon Adenocarcinoma	3979	504	517	5000
Colon Benign Tissue	4000	514	486	5000
Lung Adenocarcinoma	4011	488	501	5000
Lung Benign Tissue	4015	503	482	5000
Lung Squamous Cell Carcinoma	3995	491	514	5000
Total	20000	2500	2500	25000

2.3. Augmentation Dataset

Data augmentation was applied only to the training dataset to increase data variability while preserving class labels. This strategy is designed to minimize overfitting and enhance the model's ability to perform effectively on previously unseen data, without influencing the validity of the validation and testing procedures [21]. The augmentation techniques used in this study involve applying rotations, shifting the images horizontally and vertically, applying shear transformations, adjusting the zoom level, performing horizontal flips, and using a designated fill mode. Augmentation methods based on geometric transformations have been proven to be able to improve medical classification models without adding real data.

2.4. CNN Baseline Model

A CNN is a neural model structured to interpret visual data arranged in two dimensional form. In this study, the baseline CNN is assembled using a standard sequential design, consisting of several essential parts among them feature extracting convolution operations, dimensionality-reducing pooling layers, and dense layers that integrate the learned representations [22].

2.5. EfficientNetB2 Model

EfficientNetB2 is part of the EfficientNet family, a CNN-based model that uses a compound scaling strategy to expand its overall capacity in a balanced and coordinated way. This scaling method keeps the model efficient in computation, as well as improving accuracy [23]. This study utilizes three AdamW optimizers, RMSprop, and Nadam, tested separately by ensuring each training starts from the same model. The training process lasted for 20 epoches and used EarlyStopping and ReduceLROnPlateau to prevent overfitting and adjust the learning pace. After the training phase concludes, each architecture undergoes assessment using the test set to reveal how the three optimization approaches vary in terms of their overall effectiveness. The EfficientNetB2 model consists of several important elements.

1. Mobile Inverted Bottleneck Convolutional (MBConv) is a type of convolution block created to speed up and streamline the process of taking features on images. In it there are three main stages, namely expansion of the number of channels (expansion), processing with depthwise convolution, and the stage of resimplification through the projection layer [24].
2. The Squeeze-Excitation (SE) Block is a technique designed to emphasize the most critical features within an image, enabling the network to give priority to information that holds meaningful significance. In the SE block, the process involves three main stages: extracting the average feature values using Global Average Pooling, passing them through two fully connected layers, and finally recalibrating the feature weights. With these flows, SE Block helps the network recognize more meaningful features, so that extraction results are more accurate and efficient [25].
3. Swish is an activation function that works by multiplying a value by its own sigmoid. The way it works makes the flow of information in the network smoother because small negative values are not eliminated immediately. This approach enables the model to learn patterns more

smoothly and often yields better performance compared to conventional activation functions such as ReLU. The Swish activation function employed in this study is defined in equation 1 [26].

$$Swish(x) = x \cdot \sigma(x) \quad (1)$$

Where :
x = Input value
σ = Sigmoid function
e = Exponential value

4. Batch Normalization functions as a technique that helps maintain stable learning dynamics by adjusting the activation outputs within a mini-batch to a consistent range. The process involves calculating the mean values and variances of a batch, then normalizing it so that the data is at the same scale, and ending with a transformation using learning parameters to keep the network flexible. With these steps, training becomes faster and the model more easily achieves convergence [27].
5. Global Average Pooling (GAP) works by flattening all the values in a single feature map so that each channel leaves only one number. This operation streamlines the CNN's output without introducing additional weights and is typically applied before the classification stage to maintain the model's efficiency. The Global Average Pooling (GAP) formulation used in this study is presented in equation 2.

$$GAP = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} X_c(i,j) \quad (2)$$

Where:
$X_c(i,j)$ = Activation value on the c channel on a pixel
H = Feature map height
W = Width of the feature map
$\sum_{i=1}^{H}\sum_{j=1}^{W} X_c(i,j)$ = Sum of all pixel values in a channel

6. The Fully Connected Layer acts as the concluding component where the extracted representations are combined. At this stage, each unit is extensively linked to the units in the preceding layer, enabling the network to generate its final output by identifying the class that best matches the input image. The formulation of the fully connected layer used in this study is defined in equation 3.
Output layer:

$$y = Wx + b \quad (3)$$

Where:
y = The vector fed into the model
W = A weight matrix that links each input neuron to the corresponding output neuron
x = Input vector
b = Vector bias

2.6. Training

The model training process is carried out by calculating prediction errors using Categorical Cross-Entropy loss (CCE), then the network is updated through backpropagation to minimize the loss value. CCE is commonly applied in classification tasks because it provides stable and dependable gradients, enabling effective parameter updates throughout training [28]. The mathematical formulation of CEE used in this study is presented in equation 4.

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (4)$$

Where:
L = Value of loss

C = Number of classes in the classification
y_i = Actual label for class i
$\hat{y}_i$ = Predictive probability of the model
$\log(\hat{y}_i)$ = Logarithm of the prediction probability

2.7. Optimizer

Selecting an appropriate optimization method greatly influences how well the learning procedure progresses. Therefore, this study uses various types of optimizers, namely AdamW, RMSprop, and Nadam AdamW is a momentum-based adaptive optimizer that utilizes the first and second moment estimation to stabilize the weight update. Momentum-based optimizers like AdamW have an inertial pattern on the gradient so that it can accelerate convergence. The estimated AdamW moment is calculated as $m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$ and the adaptive variance is updated by $v_t = \beta_2 m_{t-1} + (1-\beta_2) g_t^2$ then a bias correction is performed with $m_t \leftarrow \frac{m_t}{\alpha_1 \mu_t}$ and $v_t \leftarrow \frac{v_t - m_t^2}{\alpha_2 \sigma_t}$ [29]. The update of the AdamW parameter uses a form of decoupled weight decay, namely:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \varepsilon} - \eta \lambda \theta_t \tag{5}$$

Where:
θ_t = Parameter of the model
η = Learning rate
m_t = Average momentum
v_t = The average of the squared gradients
ε = Small value
λ = Weight decay

RMSProp is an adaptive optimization method developed to address the issue of vanishing or unstable learning rates often encountered in SGD. In this technique, the learning rate for each parameter is continuously updated according to the averaged squared gradients over time. In this way, parameters that often result in large gradients will get smaller update steps, while parameters with small gradients will have larger update steps so that the learning process becomes more stable [30]. RMSprop has two basic mechanisms, namely:

1. Mean Squared Gradient

$$v_t = \beta v_{t-1} + (1-\beta) g_t^2 \tag{6}$$

Where:
v_t = The mean value of the gradients raised to the second power
β = Coefficient
v_{t-1} = The average of the squared gradients from the previous step
g_t^2 = Gradient square.

2. Weight update

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_t} + \varepsilon} g_t \tag{7}$$

Where:
w_t = Weight of the model
w_{t+1} = Weight that has been updated
η = Learning rate
v_t = Average value of the square of the gradient
ε = Small value to prevent division by zero
g_t = Gradient
$\frac{\eta}{\sqrt{v_t}+\varepsilon}$ = Adjustment factor to be stable when the gradient is large or small

Nadam is an adaptive optimizer that combines Nesterov momentum with an adaptive learning rate adjustment scheme like Adam, and allows parameter updates that bypass gradient predictions. Mathematically, Nadam updates its initial momentum term as $m_t = \mu_t m_{t-1} + (1-\mu_t)g_t$, and its secondary variance estimate as $n_t = v n_{t-1} + (1-v)g_t^2$ and then the bias correction by updating the parameters as follows [31].

$$\theta_t = \theta_{t-1} - \alpha_t \frac{\hat{n}_t}{\sqrt{\hat{n}_t} + \varepsilon} \hat{m} \tag{8}$$

Where:

g_t = Gradient
m_t = Momentum estimate
n_t = Second moment estimate
$\hat{n}_t$ = Bias correction
α_t = Learning rate
ε = Constant

In this study, three modern optimization algorithms AdamW, RMSprop, and Nadam were applied independently to the EfficientNetB2 architecture to ensure a fair and consistent performance comparison. All optimizers were trained using the same initial learning rate of 0.0001, allowing performance differences to be attributed solely to the weight update mechanisms of each optimizer. The training process was conducted for 20 epochs, employing two callback mechanisms, namely EarlyStopping and ReduceLROnPlateau, to maintain training stability and reduce the risk of overfitting. The EarlyStopping mechanism was set to track changes in validation error, allowing training to continue for up to ten additional epochs before termination and reverting the network parameters to those associated with the optimal validation outcome. In parallel, the ReduceLROnPlateau scheduler dynamically adjusted the step size by halving it whenever the validation performance stagnated for five successive epochs, while enforcing a lower bound of 0.00001 on the learning rate.

To ensure the independence of each experiment, the EfficientNetB2 model was re-cloned prior to training with each optimizer, preventing the carryover of learned weights from previous training runs. Upon completion of the learning process, performance assessment was conducted on the test set, with the reported outcomes derived from the parameter configuration that achieved the highest validation performance. This approach ensures that the evaluation reflects the optimal model performance and enhances the reliability and reproducibility of the experimental results.

2.8. Evaluation

During evaluation, predicted outputs are reviewed against ground-truth categories, allowing accuracy to be quantified and indicating whether further refinement may be required. This stage becomes especially important when dealing with imbalanced datasets, as the choice of evaluation metrics greatly influences how performance is interpreted and certain metrics may provide overlapping insights [32]. In this study, the evaluation of the model is carried out using a matrix-based measurement framework that yields metrics such as overall accuracy, precision levels, recall rate, and the computed F1 value. The table contains four essential outcomes: samples recognized as belonging to the positive group (TP), data points assigned accurately to the negative group (TN), entries mistakenly placed in the positive category (FP), and observations wrongly assigned to the negative category (FN) [33].

$$Accuracy\ (\%) = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \tag{9}$$

$$Precision\ (\%) = \frac{T_p}{T_p + F_p} \times 100 \tag{10}$$

$$Recall\ (\%) = \frac{T_p}{T_p + F_n} \times 100 \tag{11}$$

$$F1\ score\ (\%) = 2\frac{P \cdot R}{P + R} \tag{12}$$

3. RESULTS AND DISCUSSION

The effectiveness of the image-recognition approach was measured across five distinct classes. EfficientNetB2 was chosen for this study due to its architecture, which uses a compound scaling strategy to expand the network in a balanced manner, allowing for efficient feature extraction with fewer parameters compared to conventional CNN models. The fine-tuned model uses ImageNet pre-trained weights to speed up the process of adaptation to the characteristics of medical images. Table 2 presents the best results from training the EfficientNetB2 model with all three optimizers over 20 epoches

Table 2. Training Outcomes Related to Performance Metrics for Each Optimizer

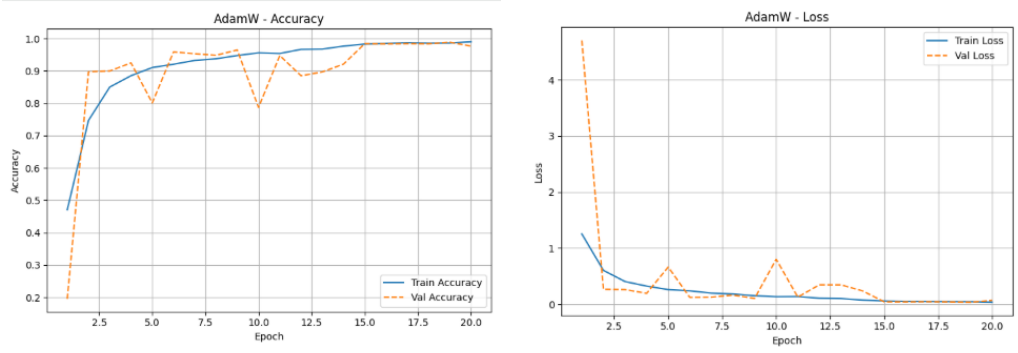
	Training		Validation		Testing	
Optimizer	Acc	Loss	Acc	Loss	Acc	Loss
AdamW	0.9888	0.0387	0.9888	0.0352	0.9856	0.0418
RMSprop	**0.9905**	**0.0328**	**0.9916**	**0.0282**	**0.9872**	**0.0335**
Nadam	0.9851	0.0491	0.9844	0.0442	0.9832	0.0465

Based on the results presented in Table 2, RMSProp demonstrates the highest overall capability, reaching 99.05% during training, 99.16% on the validation set, and 98.72% on the test set, while also producing the smallest loss value compared to the other models. This shows that RMSProp is able to achieve convergence faster while maintaining stable generalizations. AdamW also showed good performance, while Nadam was slightly lower but remained in the high accuracy category.

Performance evaluation was conducted to see how differences in the weight adjustment mechanism on each optimizer affected training stability, convergence speed, and model generalization. The analysis was carried out using accuracy, loss, and confusion matrix values for each optimizer.

A. AdamW

AdamW showed a steady trend of increasing accuracy since the beginning of the training. The loss curve declined sharply in the first five epochs and stayed below 0.05. The adaptive learning rate mechanism allows the gradients to be adjusted more effectively, while the decoupled weight decay helps minimize overfitting, keeping the model stable throughout training, as illustrated in Figure 3.

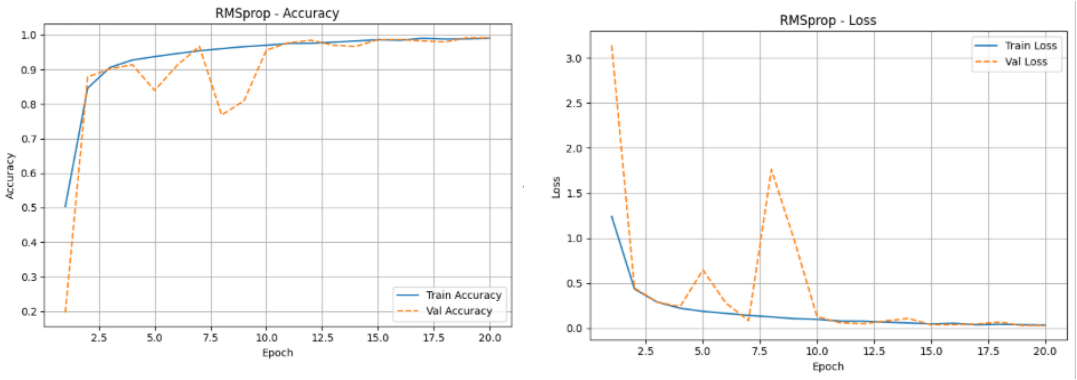


(a) (b)

Figure 3. Graph of the training results of the EfficientNetB2 model on the AdamW Optimizer (a) Accuracy Comparison, (b) Loss Comparison

B. RMSprop

RMSprop is the best-performing optimizer. The accuracy curve has been steadily increasing and reaching stability after the 10th epoch. Losses decreased the fastest and showed high stability, both in training data and validation seen in Figure 4. The lowest error generalization value indicates that RMSprop can optimally balance convergence speed and generalization capabilities.

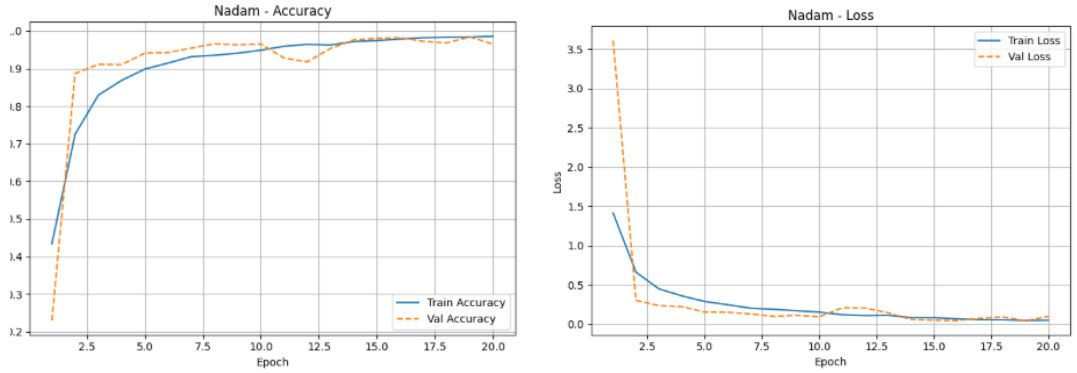


(a) (b)

Figure 4. Graph of the training results of the EfficientNetB2 model on the RMSprop Optimizer (a) Accuracy Comparison, (b) Loss Comparison

C. Nadam

Nadam achieved high initial accuracy thanks to Nesterov's momentum, but the validation graph showed slight fluctuations in the middle of the epoch, signaling sensitivity to gradient variations. Losses are consistently decreasing, although the value of the validation loss is slightly higher than AdamW and RMSprop seen in Figure 5. This shows that the convergence stability of Nadam is still lower in data with complex patterns such as histopathological images.



(a) (b)

Figure 5. Graph of the training results of the EfficientNetB2 model on the Nadam Optimizer (a) Accuracy Comparison, (b) Loss Comparison

Based on the performance analysis of the three optimizers, it can be seen that the difference in performance is influenced by the way each optimizer adjusts the learning rate and handles unstable gradient dynamics. Nadam was indeed able to achieve high accuracy quickly, but showed fluctuations in weight updates in the early phases of training. In contrast, RMSprop provides the best balance between stability, convergence speed, and generalization capabilities, especially in histopathological images that have complex and varied structures. The latest gradient-based learning rate adjustment mechanism makes it more consistent than AdamW and Nadam. These

findings confirm that RMSprop is the most appropriate choice for the EfficientNetB2 architecture in this classification.

In this study, model performance evaluation is not limited to accuracy and loss values but is further analyzed using precision, recall, and F1-score for each class. This evaluation strategy is particularly important in medical image classification tasks, where accuracy alone may not adequately reflect class-wise predictive reliability. Tables X, Y, and Z present the class-wise evaluation results on the test dataset for the AdamW, RMSprop, and Nadam optimizers, respectively. All metrics were obtained from the classification reports generated based on the model predictions on the test set.The results indicate that RMSprop delivers the most consistent performance across all classes, achieving high and balanced precision, recall, and F1-score values. The Colon Adenocarcinoma and Lung Benign Tissue classes were classified perfectly (F1-score = 1.00), while misclassifications among lung cancer classes were lower compared to the other optimizers.

The AdamW optimizer also demonstrates strong performance; however, a slight decrease in F1-score is observed in the Lung Adenocarcinoma class, suggesting confusion with Lung Squamous Cell Carcinoma due to morphological similarities. Meanwhile, Nadam achieves relatively high overall metrics but shows reduced stability in distinguishing lung cancer classes, although benign tissues are consistently identified with high accuracy. These findings confirm that class-wise evaluation metrics provide a more comprehensive assessment of model performance and further support the conclusion that EfficientNet-B2 combined with RMSprop offers the most effective configuration for histopathological image classification in this study.

Table 3. Classification Report of EfficientNetB2 Using the AdamW Optimizer

Class	Precision	Recall	F1-score	Support
Colon Adenocarcinoma	1.00	1.00	1.00	517
Colon Benign Tissue	1.00	1.00	1.00	486
Lung Adenocarcinoma	0.99	0.94	0.96	482
Lung Benign Tissue	1.00	1.00	1.00	514
Lung Squamous Cell Carcinoma	0.94	0.99	0.97	501
Accuracy			0.99	2500
Macro avg	0.99	0.99	0.99	2500
Weighted avg	0.99	0.99	0.99	2500

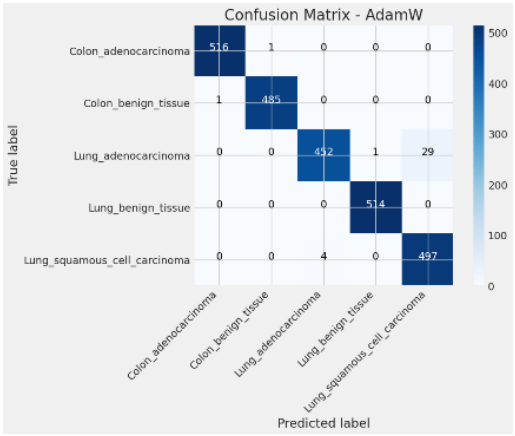


Figure 6. Classification Outcome Table for the AdamW Optimizer

Table 4. Classification Report of EfficientNetB2 Using the RMSprop Optimizer

Class	Precision	Recall	F1-score	Support
Colon Adenocarcinoma	0.98	1.00	0.99	517
Colon Benign Tissue	1.00	0.98	0.99	486
Lung Adenocarcinoma	0.98	0.98	0.98	482
Lung Benign Tissue	1.00	1.00	1.00	514
Lung Squamous Cell Carcinoma	0.98	0.98	0.98	501
Accuracy			0.99	2500
Macro avg	0.99	0.99	0.99	2500
Weighted avg	0.99	0.99	0.99	2500

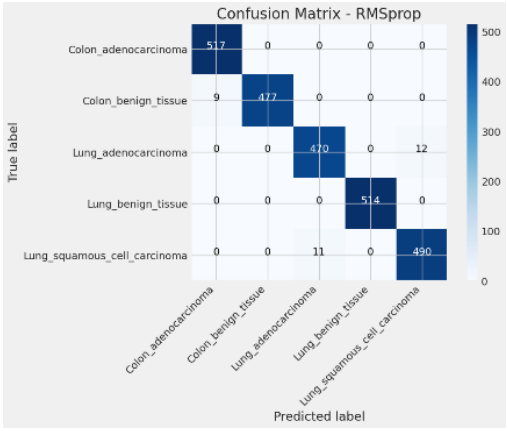


Figure 7. Classification Outcome Table for the RMSprop Optimize*r*

Table 5. Classification Report of EfficientNetB2 Using the Nadam Optimizer

Class	Precision	Recall	F1-score	Support
Colon Adenocarcinoma	1.00	0.98	0.99	517
Colon Benign Tissue	0.99	1.00	0.99	486
Lung Adenocarcinoma	0.96	0.96	0.96	482
Lung Benign Tissue	1.00	1.00	1.00	514
Lung Squamous Cell Carcinoma	0.97	0.97	0.97	501
Accuracy			0.98	2500
Macro avg	0.98	0.98	0.98	2500
Weighted avg	0.98	0.98	0.98	2500

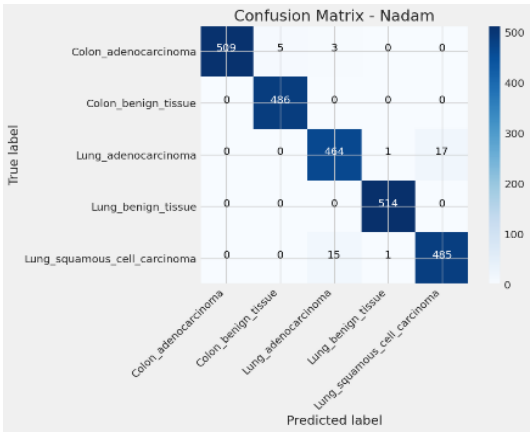


Figure 8. Classification Outcome Table for the Nadam Optimizer

Based on the evaluation results incorporating both the confusion matrix and class-wise precision, recall, and F1-score metrics, the EfficientNetB2 model demonstrates strong classification performance across five histopathological image categories. Overall, the dominance of predictions along the main diagonal of the confusion matrix, supported by high precision and recall values across most classes, indicates that both cancerous and healthy tissue patterns are consistently recognized by the model. The AdamW optimizer delivers strong performance with high precision, recall, and F1-score values for almost all classes, particularly for benign tissue categories. However, a slightly lower F1-score is observed for the Lung Adenocarcinoma class, suggesting misclassification with Lung Squamous Cell Carcinoma due to the high morphological similarity between these two lung cancer types.

The RMSprop optimizer achieves the most consistent performance, characterized by high and well-balanced precision, recall, and F1-score values across all classes. The Colon Adenocarcinoma and Lung Benign Tissue classes are classified perfectly, while misclassifications among lung cancer classes are lower compared to AdamW and Nadam. These results indicate that

RMSprop is more effective in optimizing complex features in histopathological images and provides better generalization stability. Meanwhile, Nadam also attains high overall accuracy but exhibits slightly lower precision and F1-score values in several lung tissue categories. Although benign classes are identified reliably, the misclassification rate among lung cancer classes remains higher than that of the other two optimizers. This suggests that while Nadam converges rapidly, its stability in distinguishing highly similar tissue patterns is still suboptimal.

4. CONCLUSION

This research seeks to determine which optimization method offers the greatest improvement for the EfficientNet-B2 architecture in performing histopathological image recognition, as outlined in the Introduction. The results of the analysis in Results and Discussion show that this goal was successfully achieved through a comprehensive evaluation of AdamW, RMSprop, and Nadam.The three optimizers were able to provide good classification results, but differences in learning rate adaptation mechanisms caused clear variations in performance.

RMSprop shows the most stable performance with the least number of misclassifications and the best generalization of test data. AdamW still provides high accuracy, but faces obstacles in distinguishing types of lung cancer with similar morphological characteristics. Meanwhile, Nadam showed a rapid but less stable convergence in handling classes of networks that have similar patterns. Overall, the combination of EfficientNetB2 and RMSprop met the research expectations to produce a stable, efficient, and accurate model in identifying five classes of histopathological images. These findings prove the compatibility between the initial expectations of the research and the results obtained in the discussion section.

This study also highlights opportunities for future work, such as incorporating model interpretability techniques, exploring more diverse datasets, and integrating alternative architectures like attention mechanisms or hybrid CNN Transformer models. These advancements are expected to enhance the effectiveness of medical image based diagnostic systems and broaden their applicability to other clinical tasks in the future.

ACKNOWLEDGEMENTS

The authors extend their heartfelt appreciation to FAST Ahmad Dahlan University, Indonesia, especially the Mathematics Laboratory, for the support, facilities, and meaningful assistance that greatly contributed to the completion of this research.

REFERENCES

[1] Y. Wu *et al.*, “Recent Advances of Deep Learning for Computational Histopathology: Principles and Applications,” Mar. 01, 2022, *MDPI*. doi: 10.3390/cancers14051199.

[2] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, “Deep Convolutional Neural Networks in Medical Image Analysis: A Review,” Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/info16030195.

[3] W. Yue, S. Liu, and Y. Li, “Eff-PCNet: An Efficient Pure CNN Network for Medical Image Classification,” *Applied Sciences (Switzerland)*, vol. 13, no. 16, Aug. 2023, doi: 10.3390/app13169226.

[4] K. Khotimah, S. Surono, and A. Thobirin, “Optimizing EfficientNet for imbalanced medical image classification using grey wolf optimization,” *Computer Science and Information Technologies*, vol. 6, no. 2, pp. 112–121, 2025, doi: 10.11591/csit.v6i2.pp112-121.

[5] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” Sep. 2020, [Online]. Available: http://arxiv.org/abs/1905.11946

[6] M. W. Purbandanu, R. Yanuarta, and A. Kurniawan, “Optimization of Skin Cancer Detection to Improve Accuracy with the Application of Efficient Convolutional Neural Network and EfficientNetB2 Models,” *Journal of Intelligent Computing & Health Informatics*, vol. 5, no. 2, Apr. 2024, doi: 10.26714/jichi.v5i2.14338.

[7] M. Latha, P. S. Kumar, R. R. Chandrika, T. R. Mahesh, V. V. Kumar, and S. Guluwadi, “Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and Explainable AI,” *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01404-3.

[8] K. Natarajan *et al.*, "A novel method for the detection and classification of multiple diseases using transfer learning-based deep learning techniques with improved performance," *Neural Comput Appl*, vol. 36, no. 30, pp. 18979–18997, Oct. 2024, doi: 10.1007/s00521-024-09900-x.

[9] M. Alruwaili and M. Mohamed, "An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification," *Diagnostics*, vol. 15, no. 5, Mar. 2025, doi: 10.3390/diagnostics15050551.

[10] H. Sun *et al.*, "An Improved Medical Image Classification Algorithm Based on Adam Optimizer," *Mathematics*, vol. 12, no. 16, Aug. 2024, doi: 10.3390/math12162509.

[11] A. M. J. M. D. Zubair Rahman, R. Mythili, K. Chokkanathan, T. R. Mahesh, K. Vanitha, and T. E. Yimer, "Enhancing image-based diagnosis of gastrointestinal tract diseases through deep learning with EfficientNet and advanced data augmentation techniques," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01479-y.

[12] G. Latif, J. Alghazo, M. A. Khan, G. Ben Brahim, K. Fawagreh, and N. Mohammad, "Deep convolutional neural network (CNN) model optimization techniques—Review for medical imaging," 2024, *American Institute of Mathematical Sciences*. doi: 10.3934/math.2024998.

[13] X. Wang and L. Aitchison, "How to set AdamW's weight decay as you scale model and dataset size," Jun. 2025, [Online]. Available: http://arxiv.org/abs/2405.13698

[14] N. Shi, D. Li, M. Hong, and R. Sun, "RMSPROP CONVERGES WITH PROPER HYPER-PARAMETER."

[15] S. Mulla and N. F. Shaikh, "Over comparative study of text summarization techniques based on graph neural networks," *Web Intelligence*, vol. 22, no. 2, pp. 231–248, Apr. 2024, doi: 10.3233/WEB-230014.

[16] Z. Zhuang, M. Liu, A. Cutkosky, and F. Orabona, "Understanding AdamW through Proximal Methods and Scale-Freeness," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2202.00089

[17] Q. Zhang, Y. Zhou, and S. Zou, "Convergence Guarantees for RMSProp and Adam in Generalized-smooth Non-convex Optimization with Affine Noise Variance," Mar. 2025, [Online]. Available: http://arxiv.org/abs/2404.01436

[18] R. Abdulkadirov, P. Lyakhov, and N. Nagornov, "Survey of Optimization Algorithms in Modern Neural Networks," Jun. 01, 2023, *MDPI*. doi: 10.3390/math11112466.

[19] S. Anjum *et al.*, "Lung Cancer Classification in Histopathology Images Using Multiresolution Efficient Nets," *Comput Intell Neurosci*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/7282944.

[20] N. Guetta, A. Shabtai, I. Singh, S. Momiyama, and Y. Elovici, "Dodging Attack Using Carefully Crafted Natural Makeup," Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.06467

[21] H. Naveed, S. Anwar, M. Hayat, K. Javed, and A. Mian, "Survey: Image Mixing and Deleting for Data Augmentation," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2106.07085

[22] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[23] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sep. 2020, [Online]. Available: http://arxiv.org/abs/1905.11946

[24] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2104.00298

[25] J. Su, S. Anderson, and L. Mihaylova, "Holistic Self-Distillation with the Squeeze and Excitation Network for Fine-grained Plant Pathology Classification," in *2023 26th International Conference on Information Fusion, FUSION 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.23919/FUSION52260.2023.10224184.

[26] L. Deng, H. Suo, and D. Li, "Deepfake Video Detection Based on EfficientNet-V2 Network," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/3441549.
[27] H. Daneshmand, A. Joudaki, E. Zurich, and F. Bach, "Batch Normalization Orthogonalizes Representations in Deep Random Networks." [Online]. Available: https://github.com/hadidaneshmand/batchnorm21.git
[28] S. Rajaraman, G. Zamzmi, and S. K. Antani, "Novel loss functions for ensemble-based medical image classification," *PLoS One*, vol. 16, no. 12 December, Dec. 2021, doi: 10.1371/journal.pone.0261307.
[29] P. Zhou, X. Xie, Z. Lin, K.-C. Toh, and S. Yan, "Win: Weight-Decay-Integrated Nesterov Acceleration for Faster Network Training," 2024. [Online]. Available: https://github.com/sail-sg/win.
[30] R. Abdulkadirov, P. Lyakhov, and N. Nagornov, "Survey of Optimization Algorithms in Modern Neural Networks," Jun. 01, 2023, *MDPI*. doi: 10.3390/math11112466.
[31] J. Tu, A. Yasin, and N. S. Mansor, "NadamClip: A Novel Optimization Algorithm for Improving Prediction Accuracy and Training Stability," *Processes*, vol. 13, no. 7, Jul. 2025, doi: 10.3390/pr13072145.
[32] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices," *Remote Sens (Basel)*, vol. 16, no. 3, Feb. 2024, doi: 10.3390/rs16030533.
[33] N. Widya Utami, M. A. P. Putra, I. G. J. E. Putra, and G. A. Sampedro, "A CNN-Based Information System for Balinese Dance Classification with Hyperparameter Optimization," *International Journal of Advances in Data and Information Systems*, vol. 6, no. 2, pp. 376–390, Aug. 2025, doi: 10.59395/ijadis.v6i2.1402.