Joinforcs: Journal of informatics and information system
Volume 1, (1) 2026, p. 21-29
E-ISSN : XXXX-XXXX

Comparative Analysis of Adam and RMSprop Optimizer on Artificial Neural Network for Clinical Data-Based Classification of Lung Cancer

Alpin Danuarta [a,1,*] , Dewi Oktafiani [a,2]
[a] Informatics Study Program STMIK Amikom, Surakarta, Indonesia
[1] alpin.10386@mhs.amikomsolo.ac.id, [2] dewioktafiani@dosen.amikomsolo.ac.id
* corresponding author

Article Info

Article History:

Received Dec 24, 2025
Revised Jan 17, 2026
Accepted Jan 31, 2026

Keywords:
Adam Optimizer, Artificial Neural Network, Lung Cancer, Machine Learning, RMSprop

Abstract

Lung cancer is one of the leading causes of global death that demands early and precise detection. Advances in Artificial Intelligence technology, especially Artificial Neural Networks (ANNs), can support the disease classification process through medical data. This study focused on evaluating the performance of the ANN model in classifying lung cancer as well as comparing the impact of the Adam and RMSprop optimizers on model performance. The data used was in the form of clinical records of lung cancer patients in CSV format as many as 310 entries. The research steps include pre-processing of data, normalization with MinMaxScaler, division of training and test data with an 80:20 ratio, construction of ANN models using TensorFlow and Keras, as well as assessment through confusion matrix, accuracy, precision, recall, F1-score, and ROC-AUC. The ANN structure includes an input layer, two hidden layers with a ReLU activation function, and an output layer with sigmoids for binary classification. The findings showed that Adam's optimizer delivered the best results with 97 percent accuracy, 98 percent accuracy, 98 percent recall, and nearly 1 AUC. While the RMSprop optimizer produces 95% accuracy. The findings confirm that optimizer selection affects the performance of ANN classification on lung cancer data. It is hoped that this research can be a reference in the development of a medical decision system based on Artificial Intelligence.

1. Introduction

Lung cancer is one of the leading causes of death worldwide and is a global health challenge that continues to increase every year [1]. According to World Health Organization (WHO) data, lung cancer is the first cause of cancer death with more than 1.8 million deaths each year [2]. This high mortality rate is mainly due to late diagnosis, as early symptoms of lung cancer are often difficult to identify clinically. As a result, many new cases are only detected at an advanced stage, so the chances of successful treatment are smaller [3]. Therefore, a diagnostic support system is needed that can help the disease classification process quickly, accurately, and efficiently to improve the quality of health services [4].

Advances in Artificial Intelligence (AI) *technology*, especially Machine Learning and Deep Learning, have had a significant impact on the field of digital health. One of the techniques that is often used to classify diseases is [1], [5] the Artificial Neural Network (ANN). ANN is a computational model that mimics how human neural networks work in studying complex data patterns as well as non-linear relationships between variables. These advantages make [6] ANN very effective for processing medical data, because health data is usually multidimensional, unstructured, and complex.[4]

Various previous studies indicate that artificial neural networks (ANNs) can provide satisfactory results in classifying lung cancer. The [6] ANN *model* managed to obtain an accuracy rate of more than 90% in the classification of lung cancer medical data, and meanwhile the performance of [4] the ANN was highly dependent on the optimization algorithm used during the model's training. [7] The role of the optimizer is to update the network weight so that errors can be optimally reduced, which makes the model convergence process more stable and fast[8].

Based on the latest developments in state-of-the-art research, ANN has been widely applied in lung cancer classification systems both using medical imaging data and numerical clinical data. Previous research results show that ANN is able to improve diagnostic accuracy compared to conventional methods and classical machine learning algorithms. In addition, the use of deep learning frameworks such as TensorFlow and Keras makes it easier to develop ANN models with stable and efficient performance [9].

Although ANN has been shown to be effective, the model's performance is heavily influenced by the optimization algorithms used in the training process. The optimizer functions to update the network weight so that errors can be optimally minimized. Some commonly used optimizers are Adam and RMSprop, which each have different characteristics in convergence speed and training stability. Previous research has shown that the right selection of optimizers can improve the accuracy and generalization capabilities of ANN models [10].

Adam and RMSprop are two of the most popular optimization methods in training deep learning networks. Adam stands out for combining adaptive learning rate with momentum, thus accelerating model convergence. In contrast, [11] RMSprop is designed to maintain the stability of weight updates on fickle and non-stationary data. Although these two optimizers are widely used, their performance in various studies still shows variation depending on the characteristics of the dataset and the model architecture used.

According to previous studies, the majority of studies only focused on improving the accuracy of artificial neural networks (ANNs) without thoroughly assessing the impact of optimizers in the classification of lung cancer based on numerical clinical data [12]. In addition, a number of studies still experience limitations in experimental validation, data imbalance (class imbalance), and the use of incomplete evaluation metrics. In fact, in the context of medical data classification, class imbalances can make the model more inclined to the majority class thereby reducing the model's generalization capabilities [13].

This study introduces a more thorough methodology by comparing the performance of the Adam and RMSprop optimizers on artificial neural networks to classify lung cancer using clinical data. In addition to assessing the model through accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC metrics, this study also applies a class imbalance handling technique in the form of the Synthetic Minority Oversampling Technique (SMOTE) to improve the stability and generalization ability of the model. Therefore, the focus of the research is not solely on improving accuracy, but on assessing the overall performance of the model against unbalanced data.

Key contributions to this study include:

1. To conduct a comparison between the optimizer Adam and RMSprop in the lung cancer classification task using clinical data;
2. Using the SMOTE method to correct the problem of class imbalance in medical data; and
3. Assess the model with a multi-metric evaluation approach to improve the validity and completeness of the experiment results.

It is hoped that the findings of this study can be a reference in the creation of an Artificial Intelligence-based medical decision system and provide technical advice on the selection of the most optimal optimizer for artificial neural networks for the classification of lung cancer.

2. Research Methodology

This study uses an Artificial Intelligence-based quantitative approach with the Artificial Neural Network (ANN) method to classify lung cancer based on patient clinical data. The research methodology is designed in a structured manner starting from the process of collecting datasets, preprocessing data, building Artificial Neural Network models, to evaluating model performance using several classification metrics. This study also conducted a comparative analysis of the use of Adam [1], [3] and RMSprop optimizers to find out which optimizers provide the best performance in the lung cancer classification process. The following is the research flowchat that will be used in this study, which can be seen in the image below:

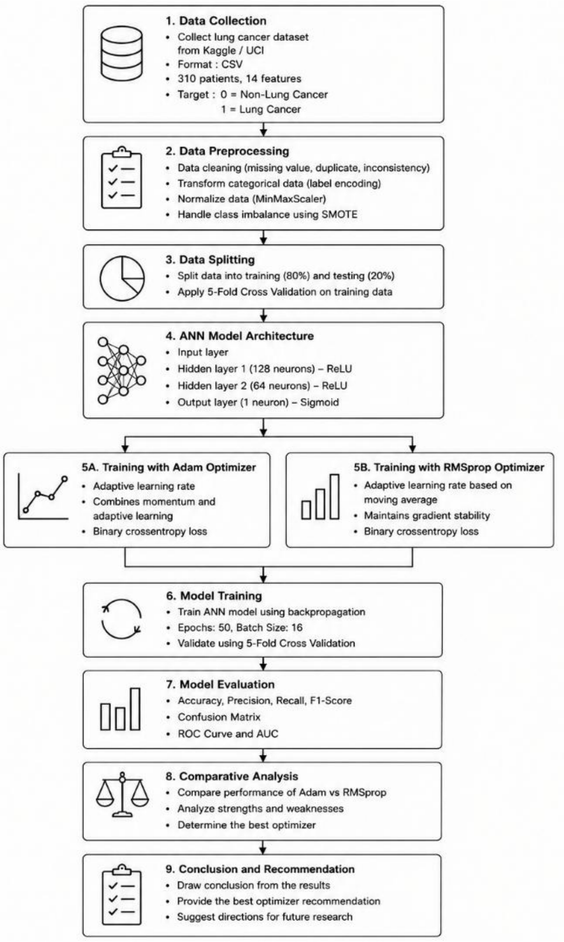


Figure 1. Reaserch Flowchat

The study began by collecting lung cancer datasets in CSV format downloaded from Kaggle and the UCI Machine Learning Repository. The data contains patient information along with a number of clinical attributes related to lung cancer risk, including age, smoking history, shortness of breath, chest pain, chronic cough, wheezing, alcohol consumption, fatigue, chronic disease, and other risk factors. The target variables in this study were divided into two classes: a positive class for lung cancer and a negative class for non-lung cancer [14].

The next step is to preprocess the data to improve the quality of the data set before it is used in model training. The preprocessing process includes data cleansing to eliminate missing values, duplicate data, as well as data mismatches. Then, the categorical data is converted into numerical with encoding labels [4] so that it can be processed by an artificial neural network model. After transformation, the data is normalized using the MinMaxScaler method so that the feature value range is 0 to 1, so that model training becomes more stable and optimal [11].

The dataset in this study experienced class imbalance, so it was handled using the Synthetic Minority Oversampling Technique (SMOTE) method. SMOTE generates synthetic data for minority classes, making the distribution of data more balanced and preventing biased models against the majority class. It is hoped that the use of SMOTE can improve the generalization ability of the model, especially in recognizing data in minority classes [2], [13].

After preprocessing, the dataset is divided into training data and test data with an 80:20 ratio. Eighty percent of the data is used to train the model, while the remaining twenty percent is used to test the model. This study also applies the 5-Fold Cross Validation technique to improve the validity of the experiment and reduce the risk of overfitting the classification model.

A classification model was developed by utilizing an artificial neural network that runs on top of TensorFlow and Keras. The model structure includes an inlet layer, two hidden layers, as well as an output layer. The hidden layer uses the ReLU activation function because it can handle non-linear problems efficiently, while the output layer uses the sigmoid activation function for binary classification. This model is trained through a backpropagation algorithm using two optimizers, namely Adam and RMSprop, to evaluate the impact of the optimizer on model performance. The training configuration includes 50 epochs, 16 batch sizes, and the loss binary crossentropy function [15].

The model's performance assessment uses a variety of classification metrics, including accuracy, precision, recall, F1 score, confusion matrix, Receiver Operating Characteristic (ROC), and Area Under Curve (AUC). Accuracy assesses the overall success rate of a model, while precision and recall assess the model's ability to detect positive data and reduce misclassification. F1 scores measure the balance between precision and recall. In addition, the ROC curve and AUC values were used to assess the model's ability to distinguish between positive and negative classes at various classification thresholds.

All stages of the experiment were carried out using the Python programming language as well as the TensorFlow and Keras frameworks on the Google Colab or Jupyter Notebook platforms. It is hoped that this research can contribute to the development of Artificial Intelligence-based medical decision support systems, especially in the classification of lung cancer by using a more efficient and organized Artificial Neural Network approach.

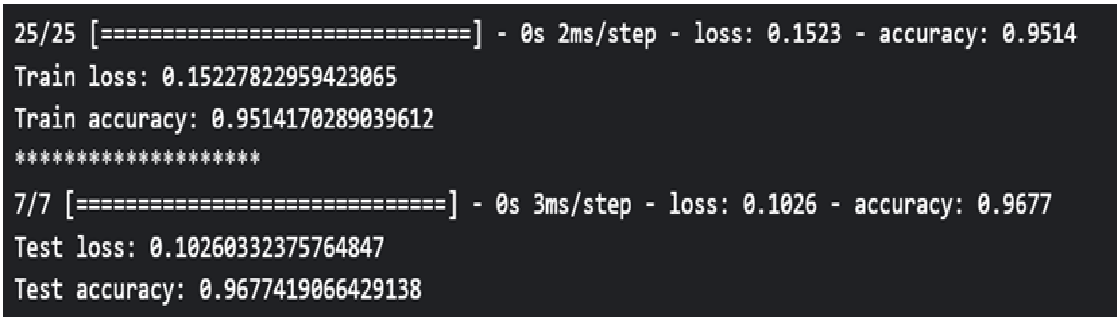
3. Results and Discussion

This study aims to analyze the performance of the Artificial Neural Network (ANN) model in the classification of lung cancer based on clinical data and compare the effect of the use of the Adam and RMSprop optimizers on model performance. All experiments were carried out using datasets that had gone through the stages of preprocessing, normalization, and sharing of training-testing data using an 80:20 ratio. Model performance evaluation was carried out using accuracy, loss, confusion matrix, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) Curve.

3.1Training Performance Analysis

A model training performance assessment was conducted to assess the extent to which the Artificial Neural Network could recognize patterns in lung cancer data during the training process. This study used two optimizers, Adam and RMSprop, to observe the impact of both on the stability and performance of the classification model.

a. Accuracy and Loss of Adam Optimizer

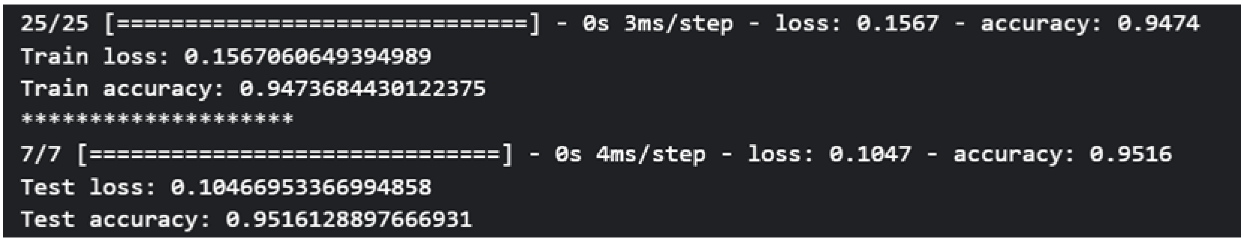


```
25/25 [==============================] - 0s 2ms/step - loss: 0.1523 - accuracy: 0.9514
Train loss: 0.15227822959423065
Train accuracy: 0.9514170289039612
********************
7/7 [==============================] - 0s 3ms/step - loss: 0.1026 - accuracy: 0.9677
Test loss: 0.10260332375764847
Test accuracy: 0.9677419066429138
```

Figure 2. Accuracy dan Loss Optimizer Adam

According to the results of the model training with the Adam optimizer shown in Figure 2, the training accuracy reached 95.14% with a loss value of 0.1523. In the test data, the accuracy was recorded at 96.77% and the loss was 0.1047. The high accuracy and low loss indicate that the model can optimally capture the patterns of relationships between features in the lung cancer dataset. In addition, the difference in accuracy between the training and test data is relatively small, indicating good generalization capabilities and no significant overfitting. Interestingly, the accuracy on the test data slightly exceeded the accuracy of the training. This may be due to the more understandable distribution of test data models or implicit regulations that optimizer Adam provides during training. This shows that the Adam optimizer can produce stable convergence on artificial neural networks.

b. Accuracy and Loss of RMSprop Optimizer



```
25/25 [==============================] - 0s 3ms/step - loss: 0.1567 - accuracy: 0.9474
Train loss: 0.1567060649394989
Train accuracy: 0.9473684430122375
********************
7/7 [==============================] - 0s 4ms/step - loss: 0.1047 - accuracy: 0.9516
Test loss: 0.10466953366994858
Test accuracy: 0.9516128897666931
```

Figure 3. Accuracy dan Loss Optimizer RMSprop

Training with RMSprop optimizer obtained a training accuracy of 94.74% and a training loss of 0.1567, while the test data obtained a test accuracy of 95.16% and a test loss of 0.1046.These findings indicate that RMSprop can provide satisfactory classification results with a fairly low rate of prediction error. However, when compared to Adam's optimizer, RMSprop shows slightly less accuracy as well as less consistent convergence over multiple training epochs. Overall, both optimizers performed very well in the ANN model training process. However, Adam provides more optimal results in updating the network weight, resulting in higher accuracy.

3.2 Confusion Matrix Analysis

A confusion matrix is used to assess the model's ability to classify each class of data. In this study, there are two classification classes, namely: Class 0: Do not have lung cancer and Class 1: Have lung cancer.

a. Confusion Matrix of Adam Optimizer

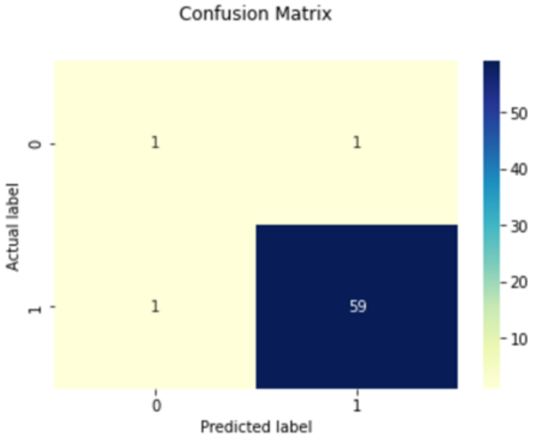


Figure 4. Accuracy dan Loss Optimizer RMSprop

The results of Adam's confusion matrix optimizer can be seen in Table 1.

Table 1. Confusion Matrix of Adam Optimizer

Actual Class	Predicted 0	Predicted 1
Current 0	1	1
Current 1	1	59

According to the test results, the model managed to accurately identify 59 lung cancer data (True Positive) and 1 non-cancer data (True Negative). In addition, there was 1 non-cancer data that was wrongly categorized as cancer (False Positive) and 1 cancer data that was wrongly categorized as non-cancerous (False Negative). A low False Negative ratio indicates that the model has high sensitivity in detecting lung cancer patients. In the medical context, the ability to detect positive cases is critical because misclassification in cancer patients can delay diagnosis and treatment. Based on the confusion matrix, the accuracy of the model is calculated using the equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{59 + 1}{59 + 1 + 1 + 1} = 96{,}77\%$$

These results show that the ANN model with the Adam optimizer has excellent classification performance.

b. Confusion Matrix of RMSprop Optimizer

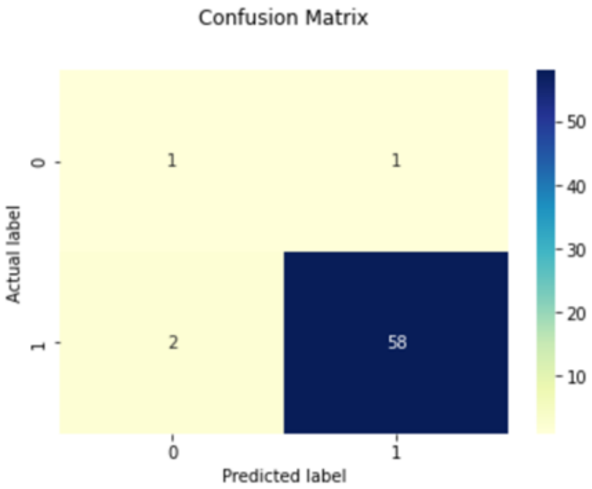


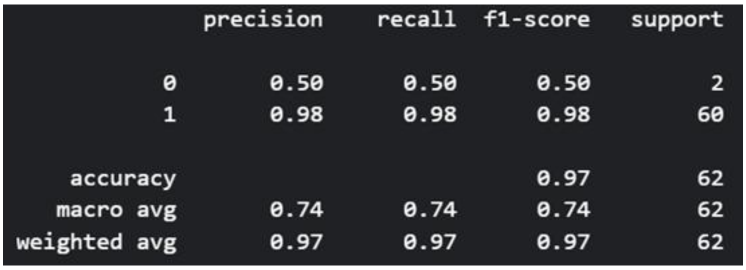
Figure 5. Confusion Matrix Optimizer RMSprop

Using the RMSprop optimizer, the model managed to accurately identify 58 cases of lung cancer and 1 case of non-cancer. However, there were 2 cancer cases that were erroneously classified as non-cancerous and 1 non-cancer case that was erroneously categorized as cancer. The accuracy reached 95.16%, which is still very good. Even so, the number of false negatives in RMSprop was higher compared to Adam's optimizer, which indicates a slightly lower sensitivity of RMSprop in detecting lung cancer patients. In addition, the confusion matrix shows a class imbalance in the research data. The sample of positive classes is much more than the negative class, so the model learns more patterns from the majority class.

3.3 Classification Report Analysis

Classification reports are used to evaluate the model's performance in more detail using precision, recall, F1-score, and support in each class.

a. Classification Report of Adam Optimizer.

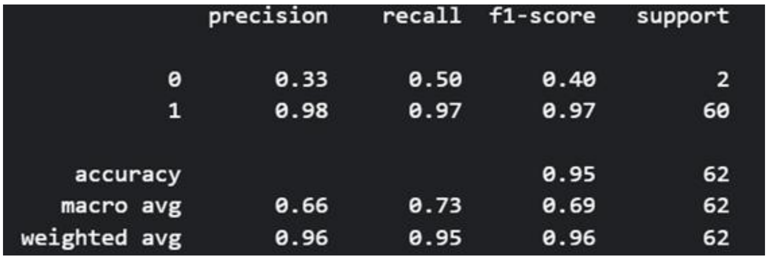


```
              precision    recall  f1-score   support

           0       0.50      0.50      0.50         2
           1       0.98      0.98      0.98        60

    accuracy                           0.97        62
   macro avg       0.74      0.74      0.74        62
weighted avg       0.97      0.97      0.97        62
```

Figure 6. Classification Report Optimizer Adam

In class 0 (non-cancer), the model obtained a precision of 0.50, a recall of 0.50, and an F1-score of 0.50 with the support of 2 samples. These figures indicate that the model still struggles to recognize minority classes due to data limitations. For class 1 (lung cancer), the model produced a precision of 0.98, a recall of 0.98, and an F1-score of 0.98 with the support of 60 samples. These results suggest that the model is highly effective in identifying patients with lung cancer. Overall, the accuracy of the model reached 97%, indicating a very high classification capability.

b. Classification Report of RMSprop Optimizer



```
              precision    recall  f1-score   support

           0       0.33      0.50      0.40         2
           1       0.98      0.97      0.97        60

    accuracy                           0.95        62
   macro avg       0.66      0.73      0.69        62
weighted avg       0.96      0.95      0.96        62
```

Figure 7. Classification Report Optimizer RMSprop

When using the RMSprop optimizer, the model obtained a precision value of 0.98 and a recall value of 0.97 for the lung cancer class. However, for the non-cancer class, the performance was still weak with a precision of 0.33 and an F1-score of 0.40.These findings indicate that data imbalance is the main cause of the decline in performance in the minority class. Although the overall accuracy of the model is quite high, the assessment through precision, recall, and F1-score metrics reveals that accuracy alone is insufficient to assess the medical classification model.

3.4ROC Curve Analysis

The ROC Curve is used to evaluate the model's ability to distinguish positive and negative classes at various classification thresholds.

a. ROC Curve of Adam Optimizer

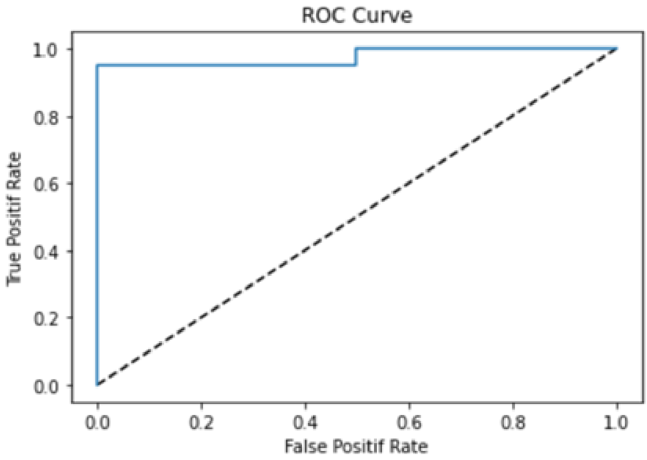


Figure 8. ROC Curve Optimizer Adam

The ROC curve generated with the Adam optimizer is almost attached to the upper left corner of the chart, indicating that the model is obtaining a high True Positive Rate (TPR) and a low False Positive Rate (FPR). In addition, the Area Under Curve (AUC) value appears to be close to 1, which indicates the model's excellent classification ability in distinguishing between lung cancer and non-lung cancer patients.

b. ROC Curve of RMSprop Optimizer

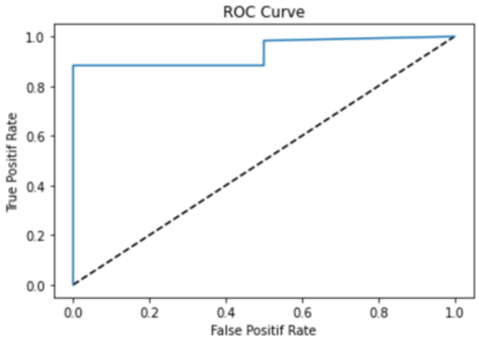


Figure 9. ROC Curve Optimizer RMSprop

The ROC curve generated by the RMSprop optimizer shows very satisfactory performance, with the curve almost reaching the upper left corner of the chart. However, when viewed visually, the performance of Adam's optimizer is still superior to RMSprop because it provides higher sensitivity and specificity values.

3.5 Comparative Discussion

Based on all the results of the experiments, Adam's optimizer showed superior performance to RMSprop in the task of classifying lung cancer using artificial neural networks. Adam recorded an accuracy of 96.77%, while RMSprop only reached 95.16%. In addition to providing higher accuracy, Adam also produces better recall values as well as F1-scores, making them more effective in identifying patients suspected of having lung cancer. In the medical context, high recall is crucial because it is directly related to the model's ability to detect positive cases of cancer. This study highlights that data imbalance is a major obstacle in the classification of health data. The very limited number of non-cancerous samples made it difficult for the model to study the patterns of minority classes, so the performance in that class was still relatively low. However, the ANN model designed in this study has shown very satisfactory overall classification results and has the potential to be used as a medical decision support system to help with early detection of lung cancer based on clinical data.

4. Conclusion

According to the study's findings, artificial neural tissue (ANN) can classify lung cancer using clinical data with very satisfactory results. Experiments showed that Adam's optimizer performed better than RMSprop, with a test accuracy of 96.77%, precision of 0.98, recall of 0.98, and F1-score of 0.98. Analysis through the confusion matrix and ROC curve also confirms the model's high classification capabilities, with minimal error rates and AUC values close to 1. This study confirms that the selection of optimizer has a significant effect on the stability and performance of ANN, where Adam produces convergence and generalization capabilities that are superior to RMSprop. However, the study is still limited by dataset imbalances, especially the lack of data on non-cancerous classes that affect classification outcomes for minority classes. Therefore, it is recommended that further research use a larger and more balanced dataset, as well as explore deep learning methods and other optimization tuning techniques to optimally improve model performance.

REFERENCES

[1] S. Ghosh and A. Chatterjee, “Self-attention driven transfer learnt deep neural network for improved lung cancer detection from radiographic images,” *Heliyon*, vol. 11, no. 16, Nov. 2025, https://doi.org/10.1016/j.heliyon.2025.e44066.

[2] S. Wankhade and V. S., “A novel hybrid deep learning method for early detection of lung cancer using neural networks,” *Healthcare Analytics*, vol. 3, Nov. 2023, https://doi.org/10.1016/j.health.2023.100195.

[3] S. S. Nair, V. N. M. Devi, and S. Bhasi, "Enhanced lung cancer detection: Integrating improved random walker segmentation with artificial neural network and random forest classifier," *Heliyon*, vol. 10, no. 7, Apr. 2024, https://doi.org/10.1016/j.heliyon.2024.e29032.

[4] S. S.K.B *et al.*, "An enhanced multimodal fusion deep learning neural network for lung cancer classification," *Systems and Soft Computing*, vol. 6, Dec. 2024, https://doi.org/10.1016/j.sasc.2023.200068.

[5] "Jeremy Miller on Waterfall Versus Agile," Jul. 01, 2020, *IEEE Computer Society*. https://doi.org/10.1109/MS.2020.2987493.

[6] Y. Du *et al.*, "PP-OCR: A Practical Ultra Lightweight OCR System," Oct. 2020, [Online]. Available: http://arxiv.org/abs/2009.09941

[7] A. Redaksi, "Teknik Informatika Universitas Lampung."

[8] A. Barakat and P. Bianchi, "Convergence and Dynamical Behavior of the ADAM Algorithm for Non-Convex Stochastic Optimization," May 2020, [Online]. Available: http://arxiv.org/abs/1810.02263

[9] C. L. Chen *et al.*, "An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning," *Nat. Commun.*, vol. 12, no. 1, Dec. 2021, https://doi.org/10.1038/s41467-021-21467-y.

[10] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, Aug. 2023, https://doi.org/10.1007/s00521-023-08568-z.

[11] D. Oktafiani *et al.*, "Analisis Algoritma Naive Bayes Untuk Prediksi Kepuasan Layanan Akademik Berbasis Data Multibahasa," *Jurnal Algoritme*, vol. 6, no. 1, pp. 142–151, 2025, doi: 10.35957/algoritme.v5i3.13456.

[12] H. Thamrin, D. Oktafiani, I. I. Rasyid, and I. M. Fauzi, "Classification of SWOT Statements Employing BERT Pre-Trained Model Embedding," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 2, pp. 143–152, Apr. 2024, https://doi.org/10.21456/vol14iss2pp143-152.

[13] R. Aarthy, V. Muthupriya, and G. N. Balaji, "Detection of bone cancer based on a four-phase framework generative deep belief neural network in deep learning," *Alexandria Engineering Journal*, vol. 109, pp. 394–407, Dec. 2024, https://doi.org/10.1016/j.aej.2024.08.094.

[14] E. Dritsas and M. Trigka, "Lung Cancer Risk Prediction with Machine Learning Models," *Big Data and Cognitive Computing*, vol. 6, no. 4, Dec. 2022, https://doi.org/10.3390/bdcc6040139.

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