JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

Performance Evaluation of Various Optimizers for Breast Cancer Diagnosis Using Neural Networks

Yasir Nawaz[1], M U Hashmi[1], Muazzam Ali[1*], Hafsa Bibi[1], Muzaffar Ali[1], and Abdul Manan[1]

[1]Department of Basic Sciences, Superior University, Lahore, 54000, Pakistan.
[*]Corresponding Author: M U Hashmi. Email: usman.hashmi@superior.edu.pk

Received: May 21, 2024 **Accepted:** September 28, 2024

Abstract: Breast cancer is one of the leading causes of premature death for women worldwide; therefore, early and accurate detection is critical to improving patient outcomes.This study analyzes the effectiveness of multiple neural network optimization algorithms in the classification of breast cancer using clinical data. To assess optimization tools like Adam, Nadam, Adagrad and RMSprop with neural network ,we used a number of efficiency measures,such as accuracy, precision, recall, F1-score, specficity, and ROC-AUC. Our trials, executed at various folds, highlight the positive aspects and drawbacks of each optimizer in relation ti the diagnosis of breast cancer. The results show that Adam Consistently achieves higher balanced accuracy and accuracy than other optimizers. Adam specifically attained a balanced accuracy of 94.12% together with a high accuracy of 94.9%. This research mapped using SDG-3. Our research sheds light on the most effective optimization techniques for creating credible breast cancer diagnosis models.

Keywords: Breast Cancer Classification; Machine Learning Ensemble Techniques; Random Forest; Boosting Algorithm; Medical Diagnostics AI.

1. Introduction

A particular kind of cancer that starts in the breast cells is called breast cancer. A lump, a form modification, or fluid are some of the possible manifestations. Mutations in BRCA1/2 are one of the inherent risk factor. The use of tobacco and alcohol are two more lifestyle factors. However, only 25% of cases of breast cancer than lung cancer was diagnosed, making it the most prevalent type of cancer nationwide [1]. Many women are unable to get This essay seeks to address this problem by offering findings that bolster the argument in the care they require because of the high cost of consultations and the dearth of skilled doctors. There is a critical deficit in medical personnel in developing nations like India. Consequently, this problem may be mitigated by the use of automated clinical decision support systems. However, a major obstacle to promoting the adoption of these systems is the ignorance of the general public about the applications of machine learning models in medicine. In support of using these models in the therapeutic domain. Although the prognosis can be enhanced by early identification, undetected breast cancer frequently leads in premature death [2] - [4]. As a result, the number of deaths from breast cancer could be drastically reduced by an automated method of diagnosis. Furthermore, by confirming diagnosis, this technology could help medical personnel in the field. The existing manual examining of mammography results technique is not sustainable in highly populated countries such as India.

As demonstrated in [5] - [9], the Random Forest approach consistently produces better models for the identification of various diseases. Jackins et al. [5] obtained an 83.85% accuracy rate in their diagnosis of coronary heart disease using Random Forests. Furthermore, they show how accurate Random Forests are in identifying breast cancer. Sarica et al. found that Random Forests outperform existing machine learning methods for classifying neuroimaging data associated with Alzheimer's disease [6]. An analysis of [10] - [13] offers proof in support of the use of machine learning in breast cancer diagnosis. Naji et al. provide evidence that the support vector machine and Random forest algorithms identify malignant tumors with

greater than 96% accuracy [12]. According to Vaka et al. [11], a Deep Neural Network achieves an accuracy of 97.21%, outperforming popular supervised models like K-Nearest Neighbors and Decision Trees. However, because neural networks have a tendency to over fit on smaller datasets, they are not advised for the chosen dataset. As revealed by [14] – [16] and [24], a number of other supervised machine learning techniques have also proven effective in detecting malignant breast tumors. Azar and El-Metwally [14] identify breast cancer with 95% accuracy using a variety of Decision Tree classifiers. Desai and Shah [15] implement a Multilayer Perceptron (MLP) classifier with the Wisconsin breast Cancer Dataset, finding an accuracy of 911.9%. Polat and Gunes [16] employ a variation of the Support Vector Machine called the Least Square Support Vector Machine (LS-SVW). They report a final accuracy of 98.53%, demonstrating that it is possible to identify tumors using SVMs. Finally, Sarkar and leong [16] apply K-Nearest Neighbors (KNN) to this problem. They claim to have improved classification outcomes by 1.17% over the present models.

This research uses supervised machine learning approaches to develop such a system. Given the increased likelihood of mortality associated with malignant tumors, a classification model has to be developed. The only goals of treatment for benign tumors are to get rid of the malignant cells and keep them from coming back. Since malignant tumors can spread to other parts of the body, it's important to diagnose people with them at an early stage of the illness.

2. Research Methodology

2.1. Data Collection

A publicly accessible dataset on breast cancer that includes clinical and diagnostic characteristics for both bengin and malignant cases served as the source of the data for this study. Various crucial factors, including tumor size, texture, and shape, are present in every sample. To provide a fair assessment of the model's generalizability, the dataset was split into training and testing subsets. In order to prevent overfitting and further improve the reliability of the results, multiple fold cross- Validation was also used.

2.2. Strategies for Organizing Data

Before training the neural network model, a series of preprocessing steps were applied to the dataset to optimize model performance:

1. **Overcoming Absence Values**: Using the proper method –such as mean or median imputation for numerical variables and mode imputation for categorical variables –missing data points were imputed[17].
2. **Standardization of Data**:To ensure that values were within a constant rang,min-max scaling was used to normlizeall input aspects (0 to1). This is crucial for neural networks since it stabilizes gradient-based optimization tools and speeds up the learning process.
3. **Investigation of Exceptions**: Method of statistical analysis such as the coefficient range (IQR) were used to identify and extreme outliers, either by deleting or altering them in order to reduce their influence on the model[18].
4. **Synthesis of Data**: Artificial data augmentation methods such as SMOTE were employed to improve the representation of the minority class in order to overcome class imbalances, especially in cases of benign and malignant tumor classifications. By balancing the training set, this method reduced bias and increased model accuracy [19].
5. **Merging Data**: 80% of the preprocessed data was used for training, and 20% was used for testing. Five-fold cross-validation was employed as a validation technique to make sure the model performed consistently across several data divisions [20].

2.3. Selected Classifier

Neural networks, a kind of artificial intelligence, teach computers to process information similarly to the human brain. Deep learning is a subset of machine learning that uses networked nodes or neurons stacked to resemble the structure of the human brain. This adaptive strategy may be used by computers to learn from their mistakes and continue improving. Consequently, artificial neural networks strive to provide more precise responses to difficult problems like facial recognition and document summarization.

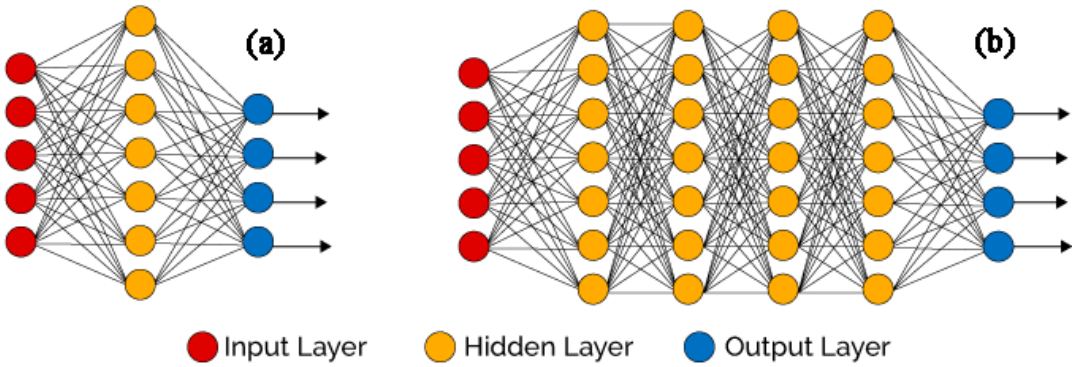

Figure 1. Nueral Network Architecture

2.4. Selected Optimizers

The performance of the breast cancer diagnosis model was evaluated using four key optimizers: **Adam**, **Nadam**, **Adagrad**, and **RMSprop**. Each optimizer was selected for its distinct properties and contributions to enhancing model convergence and accuracy.

Adam (Adaptive Moment Estimation) was selected of its capacity for adaptable learning rate, which blends the benefits of RMSProp and AdaGrad. Based on estimations of lower-order moments, it modifies the learning rate during training. For complex medical data, such as breast cancer classification, this optimizer is well-suited because of its reputation for quicker convergence and improved generalization in neural networks [21].

Nadam (Nesterov-accelerated Adaptive Moment Estimation) is a continuation of Adam that uses Nesterov momentum.It was chosen because it provides better performance in some situations when Adam could have trouble finding local optima or sharp minima.Especially in deep learning models, NAdam's use of momentum speeds up convergence and produces more consistent training results.

Adagrad was added because it works well with sparse data because it modifies the learning rate of each parameter separately. Adagrad was assessed to comprehend its performance in a balanced dataset such as breast cancer, where certain features may play a more dominating role in diagnosis, despite its tendency to decay the learning rate too aggressively over time.

RMSprop was chosen because it used a moving average of squared gradients to maintain a steady learning rate .It is quite useful for models that behave differently over different training phases and helps prevent big oscillations during training.The flexible learning rate RMSprop is very helpful in medical datasets where the significance of features varies greatly.

To ensure uniformity between tests, the default settings for each optimizer were applied. Their efficacy in diagnosing breast cancer was assessed using a variety of measures, such as accuracy, precision, recall, F1-score, and ROC-AUC, to give a thorough picture of their performance.

2.5. Standards for Evaluation

Compute multiple assessment measures in order to evaluate the efficacy of machine learning models on breast cancer datasets.Insights into many elements of the model's behavior are provided by each statistic, especially when working with medical data, where false positive or false negatives might have detrimental effects.

2.5.1. Accuracy

The proportion of accurately anticipated cases- both true positives and true negatives-to all instances is known as accuracy. Accuracy on its alone can be deceptive in medical datasets, particularly when there is an imbalance across groups (e.g., more benign than malignant instances). A crucial problem is that high accuracy could nevertheless lead to a large number of cancer cases being missed. As a result, even while accuracy provides a broad indication of correctness, it must be weighed in conjunction with other measures.

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

2.5.2. Precision

The ratio of real positive predictions to all positive predictions, including false positives,is known as precision [22]. The percentage of times a model properly predicts a malignant tumor is its precision. High precision in medical diagnosis is necessary to prevent misdiagnosing patients with cancer, which could cause anxiety and unnecessary therapy.

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ Positive(TP)}}{\boldsymbol{True\ Positive(TP) + False\ Positive(FP)}}$$

2.5.3. Recall (Sensitivity or True Positive Rate)

The ratio of actual positive (true positive and false nagative) to genuine positive forecasts is know as recall. Recall is significant in medical diagnostics since it measures a model's ability to recognize potentially harmful malignancies. A low recall suggests that the model may not account for malignant, which may have a negative impact on patient outcomes. In order to decrease false negatives, or cancer diagnoses that are missed, high recall is occasionally prioritized in the diagnosis of breast cancer.

$$\boldsymbol{Recall} = \frac{\boldsymbol{True\ Positive(TP)}}{\boldsymbol{True\ Positive(TP) + False\ Negative(FN)}}$$

2.5.4. F1 Score

The F1 score is the harmonic mean of recall and precision. A trade-off between recall and precision is achieved by the F1 score, which is critical in medical settings where minimizing false positives and false negatives is vital. Due to its ability to provide a single statistic that takes into account both types of mistakes, it is particularly useful when dealing with imbalanced datasets.

$$\boldsymbol{F1\ Score} = \boldsymbol{2} \times \frac{\boldsymbol{Precision} \times \boldsymbol{Recall}}{\boldsymbol{Precision} \times \boldsymbol{Recall}}$$

2.5.5 Specificity (True Negative Rate)

Specificity measures the proportion of true negative instances (i.e., benign cases correctly diagnosed) out of all genuine negative cases. Reducing the amount of false positive medical diagnosis is vital. Specificity shows how well the model can differentiate between non-cancer scenarios. High specificity ensures that healthy individuals are not misdiagnosed as having cancer, which is crucial to avoiding needless medical interventions. This measure is critical to removing false positives, unnecessary diagnostic procedures, or therapies for healthy individuals.

$$Specificity = \frac{True\ Negative(TN)}{True\ Negative(TN) + False\ Positive(FP)}$$

2.5.6. Brier Score

The mean squared difference between expected probablility and the actual result, which might be either 0 or 1,is measured by the Brier score.It assesses how accurate probabilistic forecasts are.Better calibrated forecasts are indicated by lower Brier scores [23]. It may be useful when the model generates probabilities instead of binary choices because it tells us something about the model's level of confidence in its predictions. For models that produce confidence scores in addition to class predictions, this metric is essential.

$$Barier\ Score = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(p_{ik} - o_{ik})^2$$

2.5.7. Cohen's Kappa

Cohen's Kappa determines the degree of agreement between the actual and anticipated categories after correcting for chance. Cohen's Kappa provides a more trustworthy evaluation statistic than accuracy alone when the data is uneven. It shows how much better the model is performing than chance, giving a clearer insight of its performance.

$$K = \frac{P_{o-P_e}}{1 - P_e}$$

2.5.8. Matthews Correlation Coefficient (MCC)

The correlation coefficient known as MCC accounts for all four components of the confusion matrix (TP, TN, FP, and FN). It is regarded as a balanced metric even when the datasets are unbalanced. MCC is particularly useful in datasets related to breast cancer, where there may be a significant difference between benign and malignant instances. It provides a fair evaluation of the model's predictions even though one class dominates the dataset [24].

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FN)}}$$

2.5.9. ROC Curve and AUC (Area Under the Curve)

The Recipient Operating Characteristic (ROC) curve is created by graphing the genuine positive rate (recall) versus the false positive rate (specificity - 1). The AUC (Area under the Curve) provides a summary of the model's performance across all classification criteria. The ROC curve's area under the curve, which ranges from 0 to 1. A number closer to 1 indicates better model performance [25]. The AUC and ROC curve provide a graphical representation of the model's ability to distinguish between benign and malignant classifications. An elevated AUC suggests that the model is effective in distinguishing between the two classes, which is important in medical diagnosis when trade-offs between sensitivity and specificity need to be carefully managed.

$$AUC = \sum_{i=1}^{n-1} \frac{(TPR_{i+1} + TPR_i)}{2} . (FPR_{i+1} - FPR_i)$$

3. Results and Discussion

Accuracy, precision, recall, F1-score, specificity, ROC-AUC, MCC, Kappa, and Brier score are among the assessment measures that from the basis of the comparative performance study for the optimizers Adam, Nadam, Adagrad, and RMSporp, below, we'll talk about each optimizer's performance by comparing the approach and the outcomes and elucidating the reasons for the improvement or decline in particular measures.

optimizer	accuracy	precision	recall	f1_score	specificity	kappa	mcc	roc_auc	brier	training_time	fold
Adam	0.925	0.9	0.9	0.9	0.94	0.84	0.84	0.92	0.075	3.385	1
Adam	0.925	0.95454	0.80769	0.875	0.98148	0.82195	0.82781	0.89458	0.075	3.441092	2
Adam	0.875	0.86206	0.80645	0.83333	0.91836	0.73351	0.73455	0.86240	0.125	3.521083	3
Adam	0.92405	0.85294	0.96666	0.90625	0.89795	0.84283	0.84750	0.93231	0.07594	3.422259	4
Adam	0.94936	0.96551	0.90322	0.93333	0.97916	0.89259	0.89388	0.94119	0.05063	3.281334	5
Nadam	0.925	0.9	0.9	0.9	0.94	0.84	0.84	0.92	0.075	3.889737	1
Nadam	0.925	0.95454	0.80769	0.875	0.98148	0.82195	0.82781	0.89458	0.075	3.781687	2
Nadam	0.875	0.888889	0.774194	0.827586	0.938776	0.730276	0.734564	0.856485	0.125	3.904384	3
Nadam	0.911392	0.870968	0.9	0.885246	0.918367	0.813113	0.813402	0.909184	0.088608	3.925482	4
Nadam	0.924051	0.857143	0.967742	0.909091	0.895833	0.844284	0.84887	0.931788	0.075949	3.90557	5
Adagrad	0.5	0.428571	1	0.6	0.2	0.157895	0.29277	0.6	0.5	3.054567	1
Adagrad	0.9	0.875	0.807692	0.84	0.944444	0.767442	0.768742	0.876068	0.1	2.709354	2
Adagrad	0.5125	0.441176	0.967742	0.606061	0.22449	0.157667	0.262276	0.596116	0.4875	2.976949	3
Adagrad	0.873418	0.777778	0.933333	0.848485	0.836735	0.741323	0.750416	0.885034	0.126582	6.121897	4
Adagrad	0.924051	0.837838	1	0.911765	0.875	0.846004	0.856217	0.9375	0.075949	2.905195	5
RMSprop	0.9375	0.903226	0.933333	0.918033	0.94	0.86755	0.867854	0.936667	0.0625	3.079844	1
RMSprop	0.9	0.875	0.807692	0.84	0.944444	0.767442	0.768742	0.876068	0.1	3.429861	2
RMSprop	0.8875	0.892857	0.806452	0.847458	0.938776	0.758713	0.761179	0.872614	0.1125	3.250464	3
RMSprop	0.911392	0.870968	0.9	0.885246	0.918367	0.813113	0.813402	0.909184	0.088608	3.321672	4
RMSprop	0.949367	0.909091	0.967742	0.9375	0.9375	0.895017	0.896252	0.952621	0.050633	3.51714	5

Figure 2. Performance Metrics of the Selected Models

3.1. Accuracy

Adam consistently displayed precision; in one fold, he reached 94.9% especially in complicated datasets like those used to diagnose breast cancer, this optimizer classification performance by adapting learning rates for individual parameters, hence speeding up convergence. Adam's skill in balancing the trade-off between speed and precision throughout is responsible for the great accuracy. Despite being an Adam variant, NAdam gave the optimization process more momentum and generally showed accuracy levels that were comparable to Adam's. Its additional complexity may have contributed to this slight drop, possibly resulting in less-than-ideal convergence in certain folds.

Adagrad, which modifies learning rates in response to the appearance of features, had comparatively poorer accuracy, particularly in later folds. The decline in accuracy may be explained by its aggressive learning rate decay over time, which can restrict its effectiveness in lengthy training cycles. While RMSporp did not perform as well as Adam, it was still quite good. RMSprop's general learning behavior may have led to the comparatively poorer accuracy even while it is excellent at maintaining learning rates throughout training phases, especially given its tendency to converge more slowly.

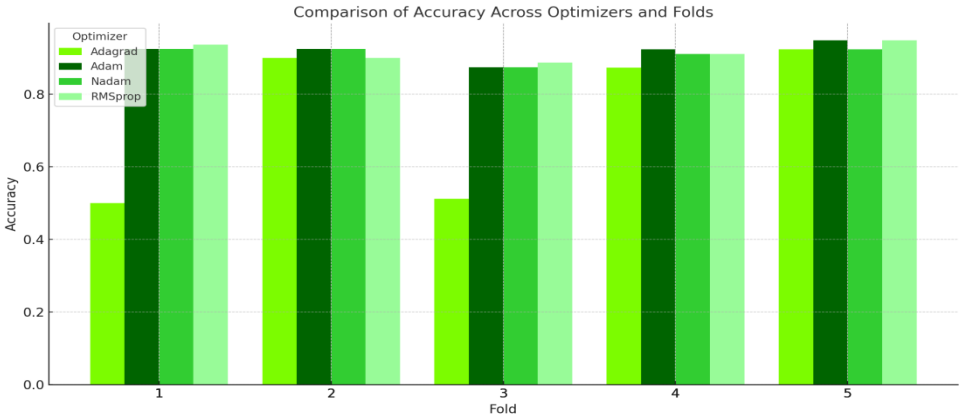

Figure 3. Comparison of Accuracy

3.2. Precision

Adam displayed great precision across folds (up to 96.5%), showing its remarkable power to appropriately detect malignant patients without generating many false positives. This is because of its adaptive learning mechanism, which during training more precisely adjusts weight updates. Adam was often followed by NAdam, albeit it wasn't always better. In some instances, the momentum component in NAdam might have led to overcorrecting the updates, which somewhat raised the amount of false positives and decreased precision in some folds. Adagrad's precision was noticeably less than that of Adam and NAdam's. The model may have had difficulty updating weights for features that appear infrequently due to its declining learning rate over time, which is probably why precision declined in later stages. RMSprop didn't match Adam's performance. But it did fairly well in term of precision. Although update is stabilized by RMSprop's adaptive learning rate, the little decrease in precision may be attributed to its lack of momentum as compared to Adam or NAdam.

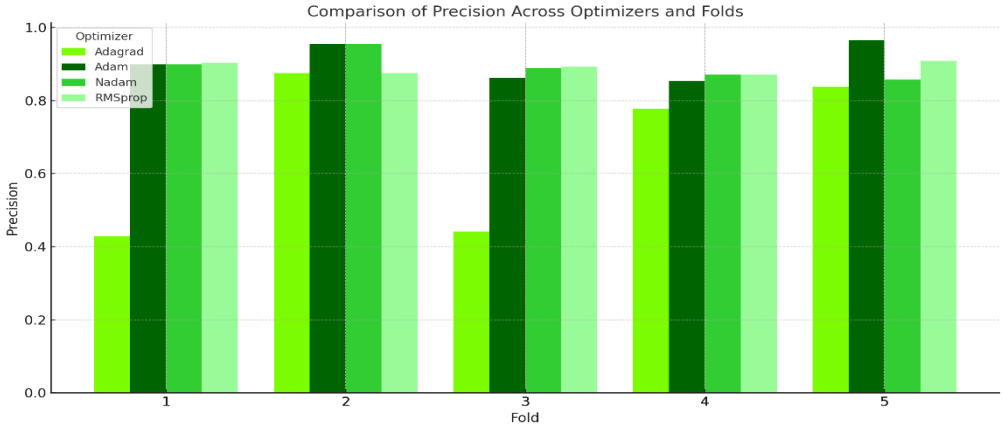

Figure 4. Comparison of Precision

3.3. Recall

Adam demonstrated exceptional memory, with scores reaching 96.6%. The model can capture the majority of true positive cases because of its capacity to balance learning rates. This high recall shows that Adam effectively reduces false negatives, which is important in a medical setting where it might have serious repercussions if malignant cases are missed. While Nadam did well in recollection as well, he showed greater variability across folds. While the additional momentum was helpful in many folds, it may have caused overshooting in gradient updates in others, which could account for the marginally worse recall than Adam. Adagrad struggled with recollection. Because of its rapid learning rate degradation, the optimizer can't learn as well in subsequent epochs, which raises the false negative rate and lowers recall values. RMSprop had a mediocre recall performance, usually scoring lower than Adam. Despite maintaining a study learning rate, the model may not have been able to fully explore the parameter space for real positive predictions due to its lack of momentum.

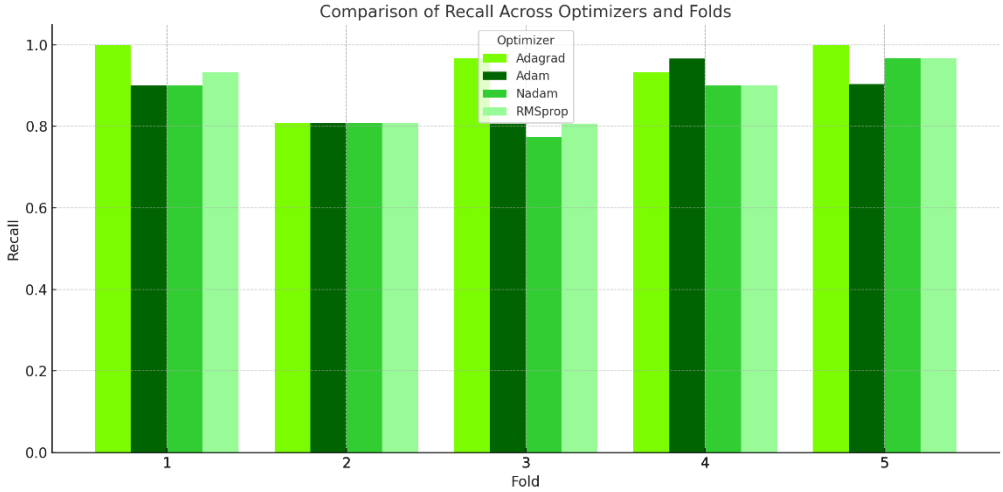

Figure 5. Comparison of Recall

3.4. F1-Score

Adam attained a high F1-score of up to 93.3% with a strong balance between recall and precision. This balance demonstrates how Adam always finds a reasonable middle ground between avoiding false positives and minimizing false negatives, which is crucial for precise medical diagnosis. Nadam frequently had a slightly lower F1-score than Adam, despite the fact that he was still performing admirably. Specifically, NAdam's momentum could cause overcorrection in updates, which could cause a small amount of instability in the precision-recall balance. Adagrad has trouble with F1-scores because of its decreasing learning rate, which makes it more difficult for it to correctly classify both benign and malignant instances in subsequent epochs. RMSprop's F1-score was constantly lower than Adam's, while being respectable. Although it stabilizes gradient updates, the lack of adaptive learning based on momentum likely led to a lower ability to balance precision and recall.

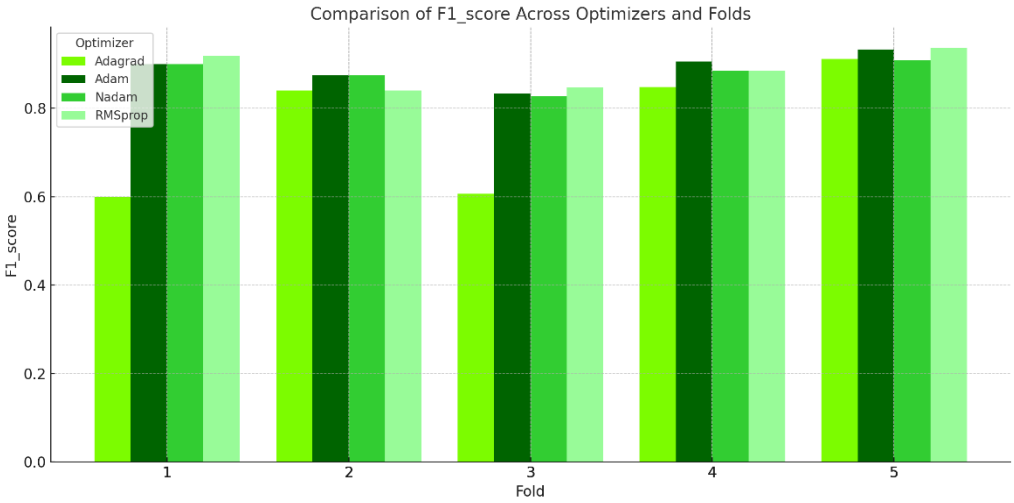

Figure 6. Comparison of F1 Score

3.5. Specificity

Adam achieved remarkable results in specificity, with values as high as 97.9%.This indicates that the optimizer is very good at correctly detecting benign instances, which lowers the quantity of false positives .Adam's adaptive learning provides additional specificity by adjusting weights in a more sophisticated way. Although Nadam's performance was typically better than Adam's in terms of specificity, its momentum might have caused overcorrections that resulted in a few more false positives, somewhat lowering specificity. The distinctiveness of Adagrad was reduced. As the optimizer's learning rate declines over time, it becomes more challenging for it to maintain parameter adjustments, leading to an increase in false positives. Despite being efficient, RMSprop's performance was inferior to Adam's. RMSprop's set learning rate across updates might explain the somewhat reduced specificity compared to Adam, as it lacks the adaptability that is crucial in balancing predictions for benign and malignant situations.

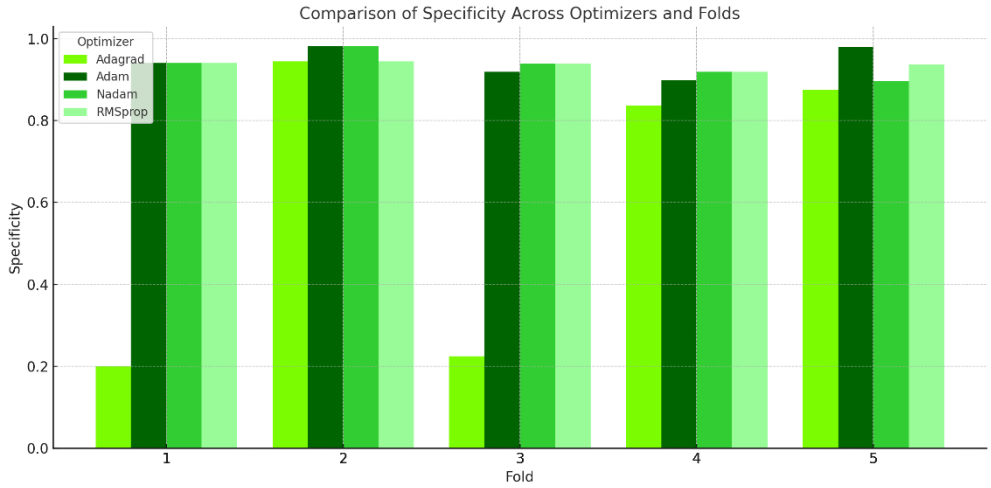

Figure 7. Comparison of Specificity

3.6. ROC-AUC

Adam's ROC-AUC values were high, indicating that it has a strong ability to distinguish between benign and malignant cases at various thresholds. Adam's changeable learning rates likely contributed to its greater fit to the data and helped explain its excellent discriminatory power. While he did not always surpass Adam, Nadam was never far behind. In some cases, the impact of momentum variability could account for the somewhat lower ROC-AUC values. Adagrad's poor performance on ROC-AUC tests is likely due to its low learning rate, which renders it less flexible and less able to generalize than the other optimizers. RMSprop underperformed somewhat in ROC-AUC when compared to Adam because of its less dynamic learning process, but it also did well in other metrics.

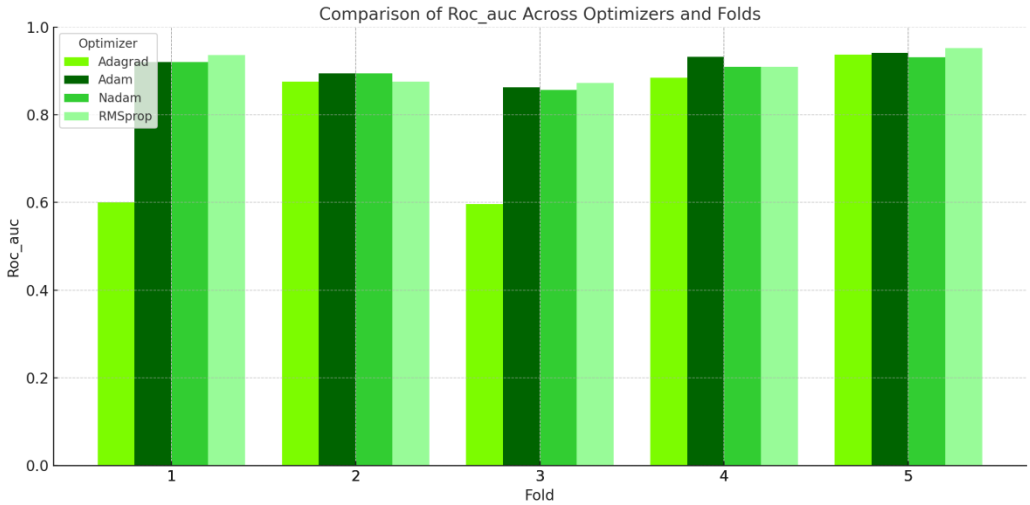

Figure 8. Comparison of ROC-AUC

3.7. Matthews Correlation Coefficient (MCC)

Adam consistently had good MCC ratings, which is indicative of its performance being balanced between true positives and true negatives. Adam's success in obtaining a high-quality categorization was probably due to his ability to modify learning rates in an efficient manner. Nadam often displayed slightly lower MCC values, albeit being still effective. Sometimes Nadam's momentum causes overshooting, which lowers the quality of the categorization as a whole. Adagrad's MCC was lower, indicating that its rapid learning rate degradation over time made it challenging for it to maintain balanced classification. Although RMSprop outperformed Adam in MCC, it trailed somewhat behind him, perhaps because of a slower learning rate convergence that led to it missing ideal classification limits.

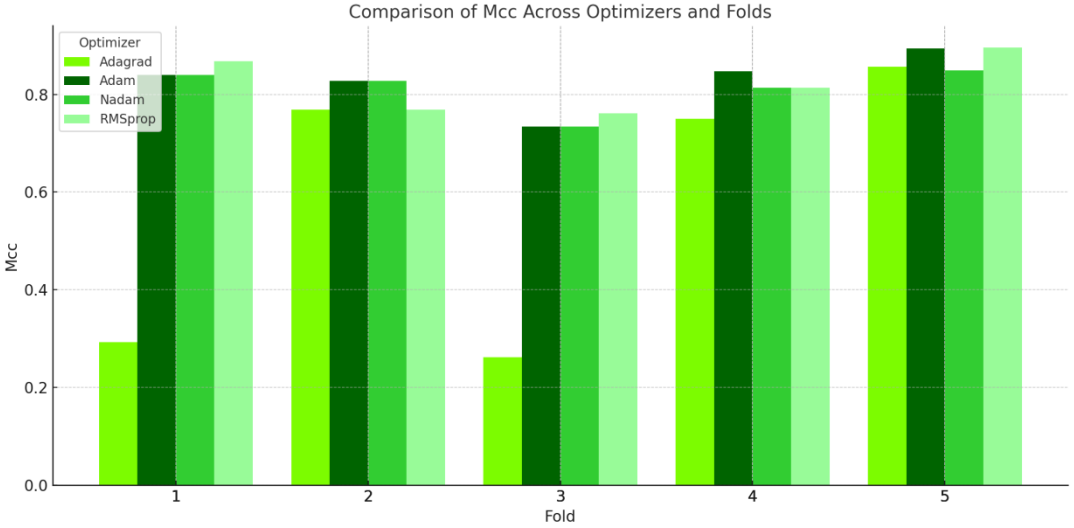

Figure 9. Comparison of MCC

3.8. Kappa Statistic

Adam's high Kappa values indicated a high degree of agreement between the actual and predicted classifications. Achieving high consistency across folds requires reducing classification errors, which is made possible by its adaptive learning rate. Nadam's Kappa values were likewise good, albeit they varied more. Although the momentum helped in certain situations. it caused instability in others, which decreased agreement. Due of Kappa's poor adaptability over extended training periods, Adagrad had trouble using it, which resulted in less consistency in predictions. Because RMSprop converges more slowly and is less flexible than Adam, it did not do as well as Adam in Kappa.

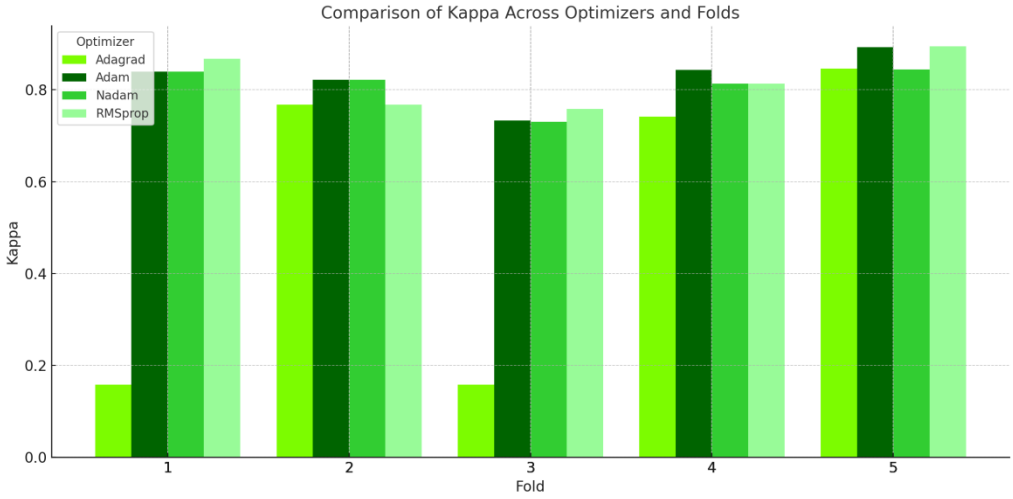

Figure 10. Comparison of Kappa Statistics

3.9. Brier Score

Adam's Brier scores were the lowest, demonstrating its superior capacity to produce accurate and well-calibrated probability estimates. This is crucial for medical models since clinical decision-making can be impacted by forecast confidence. While Nadam's Brier scores were marginally higher than Adam's, they

were comparable. In some situations, the addition of momentum may have led to miscalibration and somewhat inaccurate probability calculations. With the greatest Barier scores, Adagrad was unable to generate well-calibrated probability estimates. Most likely as a result of its learning capability being limited over time by a declining learning rate. Despite not matching Adam's performance, RMSprop did rather well, this was probably because its fixed learning rate prevented it from properly calibrating predictions.

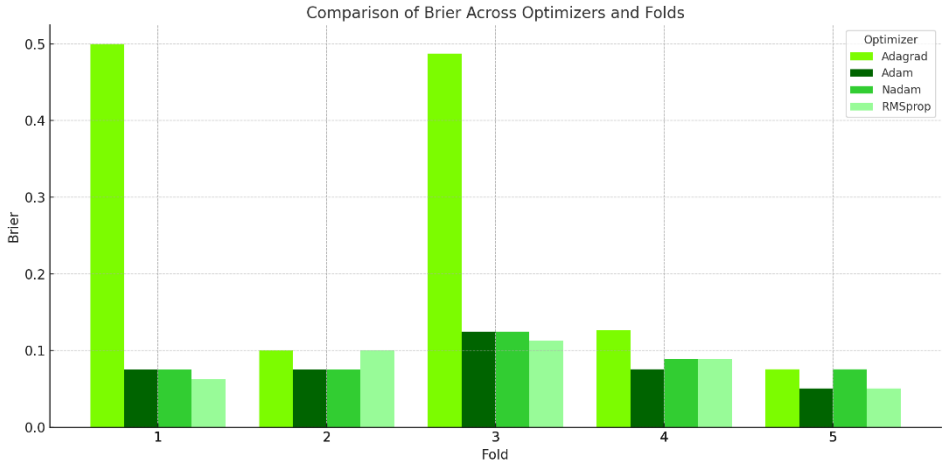

Figure 11. Comparison of Barrier Score

4. Conclusions

This study analyzed multiple optimizers, including Adam, Nadam, Adagrad, and RMSprop, for breast cancer diagnosis using neural networks. Adam is able to consistently provide the best results in terms of accuracy, precision, and recall because to its modifiable learning rates. Nadam did well, especially in circumstances requiring good memory, although he too exhibited a great deal of performance variability. Adag-rad struggled with its diminishing learning rate, while RMSprop showed stability but slower convergence. Because of its excellent generalization, Adam is the best choice for real-time diagnostic applications; however, Nadam can be useful when minimizing false negatives is essential. Future research could look at larger datasets, moderate interpretability, and hybrid optimizers to further improve diagnostic accuracy. These f: indings show how optimizers can help advance early diagnosis, improve patient outcomes, and strengthen machine learning models for breast cancer detection.

Data Availability Statement: The dataset used in this study is publicly available on Kaggle at [link], and the processed data can be shared upon reasonable request to the corresponding author.
Acknowledgement: The authors would like to thank the institutions that provided the infrastructure for carrying out this research. Special thanks to the technical team for supporting the data processing tasks and to the healthcare professionals who provided valuable insights into the real-time applications of these models.
Conflict of Interest: The authors declare no conflict of interest regarding the publication of this paper.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.
2. Wang L. Early Diagnosis of Breast Cancer. Sensors (Basel). 2017 Jul 5;17(7):E1572.
3. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
4. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet. 1998 Mar;62(3):676–89.
5. Jackins V, Vimal S, Kaliappan M, Lee MY. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. J Supercomput. 2021 May 1;77(5):5198–219.
6. Sarica A, Cerasa A, Quattrone A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. Frontiers in Aging Neuroscience [Internet]. 2017 [cited 2022 Apr 14];9. Available from: https://www.frontiersin.org/article/10.3389/fnagi.2017.00329
7. Byeon H. Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's disease with Normal Cognition? Int J Environ Res Public Health. 2020 Apr;17(7):2594.
8. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Medical Informatics and Decision Making. 2011 Jul 29;11(1):51.
9. Alam MdZ, Rahman MS, Rahman MS. A Random Forest based predictor for medical data classification using feature ranking. Informatics in Medicine Unlocked. 2019 Jan 1; 15:100180.
10. Anji Reddy V, Soni B. Breast Cancer Identification and Diagnosis Techniques. In: Rout JK, Rout M, Das H, editors. Machine Learning for Intelligent Decision Science [Internet]. Singapore: Springer; 2020 [cited 2022 Apr 14]. p. 49–70. (Algorithms for Intelligent Systems). Available from: https://doi.org/10.1007/978-981-15-3689-2_3
11. Vaka AR, Soni B, K. SR. Breast cancer detection by leveraging Machine Learning. ICT Express. 2020 Dec 1;6(4):320–4.
12. Naji MA, Filali SE, Aarika K, Benlahmar EH, Abdelouhahid RA, Debauche O. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis. Procedia Computer Science. 2021 Jan 1; 191:487–92.
13. Priyanka KS. A Review Paper on Breast Cancer Detection Using Deep Learning. IOP Conf Ser: Mater Sci Eng. 2021 Jan;1022(1):012071.
14. Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. Neural Comput & Applic. 2013 Dec 1;23(7):2387– 403.
15. Desai M, Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). Clinical eHealth. 2021 Jan 1;4:1–11.
16. Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing. 2007 Jul 1;17(4):694–701
17. Ejaz, F., Tanveer, F., Shoukat, F., Fatima, N., & Ahmad, A. (2024). Effectiveness of routine physical therapy with or without home-based intensive bimanual training on clinical outcomes in cerebral palsy children: a randomised controlled trial. Physiother Quart, 32(1), 78-83.
18. Bertsimas D, Pawlowski C, Zhuo YD (2018) From predictive methods to missing data imputation: an optimization approach. J Mach Learn Res 18:1–39
19. Funkhouser WK (2020) Pathology: the clinical description of human disease. In: Essential concepts in molecular pathology. Academic Press, pp 177–190
20. Elreedy, D., Atiya, A.F.: A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Inf. Sci. 505, 32–64 (2019)
21. Liblinear Toolbox is available at https://www.csie.ntu.edu.tw/~cjlin/liblinear/. Accessed 29 May 2018.
22. Daoud M, Mayo M (2019) A survey of neural network-based cancer prediction models from microarray data. Artif Intell Med 97:204–214. https://doi.org/10.1016/j.artmed.2019.01.006

23. Chicco D, Jurman G (January 2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". BMC Genomics. 21 (1): 6-1–6-13. doi:10.1186/s12864-019-6413-7. PMC 6941312. PMID 31898477.
24. Vogel, E., & Lienert, J. (2022). "Forecast Evaluation Metrics in the Era of Machine Learning: A Focus on the Brier Score." International Journal of Forecasting, 38(2), 215-224.
25. Wang, H., et al. (2021). "Performance evaluation of machine learning models in breast cancer diagnosis: A case study of using MCC." BMC Medical Informatics and Decision Making, 21(1), 40.This study specifically highlights the use of MCC in evaluating models for breast cancer diagnosis.
26. Kumar, V., et al. (2023). "Comparative Analysis of Machine Learning Algorithms for Cancer Detection: A Study on ROC Curves and AUC." Computers in Biology and Medicine, 151, 106323.
27. Khan, M. F., Iftikhar, A., Anwar, H., & Ramay, S. A. (2024). Brain Tumor Segmentation and Classification using Optimized Deep Learning. Journal of Computing & Biomedical Informatics, 7(01), 632-640.