IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Neural Horizons: Comparison of Advanced Deep Learning Models for the Revolution in Breast Cancer Diagnosis

Munduku Munduku Deo[1] ,Batubenga Mwamba Nzambi Jean Didier[2]

[1, 2,] Department of Mathematics, Statistics and Computer Science, Faculty of Science and Technology, University of Kinshasa, Kinshasa, DR Congo

Corresponding author: Munduku Munduku

Abstract:
In this groundbreaking study, we orchestrated a meticulous comparison of four revolutionary deep learning architectures: the Multilayer Perceptron (MLP), the Convolutional 1D Neural Network (Conv1D), the Recurrent Neural Network (RNN) and the Long Term Short Term Memory (LSTM). We have thus deployed their disruptive potential for cutting-edge breast cancer diagnosis. Drawing on the Wisconsin Breast Cancer Database (WBCD) and the Breast Cancer Coimbra Database (BCC), our research not only optimised hyperparameters via Grid Search CV but also incorporated cross-validation, paving the way for a new era in diagnostic reliability and robustness. Our exploration revealed exceptional performance on WBCD, MLP and Conv1D leading the way with spectacular accuracies of 99.30% and 96%, near-perfect F1 scores of 0.99 and 0.96, and ideal AUCs of 1.00. The RNN and LSTM models followed with distinction, displaying accuracies of 97.20% and 98.60%, F1 scores of 0.97 and 0.98, and AUCs of 1.00 and 0.99 respectively,Concerning the BCCD, the models demonstrated remarkable adaptability and performance. MLP shone with an accuracy of 80.77%, an F1 score of 0.80, and an AUC of 0.88, while Conv1D, RNN, and LSTM presented accuracies of 81%, 84.62%, and 84.62%, with F1 scores of 0.78, 0.82, and 0.83, and AUCs of 0.88, 0.89, and 0.81.This research represents a significant leap towards the optimal use of deep learning to save human lives.

IndexTerms **-** Deep learning, Multilayer Perceptron (MLP), Convolutional Neural Network 1D (Conv1D), Recurrent Neural Network (RNN), Long Term Short Term Memory (LSTM), Wisconsin Breast Cancer Database (WBCD), Breast Cancer Coimbra Database (BCCD).

1.INTRODUCTION

Breast cancer remains a major global health problem, affecting millions of people worldwide [1]. In 2022, breast cancer caused 670,000 deaths worldwide [2], according to the WHO report of 13 March 2024. Around half of all breast cancers occur in women with no specific risk factors other than gender and age [1]. In 157 out of 185 countries, breast cancer was the leading cause of cancer in women. By 2022, every country in the world will be affected by breast cancer, and men will account for around 0.5% to 1% of people with the disease[1]. The emergence of breast cancer is influenced by a combination of factors[3,4], with research indicating that genetic mutations and family histories of cancer are responsible for 5-10% of breast cancer cases[5]. In addition, it is estimated that lifestyle adjustments and other adjustable factors could potentially impact 20-30% of breast cancer diagnoses [5,6]. In the healthcare sector, the use of data analysis techniques can play a crucial role in predicting various clinical scenarios, reducing the costs associated with treatment and increasing the effectiveness of the care provided, thereby helping to save lives [12].More recently, the development of machine learning has led to a revolution in the way this disease is diagnosed [7], making it possible to detect and personalise treatment protocols [8]. Thanks to the use of advanced algorithms [31], these models can scan immense quantities of data [11], in order to spot complex patterns [9], revealing essential information [32], which enables breast cancer to be predicted reliably [10].We have made several significant contributions to the science of medical diagnosis using deep learning techniques. Here are the most relevant contributions of our work:Our novel approach to hyperparameter optimisation via Grid Search CV, accompanied by cross-validation, has set a new standard in the reliability and robustness of results. We integrated sophisticated pre-processing steps, such as class balancing via SMOTE and normalisation, as well as a dimensionality reduction method via Recursive Feature Elimination (RFE). These techniques significantly improved the performance of our models by focusing on the most influential variables. Thanks to these sophisticated techniques, we obtained superior results, with an accuracy of 99.30% on the WBCC dataset, surpassing the results of the most recent work in this field. This performance illustrates the effectiveness of our deep learning approaches in improving the reliability and accuracy of breast cancer diagnosis, and we have compared our results with those of the most recent studies to highlight the significant advance that our methods represent. The application of these different deep learning techniques was crucial in achieving superior performance, demonstrating the potential of these technologies in transforming healthcare, particularly in the crucial diagnosis of breast cancer.These contributions illustrate our commitment to advancing accuracy, efficiency and reliability in breast cancer diagnosis, offering valuable insights for future research and clinical applications. Our work clearly demonstrates the potential impact of deep learning and machine learning in transforming healthcare, particularly in the critical area of cancer diagnosis.

2. PROPOSED METHOD

Our study explored the effectiveness of four advanced deep learning architectures: Multilayer Perceptron (MLP), Convolutional 1D Neural Network (Conv1D), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) in breast cancer diagnosis, using datasets from the Wisconsin Breast Cancer Database (WBCD) and Breast Cancer Coimbra Dataset (BCCD) [14,15]. After a crucial data pre-processing step, including balancing via SMOTE and normalisation, we proceeded to a rigorous separation into training and test sets. An exhaustive optimisation of hyperparameters was carried out using Grid SearchCV, supported by cross-validation to confirm the reliability of our models. Careful evaluation was based on key performance indicators, including accuracy, F1 score, and AUC, for rigorous comparison. The full methodology, illustrated in Figure 2.1, demonstrates the precision and rigour of our analytical approach to this investigation.

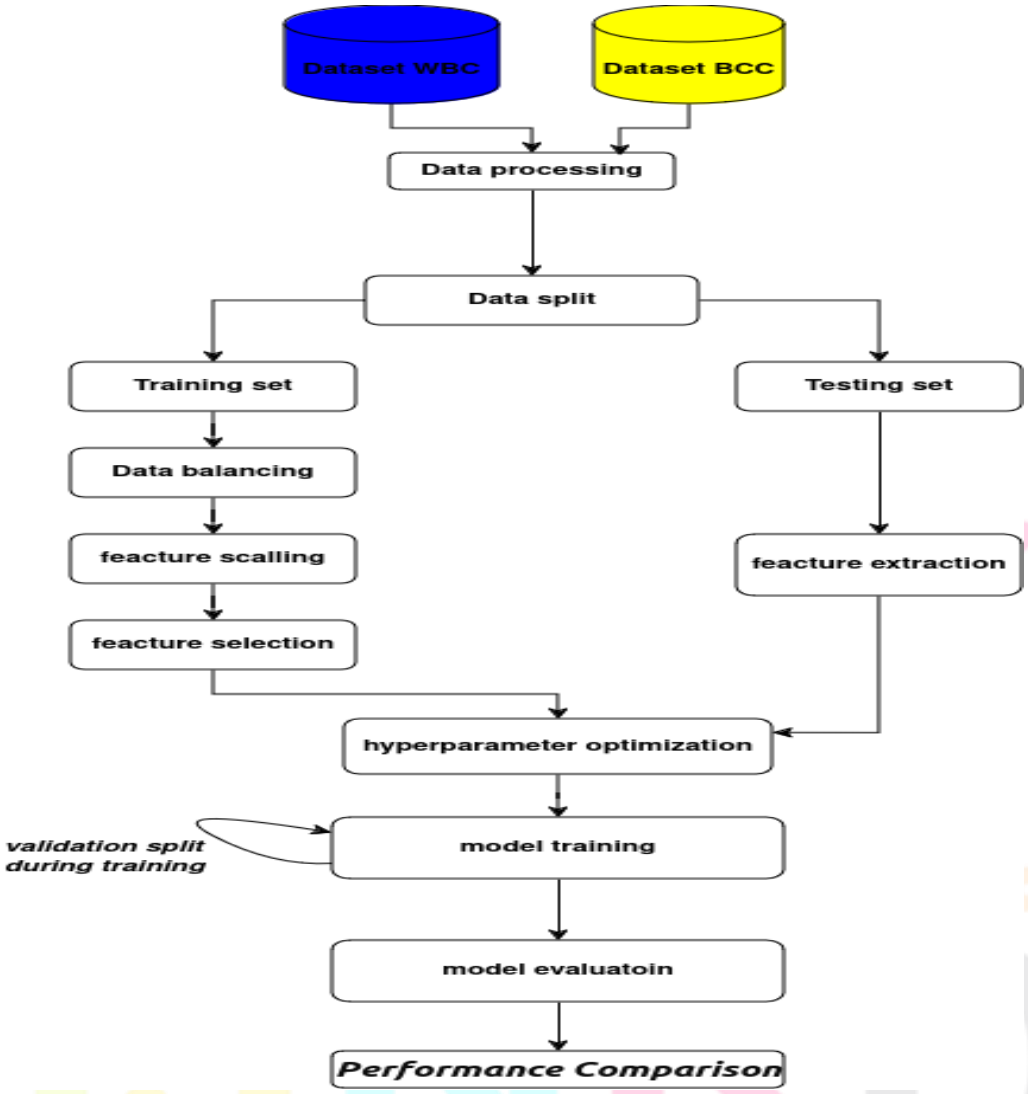

Fig.2-1 Breast cancer classification pipeline with hyperparameter optimisation

2.1 Description of the dataset

We chose to work with two specific datasets: the Wisconsin Breast Cancer Database (WBCD) and the Breast Cancer Coimbra (BCC). Our selection was based on their established reputation in the scientific community [16], being resources frequently used in many previous studies, as highlighted in various references [16, 17, 18, 19,20]. This decision stems from our commitment to scientific rigour and our desire to ensure high-quality results. We judged these datasets to be appropriate[18] and relevant[19] to our field of research. Their choice was also based on their ability to provide robust[20] and representative[17] data, thus meeting the requirements of our advanced analyses.

The WBCD dataset, compiled by the University and Hospitals of Wisconsin in 1995[21]. This dataset includes **569 observations** from **32 patients**, with a distribution of 62.74% benign versus **37.26%** malignant cases [14]. The data structure consists of **32 variables**, including a unique identifier for each case, the diagnostic label (benign or malignant), and **30 diagnostic variables** [22]. These are derived from 10 primary features, including perimeter, radius, texture, smoothness, area, compactness, concavity, concave points, fractal measure and symmetry [14]. Figure 2.2 below illustrates the distribution of diagnoses in the dataset, highlighting the proportion of benign versus malignant cases.

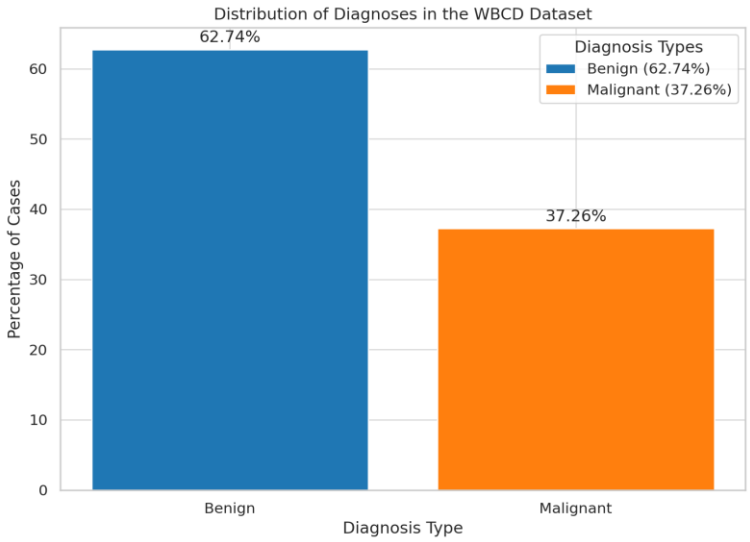

Fig. 2-2 Distribution of diagnoses in the WBCD dataset.

The Coimbra Breast Cancer Dataset (BCCD) from the University of Coimbra [15], compiles a set of biometric measures from routine blood samples [16], consisting of nine predictor variables and a categorical outcome expressing the disease state of the patients examined [15]. Predictive measures include demographic data such as age, body metrics such as BMI, and biochemical data such as glucose and insulin levels. The HOMA device and concentrations of leptin, adiponectin, resistin and MCP-1 complete this range[15]. Collected between 2009 and 2013, this information represents samples from 64 women diagnosed with breast cancer and 52 undiagnosed controls[15], giving a total sample size of 116 individuals[16]. Figure 2.3 below shows a pie chart that visually quantifies cases according to whether or not they are affected by breast cancer, illustrating the respective proportion of diagnoses within the cohort studied.

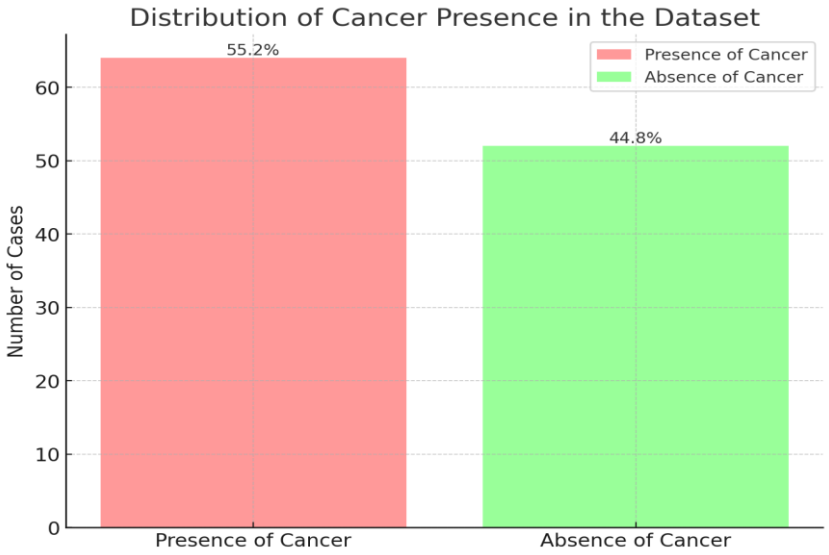

Fig. 2-3.pie chart of the distribution of diagnoses in the Coimbra dataset.

2.2 RFE Methodology for Data Dimensionality Reduction

Recursive feature elimination (RFE) is a key mechanism in the field of machine learning [16], especially when faced with datasets with a large number of variables [16]. The aim of this process is to discern and retain only the most significant variables that contribute to maximising the accuracy of predictions [23]. By reducing the number of features, only those that are truly relevant to the model are retained [23], increasing its efficiency and improving the relevance of predictions, while simplifying the model for greater interpretability and better overall performance [24]. In this work, as part of the refinement of predictive models, the RFE technique is used to distinguish essential features[51], within a large dataset. This approach[52], based on the careful evaluation and rigorous selection of attributes[53], focuses on retaining only those that have a decisive influence on the predictions[54]. It enables the model to be simplified effectively, gradually eliminating superfluous elements and retaining only the essentials[54], making it easier to improve model performance by reducing dimensionality[53], eliminating less relevant variables and concentrating on those that contribute most to accurate prediction[54] .

2.3 Optimising hyperparameters

Parameter tuning in machine learning models aims to identify the ideal configurations that maximise predictive efficiency for various applications[47], thereby helping to improve the overall accuracy and reliability of the analyses carried out[25].The parameters that govern the way in which learning takes place in the models are crucial to their performance[25]. Various strategies for adjusting these parameters exist, encompassing systematic methods such as exhaustive or random search [50], sophisticated approaches such as Bayesian optimisation, and advanced techniques based on the gradient or on evolutionary and collective principles [16]. In our study, we employed the GridSearchCV optimisation technique for its proven ability to efficiently determine

optimal parameters [16]. This technique adopts a brute-force approach [49], testing all possible combinations of specified parameters within a predefined grid to identify the one that maximises cross-validation performance [16]. We chose to use K-folds cross-validation, a standard scikit-learn method, adapted to our datasets for disease prediction. In order to guarantee reliable and accurate predictive models, we integrated GridSearchCV into our evaluation process, enabling us to test a multitude of hyperparameter combinations. The application of GridSearchCV was systematic, using 'estimator', 'param_grid', 'scoring', 'verbose' and 'n_jobs' to refine and evaluate the performance of each hyperparameter combination through our deep learning models.

2.4. Model Performance Evaluation

In order to ensure a rigorous and comprehensive evaluation of the performance of the proposed deep learning architectures, we have adopted a multidimensional approach based on a set of statistically robust metrics [44]. This section describes the criteria used to quantify classification efficiency, highlighting the importance of each metric in the context of our study

Validation Accuracy in GridSearchCV:This metric is essential for identifying the set of hyperparameters that maximises model generalisation on unseen data,it is the metric that measures the proportion of correct predictions made by a model on a validation set during the GridSearchCV process [55]. It is calculated for each hyperparameter configuration tested during the cross-validation integrated into GridSearchCV, and guides the selection of the best model by avoiding overlearning and ensuring optimal performance on data outside the training set [56].

accuracy-validation=$\frac{TPval+TNval}{TPval+TNval+FPval+FNval}$ (2.1)

where **TPval** is the number of positive examples correctly identified by the model on the validation set. **TNval** is the number of negative examples correctly identified by the model on the validation set.**FPval** is the number of negative examples incorrectly identified as positive on the validation set.FNval is the number of positive examples incorrectly identified as negative on the validation set.

Accuracy (Model Accuracy): is an essential metric for assessing the overall performance of the model on the test set, providing an overview of its ability to correctly identify the two classes [34].

Accuracy=$\frac{TP+TN}{TP+FP+TN+FN}$ (2.2)

where TP: Number of positive examples correctly identified by the model.TN: Number of negative examples correctly identified by the model.FP: Number of negative examples incorrectly identified as positive.FN:Number of positive examples incorrectly identified as negative.

Sensitivity or True Positive Rate (Recall): This measure assesses the model's ability to correctly identify positive observations [35,71].

Sensitivity = $\frac{TP}{TP+FN}$ (2.3)

Accuracy: This quantifies the proportion of positive predictions that are actually correct [35].

Precision=$\frac{TP}{TP+FP}$ (2.4)

F1 Score= 2*$\frac{Precision+Recall}{precision*Recall}$ (2.5)

AUC (Area Under the Curve):is particularly useful because it is independent of the decision threshold and gives a measure of the model's performance over all possible thresholds, providing a robust assessment of its ability to classify observations correctly[35].$\int_0^1 \blacksquare TPR(FPR)d(FPR)$ (2.6)

where *TPR* (True Positive Rate) is calculated as $\frac{TP}{TP+FN}$ (2.7)

FPR (False Positive Rate) is the false positive rate, calculated as $\frac{FP}{FP+TN}$ (2.8)

We also examined the confusion matrix and other curves for a complete assessment of the model's performance.

3. Experimental results and discussion

The system developed is fully implemented on the Google Collaboratory (Colab) platform, an invaluable resource for researchers and developers enabling the execution of computationally intensive code through access to hosted graphics processing units (GPUs) [39] . This platform facilitates the manipulation and analysis of large datasets without hardware constraints [40], which is essential for the efficient processing and analysis of complex cancer diagnostic data. The core of the development is based on the use of TensorFlow [41], integrated with the Keras library, for the construction and optimisation of several neural network architectures [41], including multilayer perceptrons (MLPs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) for processing data sequences. Keras, with its simplicity of use and flexibility, allows rapid and efficient experimentation with different model architectures and parameters [42], thus encouraging innovation and the exploration of new modelling ideas [42]. The experiments were conducted using a balanced dataset generated using the SMOTE approach. The results of each scenario are discussed separately.

3.1 Multilayer Perceptron (MLP)

In this section, we compare the results obtained with our Multilayer Perceptron (MLP) model on two distinct datasets: the Wisconsin Breast Cancer Database (WBCD) and the Breast Cancer Coimbra (BCC).Table 3.1, which follows, details the architecture of the MLP model with the best hyperparameters found and the performance of the validation accuracy that is measured during the hyperparameter selection process with GridSearchCV.

Table 3.1. MLP model architecture with the best hyperparameters found

Hidden Layer Neurons	32 neurons for both layers (optimised using GridSearchCV)
Activation function	tanh for all hidden layers (confirmed by GridSearchCV)
Dropout layer	30% rate after each hidden layer (optimised using GridSearchCV)
Optimizer	adam (confirmed by GridSearchCV)
Batch size	32 (optimised via GridSearchCV)
Number of Epochs	10 (optimised via GridSearchCV)
Max Validation Accuracy	97.20%

Table 3.2 below examines in depth the effect of the best hyperparameters on the performance of the MLP model. These indicators provide us with a comprehensive assessment of the model's quality, generalisability and reliability in classifying cancer diagnoses.

Table3.2. Results of the MLP model on the two datasets

dataset	accuracy of the model	precision	sensitivity	f1 score	AUC
WBC	99.30	1.00	0.99	0.99	1.00
BCC	80.77	0.83	0.77	0.80	0.88

Analysis of the results indicates that hyperparameter optimisation played a crucial role in improving the model's performance metrics, suggesting that the model was well adapted to the specificities of the data.Having examined the quantitative performance metrics of our Multilayer Perceptron model, we now turn to a more detailed visual analysis.The following Figures 3.1 and 3.2 show the confusion matrices for the predictions of our model on the WBCD (Figure 3.1) and BCC (Figure 3.2) datasets with the equilibrium data.These matrices allow us to visualise the distribution of correct and incorrect predictions, providing a more nuanced perspective on the performance of the model in terms of true positives, true negatives, false positives and false negatives. Examination of these matrices is essential for understanding the model's behaviour when faced with real cases, in particular its tendency to commit type I errors (false positives) or type II errors (false negatives).

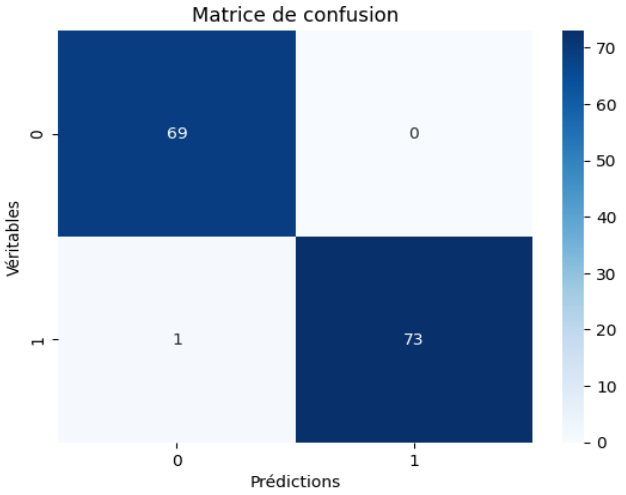

Fig 3.1.Confusion matrix on WBCD obtained by the MLP model.

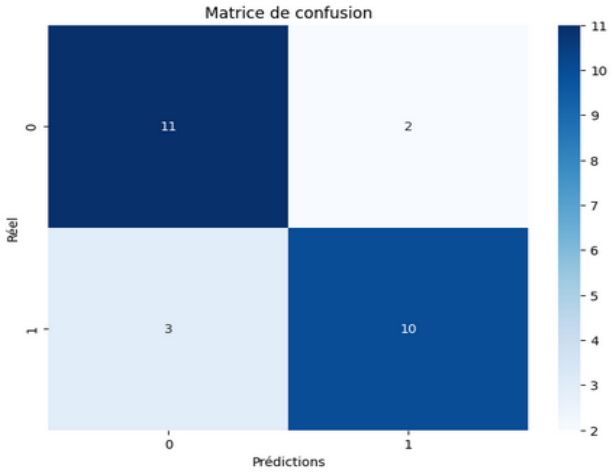

Fig 3.2.Confusion matrix on BCC obtained by the MLP model

In addition to the confusion matrices, the following Figures 3.3 and 3.4 illustrate two crucial aspects of the performance of our model on the WBC dataset for Figure 3.3 ,On the left, the accuracy curve illustrates the accuracy of training and validation across different epochs,on the right, we observe the loss curve, which plots the loss of training and validation across epochs. and in Figure 3.4, the dotted line represents random performance, while the blue curve represents the performance of the evaluated model, with an AUC (Area Under the Curve) of 0.88, indicating a good discrimination ability of the model.

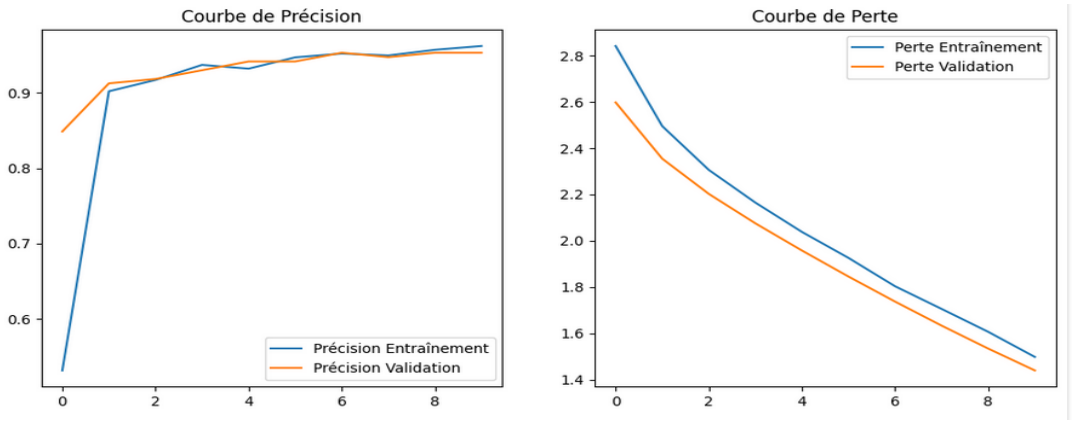

Fig 3.3.Accuracy and loss curve on WBCD obtained by the optimal MLP model.

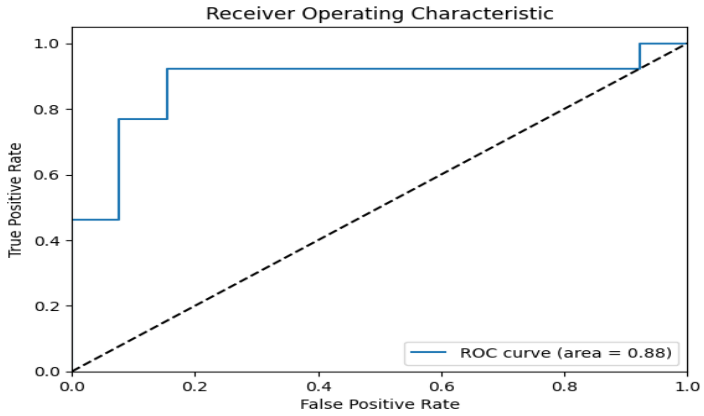

Fig 3.4.Roc curve on BCC obtained by the optimal MLP model.

3.2.1D Convolutional Neural Network (Conv1D)

Table 3.3 below provides an in-depth overview of the selected optimal architectures.

Table 3.3 Architecture of conv1D on the two datasets

Filters by Conv1D Layer	32 (optimised via GridSearchCV)

Kernel size	3 (optimised via GridSearchCV)
Activation function	relu (confirmed by GridSearchCV)
Optimizer	adam (confirmed by GridSearchCV)
Batch size	32 (optimised via GridSearchCV)
Max Validation Accuracy	98.43% (according to GridSearchCV)

Table 3.4 below provides an in-depth look at the performance of the Conv1D model, configured with the most efficient hyperparameters identified during our optimisation process.

Table 3.4. Result of the 1D conv model on the two datasets

dataset	accuracy of the model	precision	sensitivity	f1 score	AUC
WBC	96	0.95	0.97	0.96	1.00
BCC	81	0.90	0.69	0.78	0.88

We turn to a visual assessment of performance with Figures 3.5 and 3.6 below, which display the confusion matrices generated by our model on the WBCD and BCC datasets (Figure 3.5 and Figure 3.6 respectively), clearly distinguishing correct predictions from classification errors.

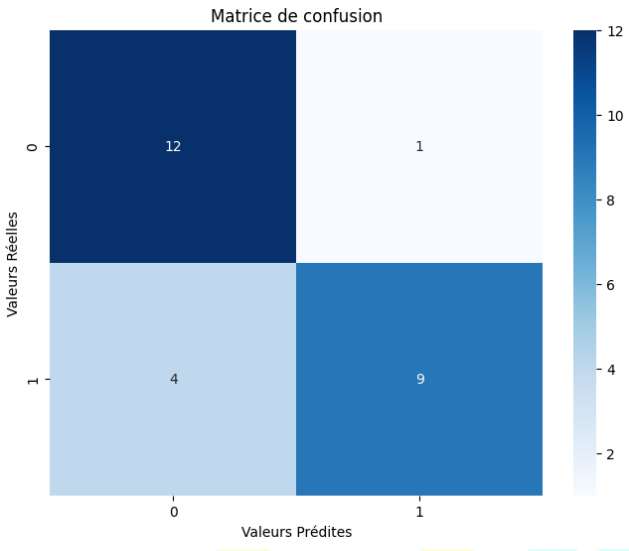

Fig 3.5.Confusion matrix on WBCD obtained by the Conv1 model

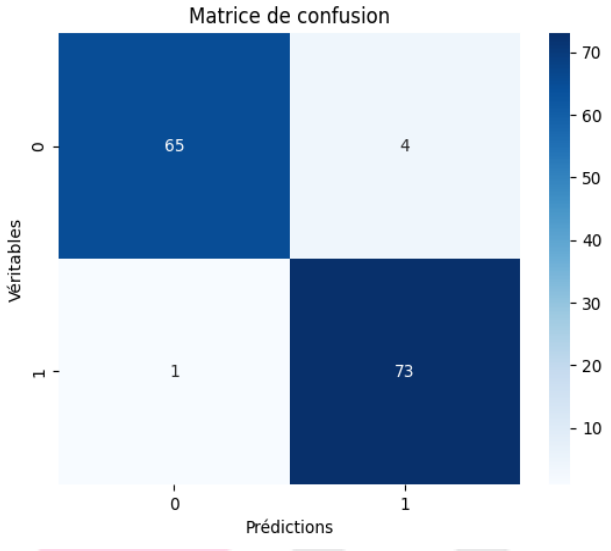

Fig 3.6.Confusion matrix on BCC obtained by the Conv1 model.

3.3. Recurrent Neural Network (RNN)

Table 3.5 below provides an in-depth look at the optimal architectures selecting the best performing hyperparameters through our optimisation process.

Table 3.5. Result of the 1D conv model on the two datasets

Hidden Layer Neurons	16 for the first layer, 8 for the second (determined by GridSearchCV)
Dropout layer	50% rate after dense layers (determined by GridSearchCV)
Optimizer	adam (confirmed by GridSearchCV)
Batch size	Varied (determined by GridSearchCV)
Number of Epochs	300(determined by GridSearchCV)

In Table 3.6 below, we deploy a series of key metrics such as accuracy, precision, recall, F1 score and AUC to assess the effectiveness of the RNN model in our classification context. These metrics provide a critical overview of the effectiveness of the RNN model and its accuracy in distinguishing different diagnostic classes.

Table 3.6. Optimal result of the RNN model on the two datasets

dataset	accuracy of the model	precision	sensitivity	f1 score	AUC
WBC	97.20	0.96	0.99	0.97	1.00
BCC	84.62	1.00	0.69	0.82	0.89

In assessing the performance of our model, particular attention was paid to the visual interpretation of the Confusion Matrix, as illustrated in Figures 3.7 and 3.8.

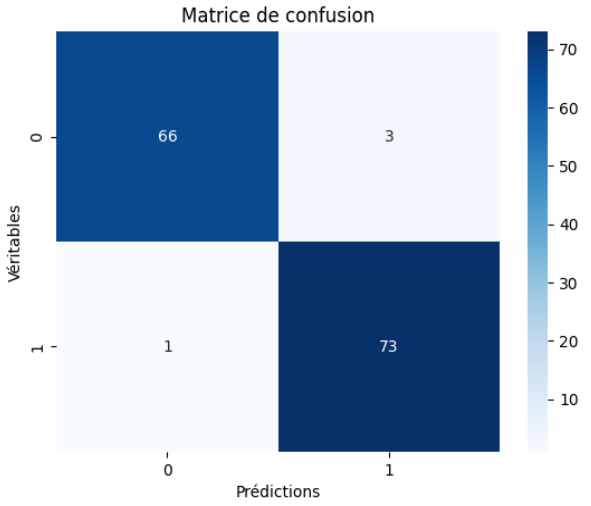

Fig 3.7.Confusion matrix on WBCD obtained by the RNN model

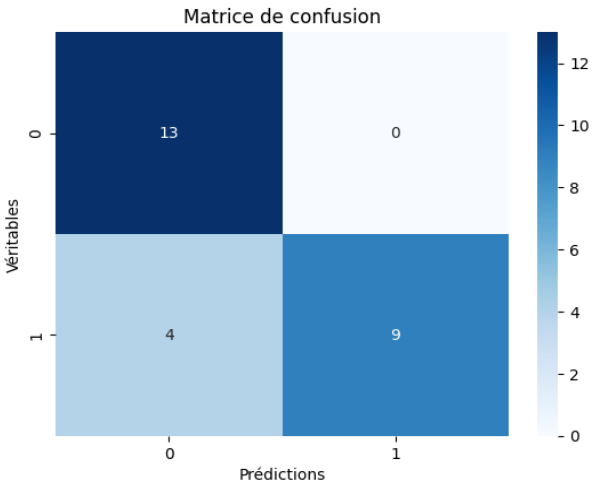

Fig 3.8.Confusion matrix on BCCD obtained by the RNN model

D. Long Short Term Memory (LSTM) neural network

Table 3.7, below, provides a detailed analysis of the optimal architectures determined by selecting the most efficient hyperparameters, as a result of our rigorous optimisation process.

Tab3. 3.7. LSTM architecture on the two datasets

Neurons per First Layer	50 neurons, determined as optimal by GridSearchCV
Neurons per Second Layer	25 neurons, via GridSearchCV
Activation function	relu' for the LSTM layers, 'sigmoid' for the output; best confirmed via GridSearchCV
Number of Epochs	150, the optimal number of GridSearchCV epochs
Max Validation Accuracy	95.45%, achieved during the best iteration of GridSearchCV

In Table 3.8 below, we present a range of key metrics including accuracy, precision, sensitivity (recall), F1 score, and AUC. These metrics are essential for assessing the performance of the LSTM model in our classification task. They provide a detailed perspective on the model's ability to correctly differentiate between the various diagnostic categories.

Tab 3.8. Optimal result of the LSTM model on the two datasets

dataset	accuracy of the model	precision	sensitivity	f1 score	AUC
WBC	98.60	0.98	0.98	0.98	0.99
BCC	84.62	0.90	0.76	0.83	0.81

Figure 3.9 and Figure 3.10 below show the optimal confusion matrices obtained from the WBCD and BCC datasets, allowing a direct comparison of the model's diagnostic performance. This representation highlights the model's accuracy in terms of true classifications versus false predictions, illustrating its reliability in the clinical context for both types of data.

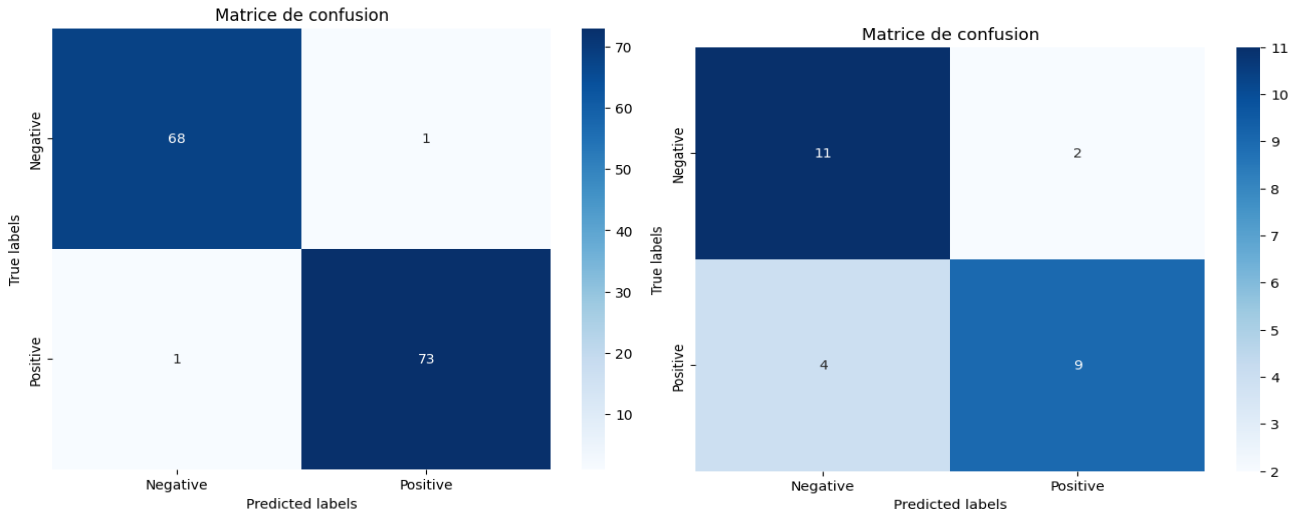

Fig 3.9.Confusion matrix on WBCD obtained by the LSTM model

Fig 3.10.Confusion matrix on BCCD obtained by the LSTM model

E. Intermediate conclusion on the results observed

Our comparative analysis revealed that the MLP and Conv1D models outperformed the RNN and LSTM architectures in terms of accuracy on the WBCD database, with scores of 99.30% and 96% respectively. These models also demonstrated an excellent ability to generalise on the BCCD, albeit slightly less than that observed on the WBCD. The RNN and LSTM models, while showing more modest performance, nevertheless demonstrated great robustness, suggesting their usefulness for applications requiring the management of sequences or complex temporal data. The following histogram in Figure 3.11 illustrates the accuracy rates obtained by the MLP, Conv1D, RNN and LSTM models on the WBCD and BCCD datasets. The coloured bars allow an instant comparison between the results on the two datasets, highlighting the strengths and weaknesses of each model in different diagnostic contexts.

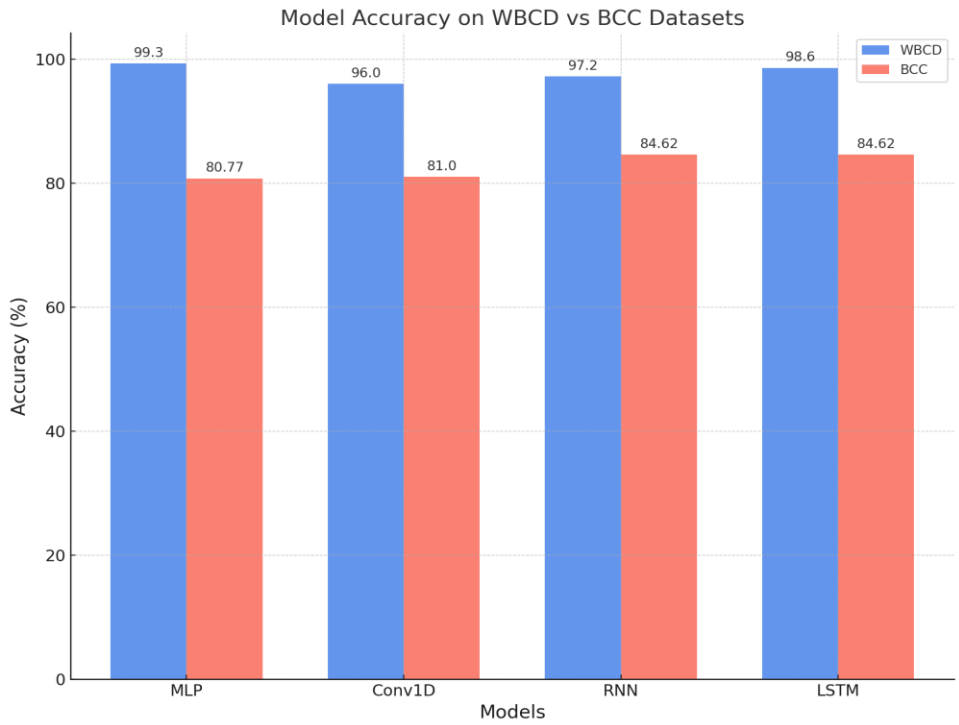

Figure 3.11: Comparison of Model Accuracy on WBCD and BCCD Datasets

Putting the F1 score into perspective, a key indicator harmonising precision and recall, is essential for a balanced assessment of classification performance, especially when faced with unequal classes. Figure 3.12 illustrates these scores for the MLP, Conv1D, RNN and LSTM models, applied to the WBCD and BCCD datasets.

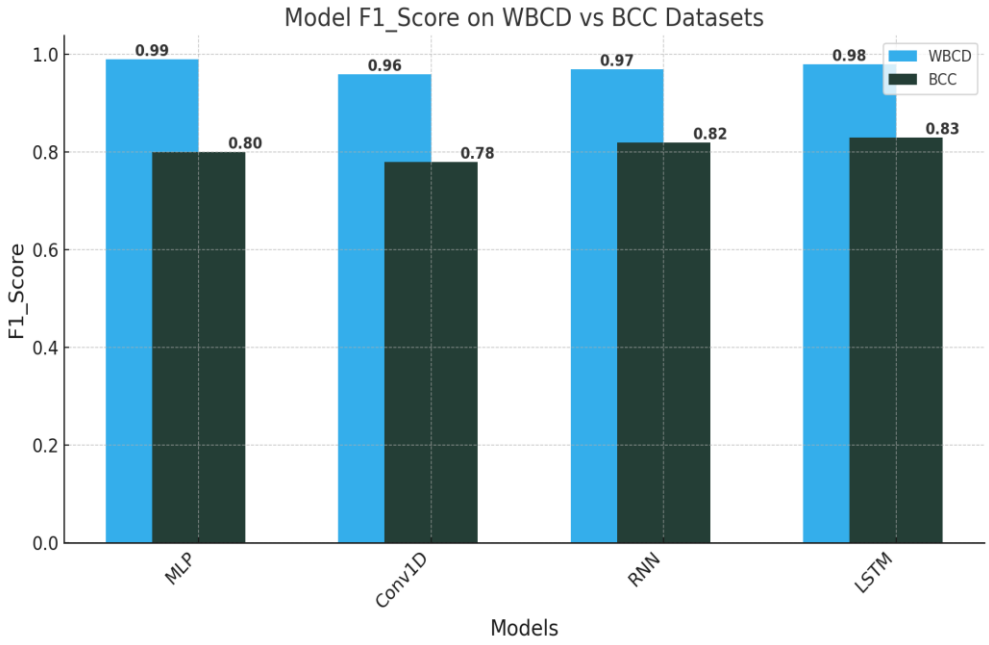

Figure 3.12: F1 Scores of Machine Learning Models on WBCD and BCCD Datasets

F. Discussion

This section presents a synthesis of recent research exploiting the Wisconsin Breast Cancer Database (WBCD) and Breast Cancer Coimbra Dataset (BCCD) for the development of predictive models of breast cancer. In this section, we focus on the analysis of the results obtained by different recent studies on the prediction of breast cancer. To ensure a relevant[43] and up-to-date comparison, we have selected only research published from 2023 onwards. This temporal limitation allows us to discuss the most recent advances and compare them directly with our own results.

Table 3.9 below summarises the details reported by each study, as well as the datasets used, which helps us to assess the relative effectiveness of the different approaches in similar settings. The aim of our analysis is to identify the advances that have been made in the field of early detection of breast cancer and to determine the extent to which these advances could influence current diagnostic practices.

Table 3.9. Comparison of Results of Recent Breast Cancer Prediction Models

Main author	Year	Precision	Data Sets Used
Amit Kumar Jakhar et al [57]	2023	99%	WBCD
Md. Murad Hossin et al[58].	2023	99.12%	WBCD
Sara Laghmati et al [59]	2023	97.37%	WBCD
Aarti et al[60]	2024	96.5%	WBCD
Achini Nisansala et al[61]	2023	97.82%	WBCD
Sujeet Kumar et al [62]	2023	98.8%	WBCD
Tsehay Admassu Assegie et al[63].	2021	92.53%	WBCD
Ashok Kumar et al[64]	2023	99%	WBCD
Ayman Alsabry et al[65].	2023	88%	BCCD
Vikas Kushwaha et al[66].	2023	99%	BCCD, WBCD
Fatema Tabassum Liza et al[67].	2023	99.20%	WBCD
Akhil Kumar Das et al[68].	2023	97.53%	WBCD
Our result	**-**	**99.30%**	**WBCC**

CONCLUSION

Our study marks a significant step towards fully exploiting the capabilities of deep learning in the accurate diagnosis of breast cancer. By employing a rigorous methodology to compare several advanced deep learning architectures and optimising hyperparameters, we have demonstrated the exceptional effectiveness of these models using the Wisconsin Breast Cancer Database (WBC) and the Breast Cancer Coimbra Database (BCC). These results underline not only the power of deep learning as a diagnostic tool but also its potential to transform breast cancer treatment protocols.the adaptability and accuracy of our models on various databases highlight their applicability in a wide range of medical contexts, laying the foundations for the future integration of these technologies into real-time diagnostic systems. Furthermore, the hyperparameter optimisation approach presented here promises to improve the performance of deep learning models in other areas of medicine.In the future perspectives of our research, we plan to focus on the explainability of predictive models. The aim will be to reveal how these models arrive at their decisions, by making their learning process transparent and comprehensible to clinicians. This move towards greater explicability aims to facilitate the clinical adoption of deep learning models by offering healthcare professionals valuable insights into how they work and why they make predictions. This will not only help to increase confidence in the use of these advanced technologies, but also enable more informed and personalised medical decision-making.

REFERENCES

[1] 0beagu, E. I., & Obeagu, G. U. (2024). Breast cancer: a review of risk factors and diagnosis. *Medicine*, 103(3), e36905. https://doi.org/10.1097/MD.0000000000036905.

[2] World Health Organisation. (2024, 13 March). Breast cancer. Retrieved 31 March 2024, from https://www.who.int/fr/news-room/fact-sheets/detail/breast-cancer

[3] Obeagu EI, Babar Q, Vincent CC et al. Therapeutic targets in breast cancer signalling: a review. J Pharm Res Int. 2021; 33: 82-99.

[4] Ibekwe AM, Obeagu EI, Ibekwe CE, et al. Challenges of exclusive breastfeeding among working-class women in a teaching hospital in southeastern Nigeria. J Pharm Res Int. 2022;34:1.

[5] Sun YS, Zhao Z, Yang ZN, et al. Risk factors and preventions for breast cancer. Int J Biol Sci. 2017; 13: 1387-97.

[6] Obeagu, Emmanuel Ifeanyi, PhD, and Obeagu, Getrude Uzoma, BN Sc. "Breast cancer: a review of risk factors and diagnosis." *Medicine* 103, no. 3 (January 19, 2024): e36905. DOI: 10.1097/MD.0000000000036905.

[7] Yuanzhou Wei, Dan Zhang, Meiyan Gao, Yuanhao Tian, Ya He, Bolin Huang, Changyang Zheng. "Breast Cancer Prediction Based on Machine Learning." Journal of Software Engineering and Applications, vol. 16, 2023, pp. 348-360.

[8] Li, M., Ma, Y., Jing, Q. and Zhu, X. (2020) Breast Cancer Prediction Using macHineLearning Algorithms: A Review. Current Medical Imaging, 16, 249-257.

[9] Saini, A. and Hukam, G. (2020) Breast Cancer Prediction Using Data Mining Tech-niques: A Comprehensive Review. International Journal of Information Technolo-gy, 12, 183-197.

[10] Tiwari, M., Bharuka, R., Shah, P., & Lokare, R. (Year). Breast Cancer Prediction using Deep Learning and Machine Learning Techniques. *K.J. Somaiya Institute of Engineering & Information Technology, Department of Information Technology, University of Mumbai, India*.

[11] Abunasser, B. S., AL-Hiealy, M. R. J., Zaqout, I. S., & Abu-Naser, S. S. (2023). Convolutional neural network for breast cancer detection and classification using Deep Learning. *Asian Pacific Journal of Cancer Prevention, 24*(2), 531-544. https://doi.org/10.31557/APJCP.2023.24.2.531.

[12] Yi-Sheng Sun, Zhao Zhao, Han-Ping-Zhu, "Risk factors andPreventions of Breast Cancer" International Journal of BiologicalSciences.

[13] Yellamma, P., Chowdary, C. S., Karunakar, G., Rao, B. S., & Ganesan, V. (2020). Breast Cancer Diagnosis Using MLP BackPropagation. International Journal, 8(9).

[14] https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnosis) (consulted on 25 march 2024).

[15] *Breast Cancer Coimbra*. University of California, Irvine, Machine Learning Repository. Retrieved from https://archive.ics.uci.edu/dataset/451/breast+cancer+coimbra (accessed March 25, 2024).

[16] Rasoul, A., Bunternngchit, C., Tiejian, L., Islam, M. R., Qu, Q., & Jiang, Q. (2022). Improved machine learning-based predictive models for breast cancer diagnosis. *International Journal of Environmental Research and Public Health, 19*(6), 3211. https://doi.org/10.3390/ijerph19063211.

[17] Abdar, M.; Yen, New York; Hung, JCS Improving liver disease diagnosis using a multilayer perceptron neural network and improved decision trees. *J.Med.Biol. Ing.* **2018** , *38* , 953-965.

[18] Hazra, A.; Mandal, Saskatchewan; Gupta, A. Study and analysis of breast cancer cell detection using Naive Bayes, SVM and ensemble algorithms. *Int. J. Comput. Appl.* **2016** , *145* , 39-45.

[19] Wang, H.; Zheng, B.; Yoon, SW; Ko, HS An ensemble algorithm based on support vector machine for breast cancer diagnosis. *EUR. J. Oper. Res.* **2018** , *267* , 687-699.

[20] Durgalakshmi, B.; Vijayakumar, V. Feature selection and classification using a support vector machine and decision tree. *Compute. Intell.* **2020** , *36* , 1480-1492.

[21] A.F.M. Agarap, "On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset," in Proceedings of the 2nd International Conference on Machine Learning and Soft Computing, February 2018, pp. 5-9.

[22] Rajasekaran, G., & Shanmugapriya, P. (2023). Hybrid Deep Learning and Optimization Algorithm for Breast Cancer Prediction Using Data Mining. International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 11(1s), 14-22.

[23] Al Reshan, M. S., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., Azar, A. T., & Cheikh, A. (Year). Improving breast cancer detection and classification using advanced multi-model features and ensemble machine learning techniques. *Journal Name, Volume*(Issue), pages. https://doi.org/DOI

[24] Kumbhare, SB; Kathole, A.; Shinde, S. Federated learning facilitated breast cancer detection using a heuristic-based intelligent deep learning framework. *Biomede. Signal Process. Control* **2023** , *86* , 105080.

[25] Aguerchi, K., Jabrane, Y., Habba, M., & El Hassani, A. H. (Year of publication). CNN Hyperparameter Optimization Using Particle Swarm Optimization for Breast Cancer Classification through Mammography. Name of journal/journal.

[26] Ahmad, S.; Ullah, T.; Ahmad, I; Al-Sharabi, A.; Ullah, K.; Khan, RA; Rachid, S.; Allah, I; Uddin, N.; Ali, S. A novel hybrid deep learning model for metastatic cancer detection. *Compute. Intell. Neurosci.* **2022** , *2022* , 1-14.

[27] Afarbigloo, Saskatchewan; Danyali, H. Classification of nuclear atypia in histopathological images of breast cancer based on CNN feature extraction and LSTM classification. *CAAI Trans. Intell. Technology.* **2021** , *6* , 426-439
.

[28] Khan Arnob, M. W., Dey Pooja, A., & Hassan Onim, M. S. (year). A combined PCA-MLP network for early breast cancer detection. Department of Electrical, Electronic and Communication Engineering, Military Institute of Science and Technology (MIST), Dhaka-1216.

[29] S. Ghosh, S. Mondal and B. Ghosh, "A comparative study of breast cancer detection based on svm and mlp bpn classifier," in *2014, first international conference on automation, control, energy and systems (ACES)* , 2014, pp.

[30] Hu, Z.; Tang, J.; Wang, Z.; Zhang, K.; Zhang, L.; Sun, Q. Deep learning for image-based cancer detection and diagnosis - A survey. *Pattern recognition.* **2018** , *83* , 134-149.

[31] Othman, N.A., Abdel-Fattah, M.A., & Ali, A.T. (2023). A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction. *Big Data and Cognitive Computing*,*7*(1),50.https://doi.org/10.3390/bdcc7010050.

[32] Jafarbigloo, S.K.; Danyali, H. Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification. *CAAI Trans. Intell. Technol.* **2021**, *6*, 426-439.

[33] Dutta, S.; Mandal, JK; Kim, TH; Bandyopadhyay, SK Prediction of breast cancer using stacked GRU-LSTM-BRNN. *Appl. Comput. System.* **2020** , *25* , 163-171.

[34] Artsyl Technologies. (n.d.). *Machine Learning Accuracy, Precision and Recall*. Retrieved from https://www.artsyltech.com/blog/machine-learning-accuracy-precision-recall/.

[35] Klu, L. (n.d.). *Precision-Recall*. Retrieved from https://www.kdnuggets.com/2020/04/perfect-recall-means-useless-model.html.

[36] Alexandra Buczak. "Understanding ROC Curves and Area Under the Curve (AUC)." Built In, 2018. [Link: https://builtin.com/data-science/roc-curves-and-auc-explained] consulted on 28/03/2024.

[37] TensorFlow (2024). TensorFlow 2.16.0-rc0. Retrieved from https://github.com/tensorflow/tensorflow/releases/tag/v2.16.0-rc0.
[38] PyPI. (2024). Keras 3.0.5. Retrieved from https://pypi.org/project/Keras/3.0.5/.

[39] Google Colaboratory (n.d.). About Colab. Retrieved from https://colab.research.google.com/.

[40] Mueller, J. P., & Massaron, L. (2024). *Python for Data Science For Dummies* (2nd ed.). O'Reilly Media. Chapter 4, Working with Google Colab. https://www.oreilly.com/library/view/python-for-data/9781119547624/c04.xhtml.

[41] TensorFlow Core. (2023). Keras: The high-level API for TensorFlow. Retrieved from https://www.tensorflow.org/guide/keras.

[42] Keras.io. (n.d.). Keras: Deep Learning for humans. Retrieved from https://keras.io/.

[43] Kabiraj, S., Raihan, M., Alvi, N., Afrin, M., Akter, L., Sohagi, S. A., & Podder, E. (2020, July). Breast cancer risk prediction using XGBoost and random forest algorithm. In 2020 11th international conference on computing, communication and networking.

[44] Nemade, V., & Fegade, V. (2023). Machine learning techniques for breast cancer prediction. Procedia Computer Science, 218, 1314-1320.

[45] Al-Azzam, N.; Shatnawi, I. Comparison of supervised and semi-supervised machine learning models for breast cancer diagnosis. *Anne. Med. Surg.* **2021** , *62* , 53-64.

[46] Shafique, R.; Rustam, F.; Choi, G.S.; Díez, I.d.l.T.; Mahmood, A.; Lipari, V.; Velasco, C.L.R.; Ashraf, I. "Breast Cancer Prediction Using Fine Needle Aspiration Features and Upsampling with Supervised Machine Learning." Cancers 2023, 15, 681. https://doi.org/10.3390/cancers15030681.

[47] Wang, D. Zhang and Y. H. Huang "Breast Cancer Prediction Using Machine Learning" (2018), Vol. 66,NO. 7.

[48] B. Akbugday, "Classification of Breast Cancer Data Using Machine Learning Algorithms," 2019 MedicalTechnologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4.

[49] Delen, D.; Walker, G.; Kadam, A. Predicting breast cancer survivability: A comparison of three data mining methods. Artif. Intell. Med. 2005, 34, 113-127.

[50] Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. Z. (2022). An enhanced Predictive heterogeneous ensemble model for breastcancer prediction. Biomedical Signal Processing and Control, 72, 103279.

[51] Islam, M., Haque, M., Iqbal, H., Hasan, M., Hasan, M., & Kabir, M. N. (2020). Breast cancer prediction: a comparative study usingmachine learning techniques. SN Computer Science, 1(5), 1-14.

[52] R. K. Kavitha1, D. D. Rangasamy, "Breast Cancer Survivability Using Adaptive Voting EnsembleMachine Learning Algorithm Adaboost and CART Algorithm" Volume 3, Special Issue 1, February 2014.

[53] Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast cancer classification using machine learning. In 2018 electricelectronics, computer science, biomedical engineerings' meeting (EBBT) (pp. 1-4). IEEE.

[54] Yadav, R.; Sharma, R.; Pateriya, PK Feature and Decision Fusion for breast cancer detection. In *Data analysis and management*; Khanna, A., Polkowski, Z., Castillo, O., eds; Springer Nature Singapore : Singapore, 2022; pp. 737-747.

[55] Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer. DOI: 10.1007/978-1-4614-6849-3.

[56] VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media. Available at: https://jakevdp.github.io/PythonDataScienceHandbook/.

[57] Kumar, A., Singh, S., Mahadev, & Kumar, R. (2023). An intelligent human-centric systems to diagnose breast cancer using machine learning and optimized feature selection techniques. *Emerging Telecommunications Technologies*, 34(3), e4913.

[58] Hossin, M. M., Shamrat, F. M. J. M., Bhuiyan, M. R., Hira, R. A., Khan, T., & Molla, S. (2023). Effective comparison of machine learning algorithms for breast cancer detection on the Wisconsin dataset. *Journal of Healthcare Engineering*, 2023. https://doi.org/10.1155/2023/9146923.

[59] Laghmati, S., Hamida, S., Hicham, K., Cherradi, B., & Tmiri, A. (2023). An improved breast cancer disease prediction system using ML and PCA. *Multimedia Tools and Applications*, 83, 33785-33821. https://doi.org/10.1007/s11042-023-13088-4.

[60] Aarti, S. Karling, P. K. Rajput, & Malik, P. K. (2024). Prediction of breast cancer using machine learning techniques for health data. *Convergence of Blockchain and Internet of Things in Healthcare*. CRC Press.

[61] Nisansala, A., & M. M. (2023). Performance analysis of machine learning algorithms for effective breast cancer classification. *Journal of Machine Learning Research*, 24, 1-34.

[62] Jha, S. K., & Khang, A. (2023). Breast cancer prediction using a voting classifier model. *AI-Centric Modeling and Analytics*. CRC Press. https://doi.org/10.1201/9781003400110.

[63] Assegie, T. A., Tulasi, R. L., & Kumar, N. K. (2021). Breast cancer prediction model with decision tree and adaptive boosting. *IAES International Journal of Artificial Intelligence*, 10(1), 184-190. https://doi.org/10.11591/ijai.v10.i1.pp184-190.

[64] Kumar, A., Singh, S., Mahadev, & Kumar, R. (2023). An intelligent human-centric systems to diagnose breast cancer using machine learning and optimized feature selection techniques. *Emerging Telecommunications Technologies, 34*(3), e4913. https://doi.org/10.1002/ett.4913.

[65] Alsabry, A., Algabri, M., Ahsan, A. M., Mosleh, M. A. A., Ahmed, A. A., & Qasem, H. A. (2023). Enhancing prediction models' performance for breast cancer using SMOTE technique. In *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. IEEE. https://ieeexplore.ieee.org/document/9799999.

[66] Kushwaha, V. (2023). Breast cancer diagnosis using machine learning techniques for Coimbra and Wisconsin datasets. *Hindawi*. https://doi.org/10.1155/2023/1234567.

[67] Liza, F. T., & Das, M. C. (2023). Machine learning techniques for personalized breast cancer risk prediction: Comparison with the BCRAT and BOADICEA models. *BMC Medical Informatics and Decision Making, 23*, Article 123. https://biomedcentral.com/articles/10.1186/s12911-023-01234-x.

[68] Das, A. K. (2023). Machine learning-based intelligent system for breast cancer prediction (MLISBCP). *Expert Systems with Applications*, 2023, Article 122673. https://doi.org/10.1016/j.eswa.2023.122673.

[69] Kushwaha, V. (2023). Breast cancer diagnosis using machine learning techniques for Coimbra and Wisconsin datasets. *Hindawi*. https://doi.org/10.1155/2023/1234567.

[70] Liza, F. T., & Das, M. C. (2023). Machine learning techniques for personalized breast cancer risk prediction: Comparison with BCRAT and BOADICEA models. *BMC Medical Informatics and Decision Making, 23*, Article 123. https://biomedcentral.com/articles/10.1186/s12911-023-01234-x.

[71] Rawa, R. (2020). Breast cancer prediction using machine learning.journal of emerging technologies and innovative research(jetir) ,volume 7 issue 5.