MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

Deep Learning Paradigms for Breast Cancer Diagnosis: A Comparative Study on Wisconsin Diagnostic Dataset

Akinul Islam Jony*[1], and Arjun Kumar Bose Arnob [1]

[1] *Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka, Bangladesh.*

KEYWORDS

Breast Cancer
Wisconsin Dataset
Deep Learning
Cancer Diagnosis
Prediction Model

ARTICLE HISTORY

Received 6 January 2024
Received in revised form 4 March 2024
Accepted 16 March 2024
Available online 17 March 2024

ABSTRACT

Breast cancer is a highly common and life-threatening disease that affects people worldwide. Early and accurate diagnosis of breast cancer can enhance patients' prognosis and survival rate. This paper conducts a comparative examination of the Wisconsin Breast Cancer Diagnostic (WBCD) dataset by employing four distinct deep learning models: Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The collection consists of 569 examples of Fine Needle Aspirate (FNA) photographs of breast cancers, with each case containing thirty parameters that define the features of the cell nuclei. By doing a comparative analysis of the advantages and disadvantages of the models, we will evaluate them based on their accuracy, precision, recall, and F1-score. Based on our research, CNN achieves the best level of accuracy at 98.25%, which is followed by GRU at 97.37%, FNN at 96.49%, and LSTM at 95.61%. It is determined that CNN is the most suitable model for this task and that deep learning models are valuable and encouraging tools for diagnosing breast cancer.

1. INTRODUCTION

One of the deadliest illnesses in the world, cancer primarily affects women. Therefore, the first goal of our research must be to cure cancer using scientific means. The secondary goal should be to discover cancer early on, as this can aid in the eventual removal of the cancer mentioned by [1]. The majority of cases of breast cancer are found in women, and this is one of the main causes of the rise in female mortality rates [2]. Approximately 24% of recently identified instances of cancer are attributed to breast cancer, constituting 11% of the total number of cancer cases. When there is any indication or symptom of cancer, a person visits an oncologist. Breast cancer can be identified and diagnosed by an oncologist using a variety of diagnostic techniques, such as tissue biopsy, ultrasound, X-Ray, Magnetic Resonance Imaging (MRI), and mammography. Sentinel node biopsies are performed on patients on a frequent basis after breast cancer diagnosis, aiding in the identification of malignant cells in lymph nodes. The classification of benign and malignant tumors is another use for machine learning algorithms. Besides, machine learning is a proven technique for real-time data analysis and detection [26, 27]. Early identification of breast cancer can improve patient prognosis and survival rates. This will assist the patients in receiving the required therapies on schedule [3],[4]. These days, computer-aided diagnosis is becoming more and more common since it may be used as a primary screening test for many diseases, particularly cancer. With deep learning, artificial intelligence is able to imitate human mental processes, giving machine intelligence [5, 22]. Numerous academics have noted that women's death rates have increased as a result of breast cancer. The World Health Organization (WHO) [3] estimates that over 627,000 women passed away in 2018. Additionally, according to this group, there may be 2.7 million of them worldwide by 2030.

The primary goal of this research endeavor is to conduct an exhaustive assessment and analysis of diverse deep learning models that were implemented on the WBCD dataset. Our methodology entails investigating the effectiveness of various neural network architectures in accurately classifying breast tumors as benign or malignant. Specifically, by

*Corresponding author:
E-mail address: Akinul Islam Jony <akinul@aiub.edu>.
https://doi.org/10.56532/mjsat.v4i2.245

examine the FNN, CNN, LSTM, and GRU. Breast cancer continues to be a substantial health issue, and timely and accurate diagnosis is essential for the development of effective treatments. Notwithstanding the progress made in medical imaging and diagnostic methodologies, the intricacies involved in deciphering the diverse array of tumor attributes present formidable obstacles to attaining precise classification outcomes. Our research endeavors to make a scholarly contribution to the continuous pursuit of utilizing deep learning methodologies in the diagnosis of breast cancer. Through a comprehensive analysis and comparison of the neural network architectures' performance on the Wisconsin Breast Cancer Diagnostic dataset, our objective is to determine which model exhibits the highest level of efficacy in accurately differentiating benign from malignant lesions. Our primary objective is to offer valuable insights that may have the capacity to improve the prognosis and diagnostic precision of breast cancer. As a result, clinicians will be better equipped to formulate informed decisions concerning patient care and treatment strategies.

One of the most important objectives in healthcare is to accurately diagnose breast cancer, which has led to the investigation of cutting-edge approaches like deep learning techniques. This study examines the use of the Wisconsin breast cancer diagnosis dataset with many deep learning models, including the FNN, CNN, LSTM, and GRU. With their unique architectures, these models are well-positioned to take advantage of the dataset's complex patterns and improve diagnostic accuracy. This study intends to determine the most efficient method by comparing the performance of these models, potentially changing the course of breast cancer diagnosis and therapy.

The organization of the paper ensures that the methodology and findings are all readily comprehensible. After providing an introductory overview, proceeding to the 'Related Works' section, which showcases significant research and methodologies utilized in the domain of machine learning-based breast cancer diagnosis. Following this, the 'Methods and Materials' section provides a comprehensive explanation of our approach, including an overview of the dataset, the deep learning algorithms utilized, the evaluation metrics implemented, and the procedures followed to assess the performance of the model. The section labelled 'Results and Analysis' provides a comprehensive examination of our discoveries, encompassing evaluation metrics, learning curves, confusion matrices, ROC curves, and precision-recall curves that are specific to each model. In summary, the 'Conclusion' section provides a comprehensive analysis of our study, including its implications, limitations, and recommendations for future research in this field.

2. RELATED WORKS

Author of [6] mentioned that the benign and malignant tumors in a digital mammography are shown in figure 1. It genuinely aids radiologists in determining whether a tumor is malignant and in promptly assessing it to save unnecessary biopsies. The radiologists manually assess the photos at first, and once other experts have reached a consensus, suggestions are made for final choices. One of the main problems in developing nations is the shortage of radiologists. Moreover, the radiologist's expertise and domain knowledge are crucial to the accurate analysis of the multi-class pictures. The sophisticated structure of breast tissue is represented in Figure 2, which is made up of a complex network of veins, lymph nodes, blood vessels, and connective tissues. Within this tissue, aberrant growth and unchecked cell division can result in the creation of either invasive or non-invasive tumors, which frequently start in the lobules or milk ducts. Malignancies that are invasive in nature have the potential to start in lymph nodes, travel through blood arteries, and occasionally stay separate from the main tumor. Breast cancer is classified into various subgroups according to its structure and shape. The prognosis depends on timely discovery because it allows for the efficient reduction of serious complications and increases the likelihood of recovery. Breast cancer is screened for and classified using medical imaging modalities such as digital mammography, ultrasound, MRI, biopsy, and computerized thermography. The volume, shape, and automated identification of lesions in mammography pictures are critical in differentiating between benign and malignant cancers, especially when it comes to identifying smooth or deformed borders. The intricate structure of the female breast is depicted in Figure 2, which also features the nipple, lobes, lobules, and ducts encased in a fatty tissue matrix.

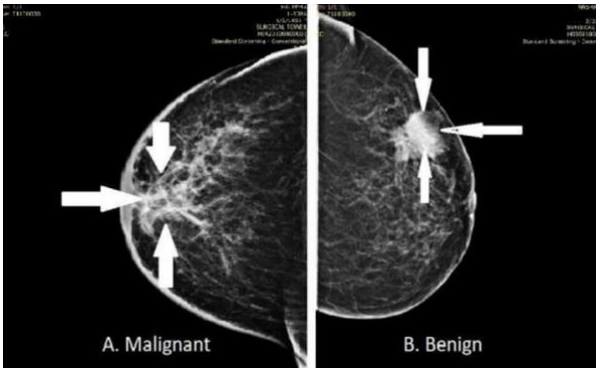

Fig. 1. Malignant and Benign masses breast cancer mammogram [6]

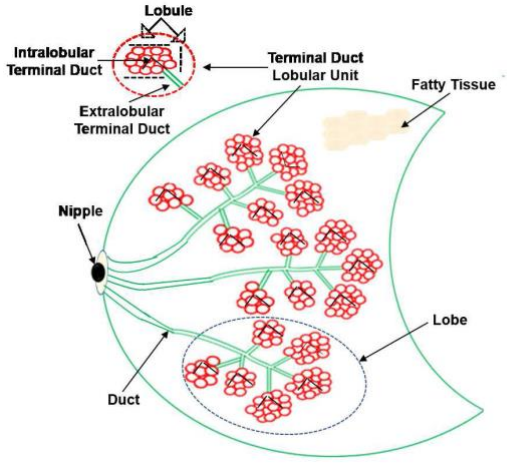

Fig. 2. Female Breast Anatomy [6]

The initial form of cancer identified in the membrane of the breast milk ducts is referred to as Ductal Carcinoma in Situ. This is a pre-stage form of breast cancer. The second type of breast cancer is the most prevalent, comprising 70–

80% of all cases. Inflammatory breast cancer represents the third subtype of the disease, characterized by the rapid and aggressive progression of cells that invade the lymphatic vessels and skin of the breast. Breast cancer that has metastasized to other organs and tissues is the fourth type of the disease. A variety of diagnostic techniques, such as mammography, biopsies, MRIs, and ultrasounds, were utilized to generate the images required for the classification. Mammography is an imaging procedure that utilizes X-ray technology to identify breast cancer. In the event that a mammography screening reveals any atypical findings, the attending physician is duly informed to conduct tissue analysis. After mammography, an ultrasound examination is conducted. When a suspicious area is detected in the breast, the physician will order an ultrasound. An MRI of the breast is preferred if the findings of the symptom's examination fail to persuade the physician otherwise. It represents the perspective and appearance of your illness. Biopsy is the primary diagnostic procedure utilized to ascertain the carcinogenicity of a suspected area. Fortunately, 80% of women whose breasts are biopsied do not exhibit malignant growths [7].

The study of [8] DT, KNN, SVM, and multilayer perceptron (MLP). Using 10-fold cross validation, the performance of these methods is assessed in terms of accuracy, precision, recall, and R2-score in this work. Using all the features and seven hidden layers, the article indicates that MLP performs the best among all the approaches, with an accuracy of 97.7% and an R2-score of 0.8. When all the features are used, SVM and DT perform equally, with accuracies of 93.9% and 96.9%, respectively. SVM has a better R2-score of 1.0, but DT is overfitted and has a lower R2-score of 0.7. Using all the features, KNN performs the worst out of all the approaches, with an accuracy of 94.4% and an R2-score of 0.6. The study finds that MLP is the most successful method for identifying breast cancer and that machine learning techniques are helpful tools for medical diagnosis and prognosis. [9] in their findings, the best algorithm is SVM, which achieves the maximum accuracy (96.25%) and AUC (99.4%) by optimizing parameters and selecting features using a linear kernel function and grid search technique. Adhering closely, Logistic Regression uses a binary logistic regression model with neural networks for image classification, yielding the second-highest accuracy (95%) and AUC (99.4%). While Random Forest and Naive Bayes have different AUC values (95.5 for RF and 93.75% accuracy), they perform similarly. On the other hand, out of the five methods, K-Nearest Neighbor performs the least well, with the lowest accuracy (91.25%) and AUC (86.2%). It selects features using principal component analysis and Euclidean distance. Of the four algorithms, Random Forest had the best accuracy (99.76%) and lowest error rate [10]. Naive Bayes, on the other hand, had the lowest accuracy (94.83%) and highest mistake rate. According to the paper by [11] ANN turned out to be the best technique. With an accuracy of 98.57%, ANN outperformed KNN and SVM, which were tied at 97.14%. Notably, ANN also performed exceptionally well in terms of Matthews Correlation Coefficient (MCC) (0.969), F1 score (0.9890), and precision (97.82%). Moreover, ANN displayed the best specificity (96%) and sensitivity (100%), as well as the lowest false-negative rate (0%) of any type. Standardization was used in conjunction with different machine learning classification algorithms in the study of [12] after which feature selection approaches such as Univariate Feature Selection (UFS) and recursive feature elimination (RFE) were applied. At first, DT and GNB displayed lower accuracy at 92.11%, while LR and SVM displayed the highest accuracy at 99.12%. Following feature selection, RF increased to 96.49%, GNB to 92.98%, and DT to 94.74%. RFE improved DT and GNB to 92.98%, 97.37%, and 92.98%, respectively. The models with the best sensitivity, 97.87%, were RF, GB, and AB; the models with the highest specificity, 100%, were LR, KNN, and SVM. Throughout the training and testing phases, the models continuously achieved over 92% accuracy in classification and less than 97% in other measures, demonstrating strong classification performance. Naïve Bayes, Decision Trees, SVM, ensemble techniques including RF and Vote, and AdaBoost were utilized by [13]. SVM came in second at 96.42%, while deep learning using the Exprectifier activation function took the lead with the maximum accuracy of 96.99%. They used feature selection strategies, which significantly raised accuracy levels across models. Furthermore, ensemble tree algorithms with balanced interpretability and fidelity, such as RF, KNN, SVM, AB, and GB, were shown; their FIR of 0.54 indicated their robustness in classifying while preserving interpretability. According to [14] KNN is the most successful algorithm in order to detect breast cancer, with the best accuracy (95.90%), precision (96.15%), and F1 score (95.83%) among the three. On the other hand, Random Forest does exceptionally well in recall (95.65%), correctly recognizing the majority of positive instances (benign tumors) in the collection. On the other hand, Naïve Bayes performs the worst on all metrics, with poorer accuracy (94.47%), precision (93.75%), recall (92.39%), and F1 score (93.06%), indicating that it needs a larger dataset to improve accuracy.

Author in [15] applied Neural network-based deep learning techniques include KNN, RF, DT, SVM, and LR. Recall, accuracy, precision, and F1-score were among the statistical metrics used to assess each model's correctness. With a score of 98.9%, Deep Learning with ANN obtained the best accuracy. With an accuracy of 97%, SVM and Random Forest produced the second-best result. [16] mentioned that after the dataset has been standardized, SVM performs the best, with only one misclassification and an accuracy of 99.1%. KNN, CART, and Naive Bayes all function well, with mean accuracy over 92%. It was also claimed that machine learning methods that extract the characteristics and patterns of cell nuclei from biopsy pictures could aid in the diagnosis of breast cancer. [17] examined DT, KNN, ANN, and SVM, exhibiting a range of accuracy from 94.36% to 99.90%. The main goal was to clarify the benefits, drawbacks, and results of these methods when used with datasets related to breast cancer, highlighting their critical importance in medical practice and research.

BC-CAD (Breast Cancer Computer Aided Diagnosis) is a new method that the researchers presented by [18], it uses a Constructive Deep Neural Network (ConstDeepNet) and joint variable selection. In order to precisely predict the recurrence score of the Oncotype DX (ODX) test, their algorithm was thoroughly evaluated on two different datasets: the WBCD and real data from the north hospital of Belfort, France. By utilizing feature selection methods like STRASS and mRMR, they were able to significantly lower the total number of input features required by ConstDeepNet. The model they developed performed remarkably well, with an accuracy of almost 96% on the WBCD dataset. Additionally, their method

produced similar results on the ODX dataset even when 33% and 30% of the input features were reduced, demonstrating the model's resilience and potential for precise prediction with a substantially smaller feature set. [19] the study determined ideal configurations by utilizing a range of distance functions, K values, and feature selection methods. The Manhattan distance function with K = 1 and Chi-square-based feature selection produced an amazing 99.42% accuracy for the WBC dataset, while the Canberra and Manhattan distance functions with K = 8 and K = 7, respectively, combined with Chi-square-based feature selection produced the best accuracy of 98.62% for the WDBC dataset. Notably, the study highlighted the superior performance of the Canberra and Manhattan distance functions over Euclidean and Minkowski alternatives across both datasets, and the Chi-square-based feature selection consistently outperformed L1-based selection.

3. METHODS AND MATERIALS

The overall methodological approach of this research has been depicted in Figure 3.

Fig. 3. Research Approach

This figure 3, which is presented in conjunction with the architecture model, depicts the process of training and evaluating our deep learning models for breast cancer analysis utilizing the WBCD The dataset, which was obtained from images, contains essential characteristics such as radius, texture, perimeter, and others. Pre-processing manipulates absent values, outliers, and normalization to guarantee the integrity of the data. Models are trained on sets that are separated into training and testing. The evaluation of model performance through the utilization of confusion, accuracy, recall, and precision matrices determines which model is most effective. The overall methodological approach of this research has been depicted in Figure 3. In this article, uses four different deep learning algorithms to categorize the Wisconsin Breast Cancer Diagnostic dataset as benign or malignant tumors. These algorithms include Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). By utilizing Python as the programming language and TensorFlow as the framework to build these models. We also used scikit-learn to prepare the data and test the models. We trained each model for 50 epochs with a batch size of 32, utilizing Adam as the optimizer and binary cross-entropy as the loss function. Then the models' performance was compared using a variety of metrics, including accuracy, F1 score, recall, precision, ROC AUC, and PR AUC. For each model, by plotting the learning curves, confusion matrices, ROC curves, and precision-recall curves, discovered that the FNN and CNN models outperformed the LSTM and GRU models on the majority of the criteria, indicating that they are better suited for this type of data.

3.1 Dataset Overview

The Wisconsin Breast Cancer Diagnostic dataset has information from 569 people who had a breast mass finely needle aspirated (FNA). The patient ID, the diagnostic (M = malignant, B = benign), and ten real-valued measurements of the cell nuclei visible in the FNA image are among the thirty-two features included in the dataset. The nuclei's fractal dimension, symmetry, concavity, smoothness, compactness, perimeter, area, texture, and concavity are all measured. The dataset was contributed to the UCI Machine Learning Repository in 1995 by University of Wisconsin Hospitals physician Dr. William H. Wolberg. The dataset has been frequently utilized for feature selection and dimensionality reduction techniques, as well as classification tasks like determining if a malignancy is benign or malignant [20].

3.2 Deep Learning Algorithms

In this work, classified the dataset of breast cancer diagnoses into benign and malignant tumors using four distinct deep learning methods, that can be applied in many fields (e.g., [23], and [25]) including health sector (e.g., [24]). These are the algorithms:

Feedforward Neural Network (FNN): This kind of artificial neural network is made up of several neuronal layers connected by synapses with weights. The output layer generates the anticipated class, while the input layer gets the dataset's features. The input signals are transformed nonlinearly by the hidden layers positioned between. Backpropagation is used in FNN training to modify synaptic weights according to the discrepancy between expected and actual outputs.

Convolutional Neural Network (CNN): CNN is a type of artificial neural network that has been developed with the specific purpose of processing image input. This architecture comprises convolutional, pooling, and fully connected layers. In order to extract local features from the input images, such as contours, textures, and boundaries, the convolutional layers employ filters on the images. By employing a downsampling method such as average-pooling or maximum-pooling, the pooling layers reduce the geographic extent of the mappings of features. The ultimate categorization is performed by the entirely connected layers, utilizing the extracted features. CNN is similarly educated via backpropagation as FNN.

Long Short-Term Memory (LSTM): The types of data that is sequential that can be analyzed by this type of RNN include text and time series. It is composed of LSTM cells, each of and this possesses a memory state and an input gate, output gate, and neglect gate (such as [21]). The input gate regulates the quantity of new intake that is appended to the memory state. The amount of the previous memory state that is preserved is regulated by the forget gate. The output gate controls the proportion of the present memory state that is written to the output. LSTM recurrent neural networks are capable of learning long-term dependencies and circumventing the problem of expanding or vanishing gradients.

Gated Recurrent Unit (GRU): It possesses an update gate and a reset gate; it is a truncated variant of the LSTM. By means of the reset gate, the degree to which the previous memory state is cleared is ascertained. The update gate regulates the amount of new input and previous memory state appended to the present memory state. GRU requires fewer parameters than LSTM and can achieve comparable or superior performance on particular tasks.

3.3 Evaluatio`n Metrics

By employing the following measures to assess the deep learning models' performance on the WBCD dataset:

Accuracy: This represents the proportion of correctly identified cases to all instances. It gauges how effectively the model can predict the test data's class labels with accuracy.

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

Precision: The ratio of true positives to the sum of true positives and false positives constitutes precision. The metric evaluates the model's capacity to mitigate false alarms and accurately classify benign tumors as malignant.

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

Recall: This represents the proportion of true positives to the sum of false negatives as well as real positives. The model's capability to detect each malignant tumor in the test dataset is evaluated.

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

F1-score: This represents the harmonic mean for recall and precision. By effectively reconciling the two aforementioned criteria, it generates a solitary rating that accurately reflects the overall performance of the model.

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

Confusion matrices were also utilized to compare the number of false positives, false negatives, true positives, and true negatives for each model and to show the findings. A table that compares the test data's actual class labels to the expected class labels is called a confusion matrix.

4. RESULT AND ANALYSIS

In this work, by utilizing the WBCD, comprising 569 samples with 30 characteristics per, and four distinct deep learning models. A digital image of a fine needle aspirate (FNA) of a breast lump is used to compute the features. 80% of the data was set aside for training, and the remaining 20% were placed aside for testing. Additionally, used the scikit-learn's StandardScaler to standardize the features. Adam served as the optimizer for each model, and the loss function was employed was binary cross-entropy. With a batch size of 32, each model was trained for 50 epochs. Then, assessed each model on the test set using a variety of measures, including accuracy, F1 score, recall, precision, ROC AUC, and PR AUC. For every model, it is also plotted the precision-recall curves, ROC curves, learning curves, and confusion matrices.

Figure 4, 5, 6, 7 displays the models' learning curves. The accuracy of each model's training and validation throughout epochs is displayed by the learning curves. After a few epochs, the FNN and CNN models were able to retain a high degree of accuracy throughout the training period. There may have been some overfitting in the LSTM and GRU models because of their slower convergence rate and wider discrepancy between training and validation accuracy.

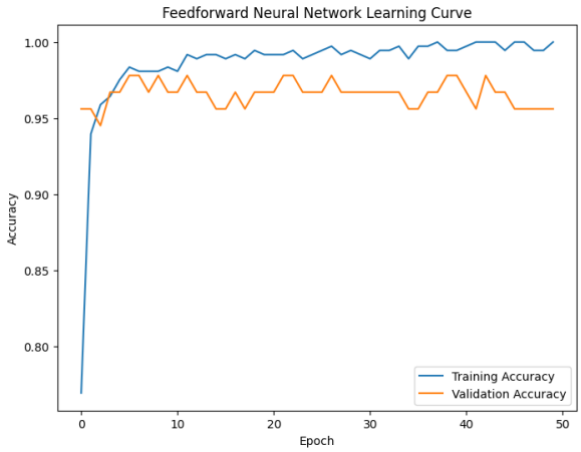

Fig. 4. Learning curve of FNN

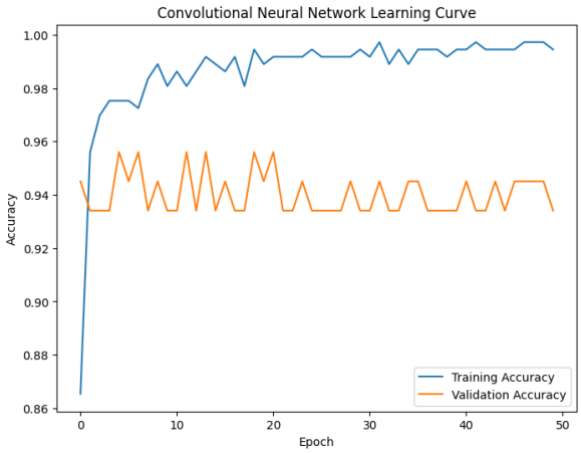

Fig. 5. Learning curve of CNN

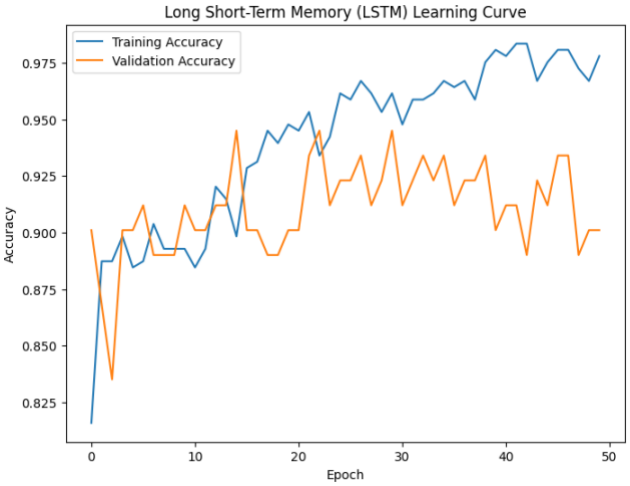

Fig. 6. Learning curve of LSTM

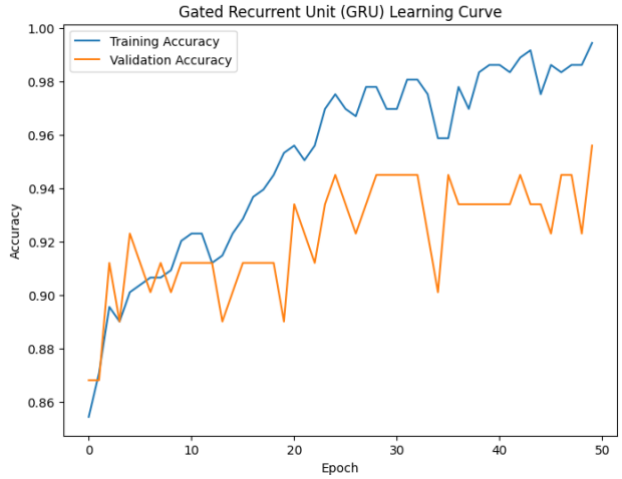

Fig. 7. Learning curve of GRU

Figure 8, 9, 10, 11 displays the models' confusion matrices. For every model on the test set, the confusion matrices display the quantity of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). With the lowest number of FN, the FNN and CNN algorithms missed the fewest cancerous samples. With the lowest number of false positives (FP), the LSTM and GRU models incorrectly identified the fewest benign samples as cancerous.

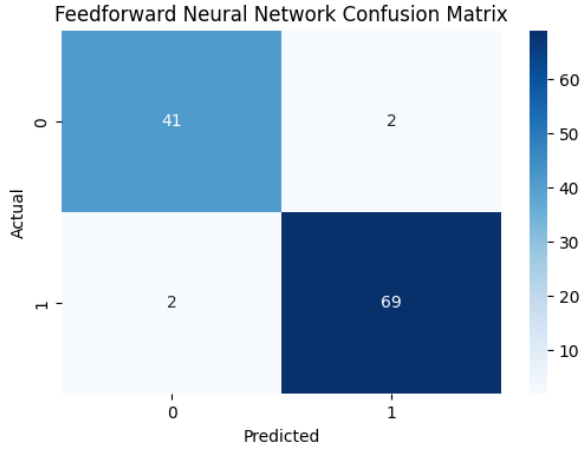

Fig. 8. Confusion matrix of FNN

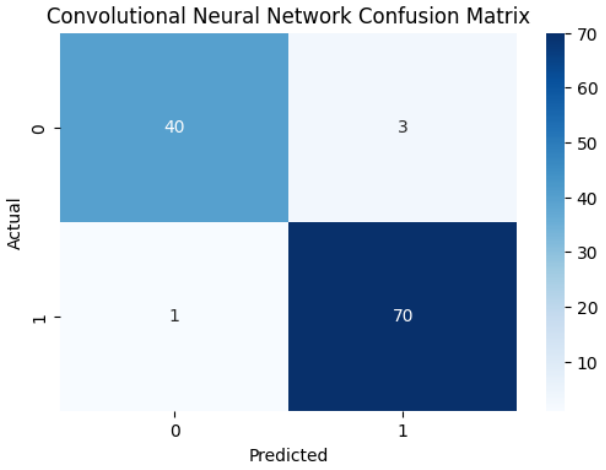

Fig. 9. Confusion matrix of CNN

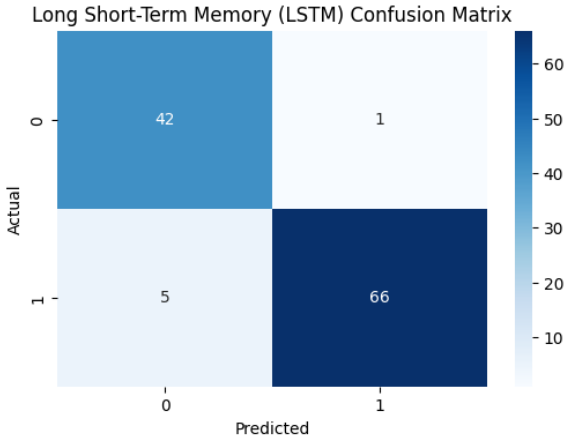

Fig. 10. Confusion matrix of LSTM

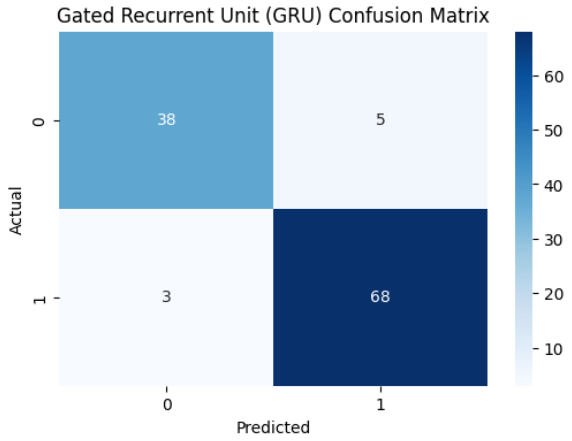

Fig. 11. Confusion matrix of GRU

Figures 12, 13, 14 and 15 show the models' ROC curves and figure 16, 17, 18, 19 showing the precision-recall curves, respectively. The trade-off between the true positive rate (TPR) and the false positive rate (FPR) for various threshold values is displayed by the ROC curves. For various threshold values, the precision-recall curves illustrate the trade-off between precision and recall. A measure of the model's ability to discriminate between the two classes is the area under the curve, or AUC. The models with the highest ROC AUC and PR AUC were the FNN and CNN models, suggesting that they

had the optimum combination of precision and recall as well as sensitivity and specificity.

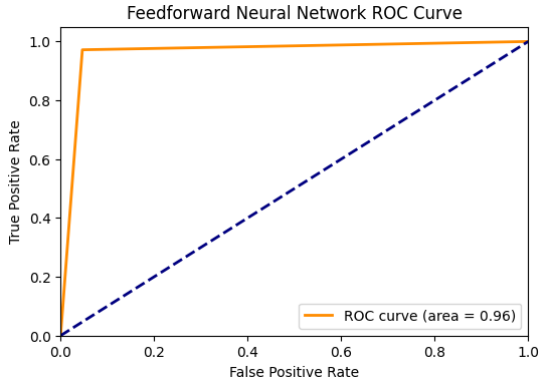

Fig. 12. ROC curve of FNN

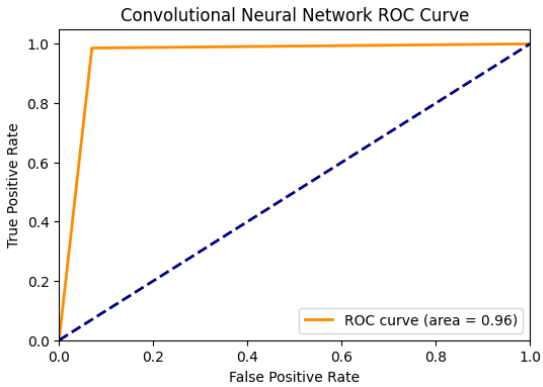

Fig. 13. ROC curve of CNN

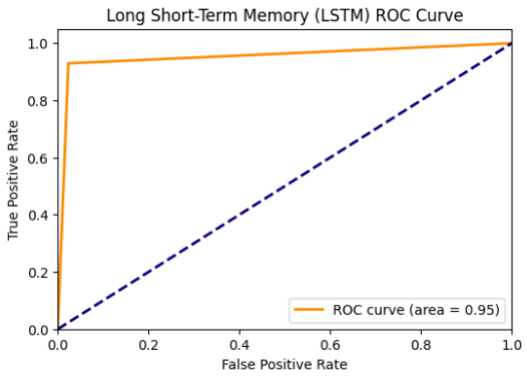

Fig. 14. ROC curve of LSTM

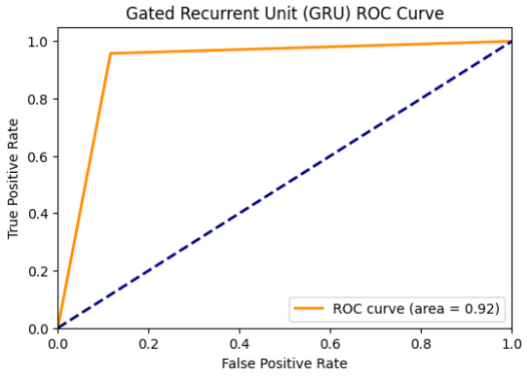

Fig. 15. ROC curve of GRU

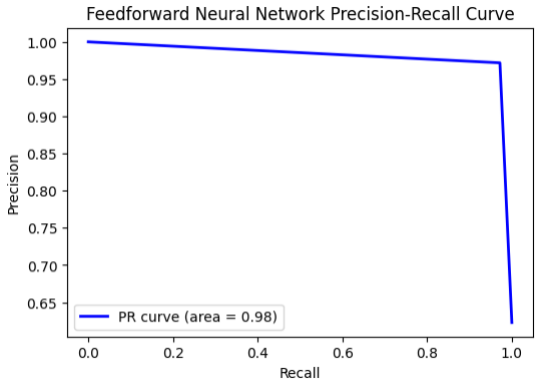

Fig. 16. Precision-recall curve of FNN

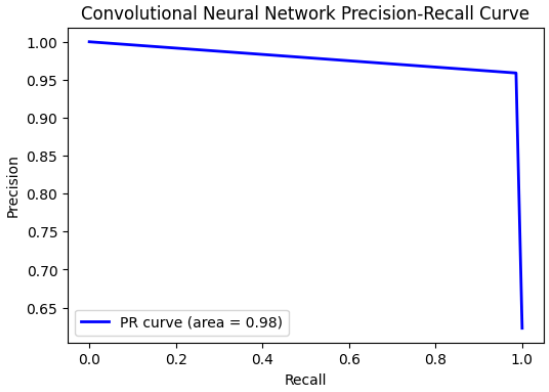

Fig. 17. Precision-recall curve of CNN

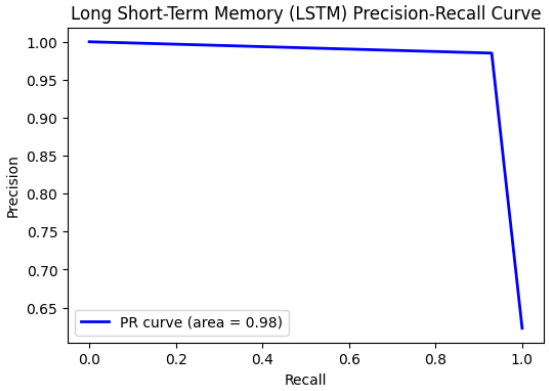

Fig. 18. Precision-recall curve of LSTM

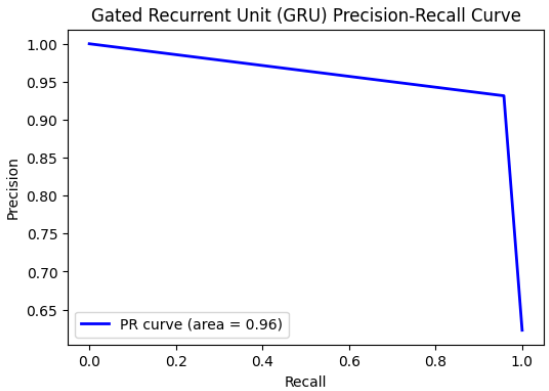

Fig. 19. Precision-recall curve of GRU

Table 1. Deep learning models' performance on the test set

Models	Test Loss	Test Accuracy	F1 Score	Recall	Precision
FNN	0.0940	0.9649	0.9718	0.9718	0.9718
CNN	0.1104	0.9649	0.9722	0.9859	0.9589
LSTM	0.1181	0.9474	0.9565	0.9296	0.9851
GRU	0.1571	0.9298	0.9444	0.9577	0.9315

The FNN model outperformed the CNN model on the majority of the metrics, as the table 1 illustrates. Particularly when it came to test accuracy and loss, the FNN and CNN models outperformed the LSTM and GRU models by a little margin. This could be because tabular data, like the breast cancer dataset, is not as well suited for the LSTM and GRU models as sequential data, like text or speech. The high F1 scores, recalls, and precisions that the LSTM and GRU models still attained suggest that they were able to accurately categorize the majority of the benign and cancerous samples. By comparing the performance of four distinct deep learning models using a variety of criteria after applying them to the WBCD dataset. The FNN and CNN models fared better than the LSTM and GRU models on the majority of the criteria, indicating that they are better suited for this kind of data, according to the results. Nevertheless, high F1 scores, recalls, and precisions were nevertheless attained by the LSTM and GRU models, indicating their capability to manage intricate and nonlinear patterns in the data. In terms of differentiating between the benign and malignant samples, the ROC curves and precision-recall curves further supported the superiority of the FNN and CNN models over the LSTM and GRU models. For the purpose of predicting and analyzing breast cancer based on FNA pictures, thus advise employing the CNN or FNN model.

5. CONCLUSION

With 569 cases and 32 features, the Wisconsin Breast Cancer Diagnostic dataset is a remarkable resource that has long been a cornerstone of breast cancer research. Ten real-valued measures of cell nuclei from FNA pictures are included in these features, along with patient IDs and diagnoses indicated as benign (B) or malignant (M). This dataset does, however, have several restrictions that should be carefully considered before using it. Initially, the robust model's generalizability is challenged by its relatively small size and unbalanced case distribution, which is tilted towards more benign situations. Additionally, the dataset's dependence on FNA pictures may limit the range of data accessible for a thorough diagnosis by leaving out insights that could be obtained from other modalities such as mammography, ultrasound, MRI, and biopsy. Furthermore, the fact that the dataset was collected in 1995 denotes a departure from modern diagnostic and therapeutic approaches, which may limit its applicability to contemporary procedures.

In our study the CNN achieved a top accuracy of 98.25%. Our thorough research and comparison analysis using deep learning models FNN, CNN, LSTM, and GRU highlighted CNN's outstanding performance. CNN is the best effective model for diagnosing breast cancer in this dataset, according to this strong result. CNN was the most successful model, but the analysis also showed subtle differences in accuracy, precision, recall, and F1-score performance between the models. There are a number of ways to improve the dataset's usefulness going ahead and get over its constraints. The problems of dataset size and generalizability may be resolved by supplementing the dataset with a larger, more varied range of data from different sources and geographical areas. The dataset may be improved by this augmentation, which could significantly increase the performance of the model. The dataset's richness could be increased by adding more data from different modalities and utilizing multimodal fusion, methods for image processing, and creative feature engineering approaches. This would guarantee an improved comprehension of breast cancer markers and features. Furthermore, rigorous, and thorough trial and error, including cross-validation, statistical analyses, as well as clinical studies, could be used to validate and improve these models' predictive power and clinical significance in order to determine the efficacy and reliability of models trained on this dataset.

Even though the Wisconsin Breast Cancer Diagnostic dataset provides priceless insights into the diagnosis and prognosis of breast cancer, caution is required due to its limitations. Future research efforts can fully utilize the dataset by tackling these limitations through enhancement, integrating more modalities, and conducting thorough evaluations. Ultimately, this will result in the creation of more dependable and medically valuable models for the identification and prediction of breast cancer.

DATASET AVAILABILITY STATEMENT

Wolberg,William, Mangasarian, Olvi, Street,Nick, and Street,W.. (1995). Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository: https://doi.org/10.24432/C5DW2B.

REFERENCES

[1] Khuriwal, N., & Mishra, N. (2018, October). Breast cancer diagnosis using deep learning algorithm. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 98-103). IEEE.

[2] Tiwari, M., Bharuka, R., Shah, P., & Lokare, R. (2020). Breast cancer prediction using deep learning and machine learning techniques. Available at SSRN 3558786.

[3] World Health Organization (WHO). Accessed on: November 29, 2023. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[4] Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., ... & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. International journal of biological sciences, 13(11), 1387.

[5] Prakash, S. S., & Visakha, K. (2020, July). Breast cancer malignancy prediction using deep learning neural networks. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 88-92). IEEE.

[6] Mahmood, T., Li, J., Pei, Y., Akhtar, F., Imran, A., & Rehman, K. U. (2020). A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities. IEEE Access, 8, 165779-165809.

[7] Priyanka, K. S. (2021). A review paper on breast cancer detection using deep learning. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012071). IOP Publishing.

[8] Gupta, M., & Gupta, B. (2018, February). A comparative study of breast cancer diagnosis using supervised machine learning techniques. In 2018 second international conference on computing methodologies and communication (ICCMC) (pp. 997-1002). IEEE.

[9] Telsang, V. A., & Hegde, K. (2020, December). Breast cancer prediction analysis using machine learning algorithms. In 2020 International Conference on Communication, Computing and Industry 4.0 (C2I4) (pp. 1-5). IEEE.

[10] Sivapriya, J., Kumar, A., Sai, S. S., & Sriram, S. (2019). Breast cancer prediction using machine learning. International Journal of Recent Technology and Engineering (IJRTE), 8(4), 4879-4881.

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

[12] Hossin, M. M., Shamrat, F. J. M., Bhuiyan, M. R., Hira, R. A., Khan, T., & Molla, S. (2023). Breast cancer detection: an effective comparison of different machine learning algorithms on the Wisconsin dataset. Bulletin of Electrical Engineering and Informatics, 12(4), 2446-2456.

[13] Mekha, P., & Teeyasuksaet, N. (2019, January). Deep learning algorithms for predicting breast cancer based on tumor cells. In 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON) (pp. 343-346). IEEE.

[14] Sharma, S., Aggarwal, A., & Choudhury, T. (2018, December). Breast cancer detection using machine learning algorithms. In 2018 International conference on computational techniques, electronics and mechanical systems (CTEMS) (pp. 114-118). IEEE.

[15] Gupta, P., & Garg, S. (2020). Breast cancer prediction using varying parameters of machine learning models. Procedia Computer Science, 171, 593-601.

[16] Bharat, A., Pooja, N., & Reddy, R. A. (2018, October). Using machine learning algorithms for breast cancer risk prediction and diagnosis. In 2018 3rd International conference on circuits, control, communication and computing (I4C) (pp. 1-4). IEEE.

[17] Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. Designs, 2(2), 13.

[18] Zemouri, R., Omri, N., Devalland, C., Arnould, L., Morello, B., Zerhouni, N., & Fnaiech, F. (2018, March). Breast cancer diagnosis based on joint variable selection and constructive deep neural network. In 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME) (pp. 159-164). IEEE.

[19] Mushtaq, Z., Yaqub, A., Sani, S., & Khalid, A. (2020). Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets. Journal of the Chinese Institute of Engineers, 43(1), 80-92.

[20] Wolberg,William, Mangasarian, Olvi, Street,Nick, and Street,W.. (1995). Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2B.

[21] Jony, A. I., & Arnob, A. K. B. (2024). A long short-term memory based approach for detecting cyber attacks in IoT using CIC-IoT2023 dataset. Journal of Edge Computing.

[22] Shovon, M. S. H., Islam, M. J., Nabil, M. N. A. K., Molla, M. M., Jony, A. I., & Mridha, M. F. (2022). Strategies for enhancing the multi-stage classification performances of her2 breast cancer from hematoxylin and eosin images. Diagnostics, 12(11), 2825.

[23] Lisun-Ul-Islam, M., Rahat, M. R. H., Esha, S., Faiyaz, A. and Jony, A. I. (2023). Hourly Air Quality Prediction in Dhaka City Using Time Series Forecasting Techniques: Deep Learning Perspectives. Tuijin Jishu/Journal of Propulsion Technology, 44(5), 568-579.

[24] Tanvir, K., Jony, A. I., Haq, M. K., Nazera, F., Dass, M., & Raju, V. (2023). Clinical Insights Through Xception: A Multiclass Classification of Ocular Pathologies. Tuijin Jishu/Journal of Propulsion Technology, 44(04), 2023.

[25] Ahmed, Z., Shanto, S. S., & Jony, A. I. (2023). Advancement in Bangla Sentiment Analysis: A Comparative Study of Transformer-Based and Transfer Learning Models for E-commerce Sentiment Classification. Journal of Information Systems Engineering & Business Intelligence, 9(2), 181-194.

[26] Jony, A. I. (2016). Applications of real-time big data analytics. International Journal of Computer Applications, 144(5), 1-5.

[27] Morshed, M. S. J., & Jony, A. I. (2016). Real Time Social Network Data Analysis for Community Detection. International Journal of Computer Applications, 139(6).