



Predictive Analytics for Disease Diagnosis: A Study on Healthcare Data with Machine Learning Algorithms and Big Data

Purna Chandra Rao Chinta^{1*}, Chethan Sriharsha Moore², Laxmana Murthy Karaka³, Manikanth Sakuru⁴, Varun Bodepudi⁵ and Srinivasa Rao Maka⁶

¹Microsoft, Sr Technical Support Engineer

²Microsoft, Sr Technical Support Engineer

³Code Ace Solutions Inc, Software Engineer

⁴JP Morgan Chase, Lead Software Engineer

⁵Deloitte Consulting LLP, Senior Solution Specialist

⁶North Star Group Inc, Software Engineer

*Address for Correspondence

Purna Chandra Rao Chinta, Microsoft, Sr Technical Support Engineer
Email Id: chpurnachandrarao@gmail.com

Submission: 04 January, 2025

Accepted: 31 January, 2025

Published: 03 February, 2025

Copyright: © 2025 Chinta PCR, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: World Health Organization (WHO); Breast Cancer; Tumour; Machine Learning; Healthcare; Disease Diagnosis; Feedforward neural network (FFNN); Random Forest (RF); Decision Tree (DT); Convolution Neural Network (CNN)

Abstract

At now, breast cancer ranks second among women in terms of cancer-related deaths, making it a major epidemiological issue. The illness is not caught early enough, and half of the one million women diagnosed with breast cancer annually die from the condition. This research aims to predict the occurrence of breast cancer using various ML algorithms, including Feed forward Neural Network, Random Forest, and Decision Tree, with the goal of reducing the risk of death from this disease, which is a second most common cause of death among women globally. This research uses the Breast Cancer Wisconsin (Diagnostic) dataset to assess ML models that may diagnose breast cancer. The FNN model outperformed RF and DT, achieving the best overall performance with a precision, recall, and accuracy of 97.18%. These results highlight the FNN's robustness in minimising false positives and maximising true positives, making it a reliable tool for breast cancer diagnosis. To further enhance the accuracy of feature extraction and classification, future research may look at incorporating stronger deep learning models such transformer architectures and Convolution Neural Networks (CNNs). The model's generalisability and clinical usefulness might be further validated by using bigger and more varied datasets.

Introduction

In terms of mortality rates, cancer ranks high among both sexes worldwide. The most common malignancies are breast cancer, which kills more women than any other illness and affects more women than any other disease in the world. Breast cancer may be detected early, which could lead to a survival rate of up to 80%, according to the WHO [1]. There are almost 1.7 million new instances of breast cancer identified each year, with 500,000 people losing their lives to the condition. Unfortunately, these figures might rise in the years to come [2, 3]. Dense breast tissue, a personal or family history of breast cancer, an older maternal age, the use of certain medications or procedures during pregnancy, drinking alcohol, and other behavioral variables are all potential risk factors for this kind of cancer [4]. The impact of certain factors is substantial, while that of others is rather little. Being a woman and getting older are unchangeable facts, but may lessen our risk of breast cancer by living a healthy lifestyle.

There are three main ways to detect breast cancer: a physical exam, a mammogram, or a biopsy. Professional radiologists are required to interpret the results of these diagnostic procedures; nonetheless, mammography is by far the most prevalent [5]. The problem with having several readings of the same mammography is that various radiologists get different conclusions. There is a 65%

to 78% accuracy rate for mammography. The malignant nature of a tumour found by mammography may be determined by doing a biopsy [6]. Although the accuracy rate of a biopsy is almost 100%, the procedure is nonetheless invasive, expensive, time-consuming, and unpleasant [7]. These issues make it more challenging for clinicians to diagnose benign or malignant tumours. Because of these factors, ML techniques have the potential to greatly impact the diagnostic process.

The application of AI techniques for the early diagnosis of breast cancer has recently increased. Learning theory is one kind of AI. For the most part, healthcare organisations have used ML and DL algorithms for breast cancer diagnosis [8]. The diagnostic accuracy of a patient utilised to be entirely dependent on the knowledge and skill of the doctor [9]. The accumulation of a physician's expertise is the result of years of closely observing patients' symptoms [10]. However, the accuracy is unreliable. It is now simpler to collect and store data because of the development of computer tools. Thus, the field of intelligent healthcare systems is dependable and beneficial. These technologies may assist doctors in diagnosing patients by providing them with relevant and reliable standards. Individuals might also benefit from these developments in terms of future health planning. This is how ML can take over the laborious physical tasks that healthcare workers face every day [11, 12].

Motivation and Contributions of the Study

This project aims to explore the use of ML algorithms on the Breast Cancer Wisconsin (Diagnostic) dataset to assess predictive analytics' potential in disease diagnosis. By determining which model, out of many options including FNN, RF, and Decision Tree, performs the best, the study significantly advances medical diagnosis. The key contributions are:

Collect the Wisconsin Diagnostic Breast Cancer dataset for breast cancer detection.

Applied essential data preprocessing steps, including the removal of duplicates, handling missing values, enhance data quality and model performance.

Applied standardisation to scale the features of the dataset, transforming them to have a mean of zero and unit variance.

Apply ML models like FNN, RF, DT for breast cancer detection.

Evaluated model performance employing accuracy, precision, recall, F1-score, and AUC, focusing on a comprehensive understanding of predictive capabilities, especially for imbalanced datasets.

Organization of the paper

Presented below is the outline of the paper: Section II finds research gaps and evaluates pertinent literature. Section III details the methodology, including data collection and the machine learning models used. Section IV presents the results of the experiments and the analysis of the model's performance. Section V wraps up the report by reviewing the results and offering suggestions for further study.

Literature Review

They summarise the research on breast cancer prediction and categorisation in this section. Classification methods were the primary emphasis of the literature studied. Some reviews are:

In this study, Khuriwal and Mishra (2018) proposed using the Wisconsin Breast Cancer database in an adaptive ensemble voting method for breast cancer diagnosis. This study aims to examine and explain how logistic and ANN algorithms, in conjunction with ensemble ML algorithms, produce improved results for breast cancer diagnosis, even when the number of variables is decreased. Wisconsin Diagnosis Breast Cancer was the dataset used in this research. When contrasted with similar literature. The Artificial Neural Network (ANN) technique achieved a 98.50% accuracy rate while utilising the logistic algorithm, according to an alternate ML methodology [13].

In this study, Gecer et al. (2018) provide a method for creating five diagnostic categories from breast biopsy whole slide images (WSI). A saliency detector using four fully convolutional networks trained with data extracted from pathologists' screening records is an integral part of the WSI diagnosis process. After that, this detector will locate diagnostically important regions using multi-scale methods. Then, image patches are classified using a convolutional network based on whether they are invasive cancer, ductal carcinoma in situ, atypical ductal hyperplasia, proliferative changes, or non-proliferative. The network is trained using reference samples collected from consensus. At last, the saliency and classification maps are combined to label pixels and classify slides, respectively. Both the saliency detector and classifier networks outperformed rival algorithms in experiments, including 240 WSI. There was no significant difference between the 45 pathologists' opinions and a five-class slide-level accuracy 55%. Breast cancer diagnostic visualisations using the learnt representations are also offered [14].

In this study, Chen et al. (2017), optimise techniques for ML to accurately forecast the onset of chronic diseases in populations prone to such outbreaks. A chronic illness of the brain, cerebral infarction, is the subject of our experiments. They present a novel multimodal illness risk prediction method that utilises hospital structured and unstructured data and is based on CNNs. No prior research in

the field of medical big data analytics has, as far as they are aware, addressed both forms of data simultaneously. With a convergence time of only 94.8% and a prediction accuracy that surpasses that of most conventional methods, the proposed method significantly exceeds a CNN-based unimodal disease risk prediction algorithm [15].

In this study, Sahoo, Mohapatra and Wu (2016) a probabilistic data-gathering technique is created, and then the acquired data was analysed for correlation. Lastly, a stochastic prediction model is made to forecast the future health state of the most related folks based on their existing status. Extensive cloud-based simulations allow for the performance assessment of the suggested protocols; these simulations achieve a prediction accuracy of around 98% while reducing analysis time by 90% while maintaining 90% CPU and bandwidth utilisation [16].

In this study, Abdel-Zaher and Eldeib (2016) by combining the unsupervised route of a deep belief network with the supervised path of back propagation, a CAD technique for breast cancer diagnosis has been created. The architecture is a Backpropagation Neural Network (BPN-NN) trained using the Liebenberg-Marquardt learning function, with weights initially set using the Deep belief network (DBN-NN) route. They validated our method using the WBCD or Wisconsin Breast Cancer Dataset. A 99.68% accuracy rate from the classifier complex is encouraging when compared to other published research. As a breast cancer categorization model, the suggested approach works well. A number of train-test partitions were also considered when analysing the design [17].

In this study, Kandaswamy et al. (2016) are very interested in the use of state-of-the-art ML techniques, such as DNNs, to categorise substances involved in chemical MOAs. To classify compounds, image-based profiling techniques have been used, sometimes in conjunction with feature reduction techniques like factor or PCA. This article demonstrates how to classify MOAs based on cell input properties independently of treatment profiles and feature reduction techniques. Our best understanding is that this is the first use of DNN using single-cell data in this field. Additionally, they employ DTL to reduce the computationally strenuous and time-consuming process of scouring the vast parameter space of a DNN. The outcomes indicate that this method results in a 30% increase in efficiency and a 2% increase in accuracy [18].

Methodology

This investigation is designed to assess an efficacy of ML models in a detection of breast cancer by employing a Breast Cancer Wisconsin (Diagnostic) dataset. A following steps of research design are shown in Figure 1 flowchart. Data preprocessing is conducted to ensure the dataset is clean and ready for analysis, including removing duplicate entries and handling missing values. Feature scaling is implemented through standardisation to normalise the data, guaranteeing that all features contribute equitably to the model's efficacy. The pre-processed data was then split into training (80%) and testing (20%) sets. A variety of classification models, including FFNN, RF, and DT, were ultimately used. In order to determine the most successful model for breast cancer diagnosis, key measures such as F1-score, recall, accuracy, and precision were used to evaluate each prototype.

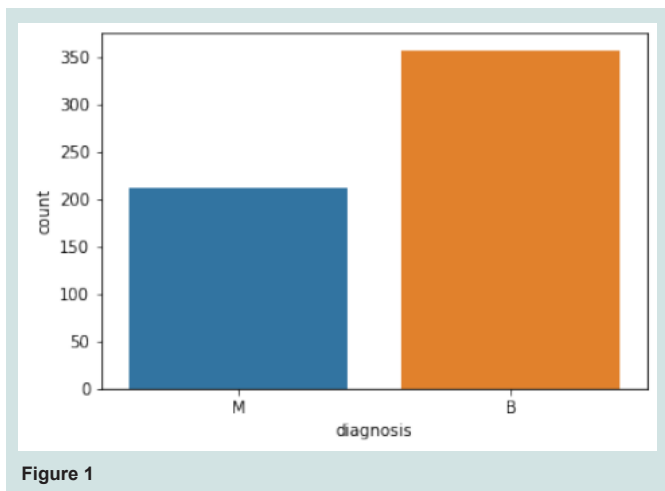


Figure 1

Flowchart for Breast cancer Diagnosis

The following steps of a flowchart are briefly explained below:

Data Collection

This study makes use of the Wisconsin Breast Cancer (Diagnostic) Data Set, which is a dataset specifically designed for this purpose. There are 569 samples in the collection, and each sample has 32 visually assessed atomic characteristics calculated from an image of a breast mass’s FNA. The distribution of benign (B) and malignant (M) tumours, as diagnosed, is displayed in (Figure 2).

This study makes use of the Wisconsin Breast Cancer (Diagnostic) Data Set, which is a dataset specifically designed for this purpose. There are 569 samples in the collection, and each sample has 32 visually assessed atomic characteristics calculated from an image of a breast mass’s FNA. The distribution of benign (B) and malignant (M) tumours, as diagnosed, is displayed in (Figure 2).

Class Distribution of data

A class distribution analysis in (Figure 2) shows an imbalanced dataset, with the majority class (“B”) having significantly more instances than the minority class (“M”). This imbalance can lead to model bias, causing poor performance in a minority class. In these cases, standard accuracy may not be a reliable statistic. Thus, it’s better to use evaluation metrics like precision, recall, F1-score, and AUC, which provide a more realistic evaluation of the model’s performance on imbalanced datasets.

Correlation Matrix of Data

(Figure 3’s) heatmap shows the relationship between several attributes, probably taken from a dataset. A color intensity varies from dark purple to light pink, indicating the strength of correlation ranging from about 0.45 to 0.90. This kind of visualisation is useful for understanding how different variables in a dataset are related to each other, which is crucial for feature selection in model-building processes.

Data Preprocessing

Data preparation in the context of breast cancer detection utilising the Wisconsin dataset entails cleaning, organising, and

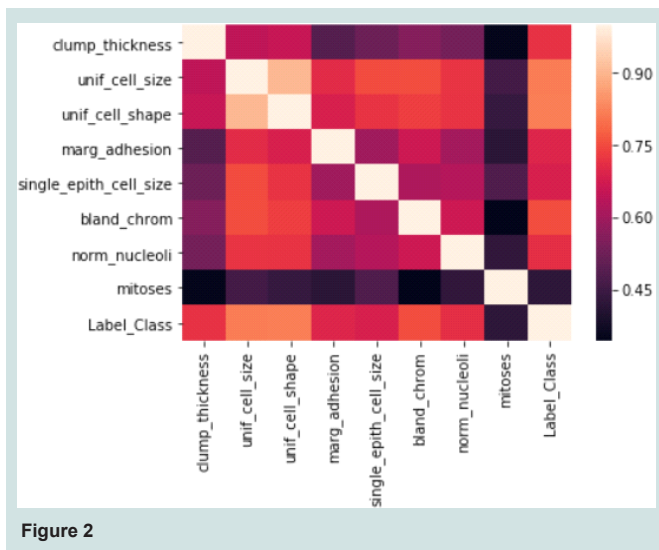


Figure 2

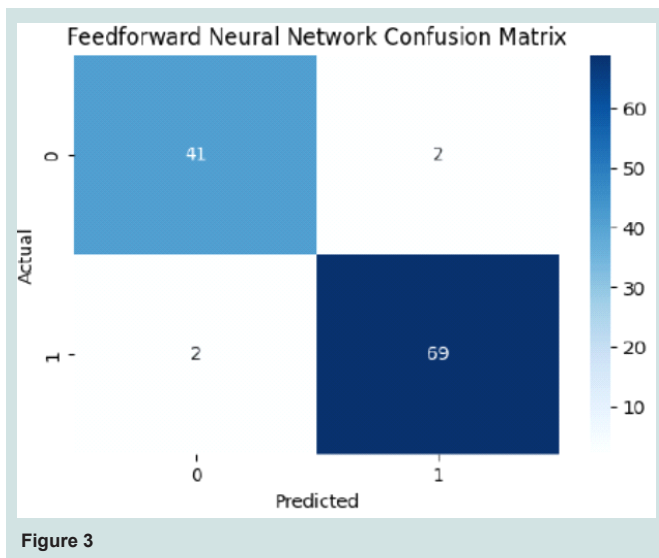


Figure 3

standardising the data so that it may be used to develop accurate and trustworthy diagnostic models [19]. A vital part of data cleaning is removing duplicates, which makes sure that there aren’t any records that are both relevant and redundant, which could affect the accuracy of analyses and models. Key pre-processing steps are listed below:

Removing duplicate Entries

Data cleaning and preparation include removing duplicates to make sure the data is correct and dependable for modelling or analysis.

Handling Missing values

Data points lacking a value for a particular variable in a dataset are called missing values [20]. Data analysis becomes much more difficult due to missing data points, which might cause results to be biased or erroneous.

Feature Scaling

Machine learning also makes use of the standard scaler, sometimes

known as standardisation, to scale features. Each feature is averaged to a zero-variance mean using this procedure [21, 3]. Although this method does not restrict the data to a certain time frame or alter its distribution, it does guarantee that the majority of data points will be close to 0. This indicates that no matter how much data is scaled, outliers will remain. Equation 1 shows the definition of standard scaling.

Where:

x_{scaled} = scaled sample point

x = sample point

\bar{x} = mean of the training samples

σ = standard deviation of the training samples

Data Splitting

There are two subsets of the dataset: the training set and the testing set. The model is trained using the training set, and its performance is evaluated using the test set. The Data was divided into the 80:20.

Proposed Feed forward Neutral Network (FNN) Models

DNNs are computer models that use a layer-by-layer architecture and a large number of neurones (node) linked together by synaptic connections (weights) [22-24]. As a result, FNNs adhere to a particular architectural arrangement in which each layer's nodes are linked to the layer below them via forward connections [25]. The limited number of neurones in a single internal hidden layer of a FNN allows it to approximate any continuous function with an activation function that is continuous and sigmoidal in nature. The connection weights provide input that a node in a FNN can process. It is possible to calculate the mathematical output y_i (excitation) of a node (node i) as (2):

Where:

\sum is a total incoming connection,

x_j is an input,

w_{ij} is a weight,

b_i is a bias, and

$f(\cdot)$ determines a range of possible values for the i -th node's output amplitude, which is controlled by the activation function.

Evaluation Parameters

F1 score, recall, accuracy, and precision are important performance indicators for evaluating a model's efficacy and helping to comprehend its predictive skills as well as finding places for development [26]. The equations of metrics, as shown in, are based on the fundamental measuring parameters of the confusion matrix, The following factors should be taken into account when measuring the parameters:

True Positive (TP): Correctly identify the presence of disease.

True Negative (TN): Correctly forecast the absence of disease.

False Positive (FP): Incorrectly forecast the disease is present when it's not.

False Negative (FN): Fail to detect the disease when it is present.

Accuracy: Equation 3 offers the formula for calculating the percentage of true outcomes (including TN and TP) relative to a total number for gauging accuracy:

Precision: the proportion of states that were considered interesting (loaded in this case) and truly exist in that state for the purpose of measuring precision. In Equation 4:

Recall: the percentage of intriguing states accurately identified as such [25]. Another name for it is Recall or Sensitivity [27]. The formula for measuring recall is mentioned in Equation 5:

F1-score: a measure of both accuracy and memory that yields the proportion of correctly identified occurrences [28]. The formula for measuring F1-score are mentioned in Equation 6:

The Findings are obtained by evaluating the model's performance using these performance metrics on the testing set.

Result Analysis and Discussion

Here, the outcomes for the various classification systems used in this study are examined. Our research employed ML techniques for an effective detection of breast cancer, specifically focusing on models such as FNN that compare with RF [29] and DT [30] shown in (Table III). An effectiveness of these algorithms was evaluated employing the Breast Cancer Wisconsin (Diagnostic) dataset. Important performance measures utilised to assess a model's utility were F1-score, recall, accuracy, and precision. (Table II). shows the results of the proposed model.

Bar Graph for Performance of FNN Model

(Table II) and (Figure 4) presents a performance of the FNN model for breast cancer prediction. The model achieved an accuracy of 96.49%, with both precision and recall scores reaching 97.18%. Additionally, the F1-score was also 97.18%, indicating a well-balanced performance according to both sensitivity and specificity. These results suggest that the FNN model is highly effective for predicting breast cancer, demonstrating strong classification performance across key metrics. [31 -40]

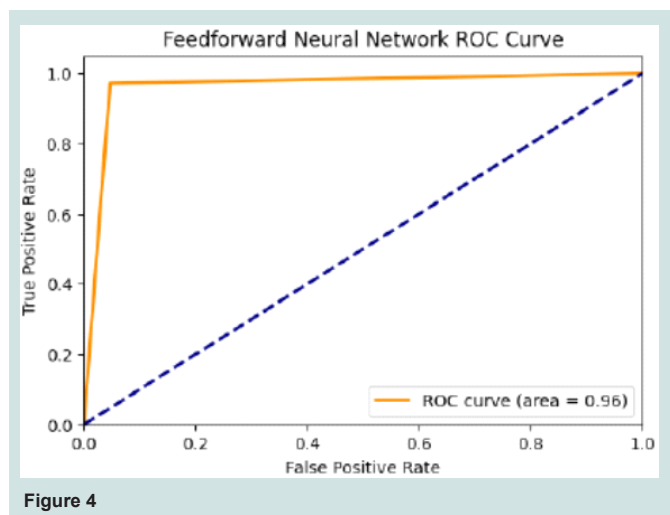


Figure 4

Table 1: provides an overview of different approaches to breast cancer classification and prediction, showcasing various methodologies, datasets, and performance outcomes.

References	Methodology	Dataset	Performance	Limitations & Future Work
Khuriwal and Mishra (2018)	Deep learning technology applied to WisconsinBreast CancerDataset	Wisconsin Breast Cancer Dataset	99.67% accuracy	Reduced variables may affect generalisation; exploring more variables for better performance.
Gecer et al. (2018)	Saliency detector with a pipeline of four fully convolutional networks; convolutional network for classification.	Whole Slide Images (WSI) of breast biopsies	Five-class slide-level accuracy of 55%, comparable to 45 pathologists.	Limited accuracy for multi-class classification; explore improvement in saliency detection and classification.
Chen et al. (2017)	Modified prediction models using a CNN-based multimodal disease risk prediction algorithm	Real-life hospital data from central China (2013-2015)	94.8% accuracy, faster convergence than unimodal CNN-based algorithms.	Future work may involve enhancing data types and exploring other chronic diseases.
Sahoo, Mohapatra, and Wu (2016)	Probabilistic data collection and stochastic prediction model	Health data from cloud simulations	98% accuracy, 90% CPU and bandwidth utilisation	Limited by cloud simulation environment, future work could explore real-world deployment.
Abdel-Zaher and Eldeib (2016)	CAD scheme using deep belief network followed by a back-propagation neural network.	Wisconsin Breast Cancer Dataset	99.68% accuracy	Limited test partitions; investigate broader dataset use for better validation.
Kandaswamy et al. (2016)	DNN with deep transfer learning for compound classification in chemical mechanisms of action.	Chemical data (MOA classification)	30% speedup, 2% accuracy improvement.	Focused on chemical mechanisms, not breast cancer; explore more diverse applications and models.

Summary of the related work on Breast cancer prediction using machine and deep learning technologies

Table 2: Finding of Feedforward Neural Network (FNN) Performance for Breast Cancer Prediction

Performance Metrics	Feed forward Neural Network (FNN) Model
Accuracy	96.49
Precision	97.18
Recall	97.18
F1-score	97.18

Confusion matrix for FNN Model

In the (Figure 5) displays a confusion matrix for a FNN, presented in a 2x2 grid format. The matrix is labelled with “Actual” on a y-axis and “Predicted” on an x-axis, indicating two classes (0 and 1). The values within the matrix are: 41 TN and TP, representing correct predictions for classes 0 and 1, respectively. There are 2 FP and 2 FN, illustrating instances where the predictions did not match the actual classes. The matrix uses a colour gradient from light to dark blue to represent the range of values from low to high, accompanied by a colour bar on the right showing values from 0 to 60.

ROC curve of FNN Model

The (Figure 6) displays a ROC curve for a FNN. It plots the TPR against the FPR across a range of threshold values. The ROC curve, represented by a solid orange line, sharply rises close to a top-left corner of a graph, suggesting high model performance. The AUC is notably high at 0.96, indicating excellent discriminative ability. A dashed blue line, representing a random classifier’s performance, diagonally divides the plot, providing a baseline for comparison.

Precision-recall curve of FNN Model

The FNN model’s Precision-Recall curve in (Figure 7) shows strong performance with an AUC of 0.98, indicating good identification of true positives. Initially, precision is high, but it gradually drops as recall increases, typical of precision-recall curves. The sharp drop in precision suggests potential class imbalance, as the model sacrifices precision to capture more true positives.

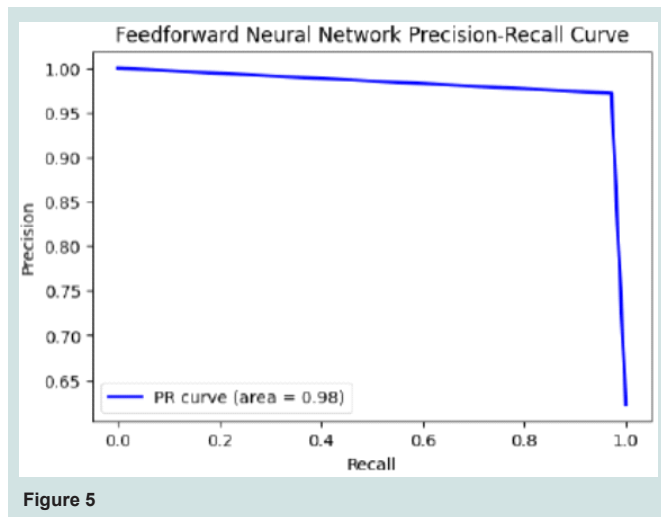


Figure 5

(Table III). above displays the outcomes of comparing a models’ performance. Among an algorithms tested, FNN performed best with an accuracy of 96.49%, surpassing both RF (94.11%) and DT (94.3%). The FNN’s remarkable recall and precision scores of 97.18% show that it is very good at reducing FP and increasing real positive detections. In comparison, the RF model demonstrated a precision of 94.97% and a recall of 94.11%, reflecting its effectiveness but slightly lower than that of the FNN. The Decision Tree model, while achieving a respectable accuracy of 94.3%, had lower precision at 91.0% and recall at 92.5%, suggesting it may face challenges in accurately identifying all positive cases. The FNN is the best model for detecting breast cancer overall, outperforming all other models in every metric, highlighting its potential to improve diagnostic precision in clinical settings [41-57]. Based on results its evident that AI technology may improve clinical care, education and training. However, clear regulation and understanding by clinicians are needed. ML is a subfield of AI creating systems that can improve predictions and decisions by exposure to

Table 3: Comparative Analysis for Breast Cancer Prediction on Breast Cancer Wisconsin (Diagnostic) Dataset

Model	Accuracy	Precision	Recall	F1-score
FNN	96.49	97.18	97.18	97.18
RF	94.11	94.97	94.11	94.04
DT	94.3	91.0	92.5	91.8

data, thereby imitating human learning [58][59]. The integration of advanced preprocessing techniques with machine learning significantly enhances the accuracy of mammography analysis, facilitating more precise differentiation between malignant and benign breast lesions [60]. Problems such as model generalization, bias, transparency, interpretability, accountability, and data privacy remain barriers for broad adoption AI in cardiology [61]. Significant potential of AI-based algorithms in enhancing the accuracy of BC survival predictions. However, further exploration and research are essential to fully understand the true impact and effectiveness of these methods. [62].

V. Conclusion and Future Scope

Cancer is a major public health concern since it is a major killer and is on the rise around the world. Breast cancer ranks high among malignancies, particularly among females, according to current studies. Early detection can reduce treatment costs and improve survival rates for people with breast cancer. Nevertheless, the early diagnosis techniques used in modern healthcare systems have disadvantages. This study uses the Breast Cancer Wisconsin (Diagnostic) dataset to assess ML algorithms that can detect breast cancer. The results showed that Feedforward Neural Networks (FNN) were more effective than RF and DT models in detecting breast cancer. With an accuracy of 96.49% and high recall, precision, and F1-scores, the FNN proved robust in minimising false positives and maximising true positives. Nevertheless, the model’s generalisability could be compromised by obstacles including possible class imbalance and the dataset’s small size and lack of diversity. To tackle these issues, future research could use bigger and more varied datasets in conjunction with sophisticated deep-learning methods like CNNs or transformers. Additionally, integrating explainable AI methods could enhance model interpretability, facilitating its adoption in clinical settings for reliable and transparent diagnostic support.

References

- Chetlen A, Mack J, Chan T (2016) "Breast cancer screening controversies: Who, when, why, and how?," *Clin. Imaging.*
- Salama GI, Abdelhalim M, Zeid MA, (2012) "Breast cancer diagnosis on three different datasets using multi-classifiers," *Breast Cancer (WDBC)* 32: 569.
- Hasan MZ, Fink R, Suyambu MR, Baskaran MK (2012) "Assessment and improvement of intelligent controllers for elevator energy efficiency," in *IEEE International Conference on Electro Information Technology.*
- Shahnaz C, Hossain J, Fattah SA, Ghosh S, Khan AI (2017) "Efficient approaches for accuracy improvement of breast cancer classification using wisconsin database," in 2017 IEEE region 10 humanitarian technology conference (R10-HTC) Pp: 792-797.
- Kumar vv, Pandey MK, Tiwari MK, Ben-Arieh D (2010), "Simultaneous optimisation of parts and operations sequences in SSMS: A chaos embedded Taguchi particle swarm optimisation approach," *J. Intell. Manuf.* 21: 335-353.
- Ibrahim Obaid O, Mohammed M, Abd Ghani MK, Mostafa S, Al-Dhief F

- (2018) "Evaluating the Performance of Machine Learning Techniques in the Classification of Wisconsin Breast Cancer," *Int. J. Eng. Technol* 7: 160-166.
- Yarlagadda VK , Pydipalli R (2018) "Secure Programming with SAS: Mitigating Risks and Protecting Data Integrity," *Eng. Int.*, vol. 6: 211-222.
- Karabatak M (2015) "A new classifier for breast cancer detection based on Naïve Bayesian," *Meas. J. Int. Meas. Confed* 2015: 33-36.
- Redig AJ , McAllister SS (2013) "Breast cancer as a systemic disease: a view of metastasis," *J. Intern. Med.*, 274: 113-126.
- Kumar VV, Yadav SR, Liou FW, BalakrishnanSN (2013) "A digital interface for the part designers and the fixture designers for a reconfigurable assembly system," *Math. Probl. Eng.*
- Meesad P , Yen GG(2003) "Combined numerical and linguistic knowledge representation and its application to medical diagnosis," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans* 33: 2.
- Yue W, Wang Z, Chen H, Payne A, Liu X (2018) "Machine learning with applications in breast cancer diagnosis and prognosis," 2.
- Khuriwal N, Mishra N (2018) "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm," in 2018 IEEE International Conference (eTechNXT) 2018: 1-5.
- Gecer B, Aksoy S, Mercan E, Shapiro LG, Weaver DL, et al. (2018) "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks," *Pattern Recognit* 2018: 345-356.
- Sahoo PK, Mohapatra SK, Wu SL (2016) "Analyzing healthcare big data with prediction for future health condition," *IEEE Access* 4: 9786-9799.
- Abdel-Zaher AM ,Eldeib AM (2015) "Breast cancer classification using deep belief networks," *Expert Syst. Appl.*, 2016: 139-144.
- Kandaswamy C, Silva LM, Alexandre LA, Santos JM (2016) "High-Content Analysis of Breast Cancer Using Single-Cell Deep Transfer Learning," *J. Biomol. Screen.*, 21: 3.
- Vennapusa SCR , Fadziso T, Sachani DK, Vamsi Krishna Yarlagadda VK (2018) "Cryptocurrency-Based Loyalty Programs for Enhanced Customer Engagement," *Technol. Manag. Rev* 3: 46-62.
- Kumar VV, Chan FTS (2011) "A superiority search and optimisation algorithm to solve RFID and an environmental factor embedded closed loop logistics model," *Int. J. Prod. Res.*, 49.
- Kumar VV, M. Tripathi M, Pandey MK, Tiwari MK (2009) "Physical programming and conjoint analysis-based redundancy allocation in multistate systems: A Taguchi embedded algorithm selection and control (TAS&C) approach," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab* 223: 215-232.
- Ojha VK, Abraham A, Snášel V (2017) "Metaheuristic design of feedforward neural networks: A review of two decades of research," *Eng. Appl. Artif. Intell.* 60: 97-116.
- Kumar VV, Liou FW, Balakrishnan SN, Kumar V (2015) "Economic impact of RFID implementation in remanufacturing: a Chaos-based Interactive Artificial Bee Colony approach," *J. Intell. Manuf.*, 26: 815-830.
- Mullangi K, Yarlagadda VK, Dhameliya NK (2018) "Integrating AI and Reciprocal Symmetry in Financial Management: A Pathway to Enhanced Decision-Making," *Int. J. Reciprocal Symmetry Theor. Phys* 5: 42-52.
- Kumar V, Kumar VV, Mishra N, Chan FTS, Gnanasekar B (2010) "Warranty failure analysis in service supply Chain a multi-agent framework," in *SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering, 2010.*
- Kumar VV, Chan FTS, Mishra N, Kumar V (2010) "Environmental integrated closed loop logistics model: An artificial bee colony approach," in *SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering, 2010: 1-7.*
- Hasan MZ, Fink R, Suyambu MR, Baskaran MK, James D, Gamboa J (2015) "Performance evaluation of energy efficient intelligent elevator controllers," in *IEEE International Conference on Electro Information Technology* 2015.

27. Zegeye WK, Dean RA, Moazzami F (2018) "Multi-layer hidden Markov model based intrusion detection system," *Mach. Learn. Knowl. Extr* 1: 265-286.
28. Fuad WM (2018) "Early detection of breast cancer using machine learning," *Brac University.*
29. Utomo C, Pratiwi P, Kardiana A, Budi I, Suhartanto H (2014) "Best-Parameterized Sigmoid ELM for Benign and Malignant Breast Cancer Detection," 2014: 50-57.
30. Patra GK, Rajaram SK, Boddapati VN, Kuraku C, Gollangi HK (2022) Advancing Digital Payment Systems: Combining AI, Big Data, and Biometric Authentication for Enhanced Security. *International Journal of Engineering and Computer Science*, 11: 25618-25631.
31. Rajaram SK, Galla EP, Patra GK, Madhavaram CR, Rao J (2022) AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance. *Educational Administration: Theory and Practice* 28: 285-296.
32. Patra GK, Rajaram SK, Boddapati VN (2019) Ai And Big Data In Digital Payments: A Comprehensive Model For Secure Biometric Authentication. *Educational Administration: Theory and Practice*, 25: 773-781.
33. Kuraku C, Gollangi HR, Sunkara JR (2020) Biometric Authentication In Digital Payments: Utilizing AI And Big Data For Real-Time Security And Efficiency. *Educational Administration: Theory and Practice*, 26: 954-964.
34. Galla EP, Madhavaram CR, Boddapati VN (2021) Big Data And AI Innovations In Biometric Authentication For Secure Digital Transactions *Educational Administration: Theory and Practice* 27: 1228-1236.
35. Sunkara JR, Bauskar SR, Madhavaram CR, Galla EP, Gollangi HR (2021) Data-Driven Management: The Impact of Visualization Tools on Business Performance, *International Journal of Management (IJM)* 12: 1290-1298.
36. Boddapati VN, Sarisa M, Reddy MS, Rajaram SK, Bauskar SR, et al.(2022) "Data migration in the cloud database: A review of vendor solutions and challenges," *Int. J. Comput. Artif. Intell.*, 3: 96-101.
37. Reddy MR, Sarisa M, Konkimalla S, Bauskar SR, Gollangi HK, et al. (2021) "Predicting tomorrow's Ailments: How AI/ML Is Transforming Disease Forecasting", *ESP Journal of Engineering & Technology Advancements* 1: 188-200.
38. Gollangi K, Bauskar SR, Madhavaram CR, Galla P, Sunkara JR, Reddy MS (2020) "ECHOES IN PIXELS : THE INTERSECTION OF IMAGE PROCESSING AND SOUND OPEN ACCESS ECHOES IN PIXELS : THE INTERSECTION OF IMAGE PROCESSING AND SOUND DETECTION," *Int. J. Dev. Res* 10: 39735-39743.
39. Gollangi HK, Bauskar SR, Madhavaram CR, Galla EP, Sunkara JR, Reddy MS (2020) Unveiling the Hidden Patterns: AI-Driven Innovations in Image Processing and Acoustic Signal Detection. (2020). *JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE)* 8: 25-45.
40. Gollangi HK, Bauskar SR, Madhavaram CR, Galla EP, Sunkara JR, Reddy MS (2020) Exploring AI Algorithms for Cancer Classification and Prediction Using Electronic Health Records. *Journal of Artificial Intelligence and Big Data* 1: 65-74.
41. Bauskar S, Boddapati VN, Sarisa M, Reddy M, Surender S, et al. (2022) Data Migration in the Cloud Database: A Review of Vendor Solutions and Challenges.
42. Chandrakanth RM, Eswar PG, Mohit SR, Manikanth S, Venkata NB, et al. (2021) Predicting Diabetes Mellitus in Healthcare: A Comparative Analysis of Machine Learning Algorithms on Big Dataset. In *Global Journal of Research in Engineering Computer Sciences* 1: 1-11.
43. Boddapati VN, Galla EP, Patra GR, Madhavaram CR, Sunkara JR (2023) AI-Powered Insights: Leveraging Machine Learning And Big Data For Advanced Genomic Research In Healthcare. *Educational Administration: Theory and Practice* 29: 2849-2857.
44. Patra GK, Kuraku C, Konkimalla S, Boddapati VN, Sarisa M (2023) Voice classification in AI: Harnessing machine learning for enhanced speech recognition. *Global Research and Development Journals* 8: 19-26.
45. Sunkara JR, Bauskar SR, Madhavaram CR, Galla EP, Gollangi HR (2023) Optimizing Cloud Computing Performance with Advanced DBMS Techniques: A Comparative Study. *Journal for ReAttach Therapy and Developmental Diversities* 6: 2493-2502.
46. Sunkara JR, Bauskar SR, Madhavaram CR, Galla EP, Gollangi HK (2023) An Evaluation of Medical Image Analysis Using Image Segmentation and Deep Learning Techniques.
47. Patra GK, Kuraku C, Konkimalla S, Boddapati VN, Sarisa M, et al. (2023) Sentiment Analysis of Customer Product Review Based on Machine Learning Techniques in E-Commerce. *Journal of Artificial Intelligence Cloud Computing* 2: 1-4.
48. Siddharth K, Kumar GP, Chandrababu K, Janardhana Rao S, Sanjay Ramdas B, et al. (2023) A Comparative Analysis of Network Intrusion Detection Using Different Machine Learning Techniques. *J Contemp Edu Theo Artific Intel: JCETAI-102.*
49. Eswar Prasad G, Hemanth Kumar G, Venkata Nagesh B, Manikanth S, Kiran P, et al. (2023) Enhancing Performance of Financial Fraud Detection Through Machine Learning Model. *J Contemp Edu Theo Artific Intel: JCETAI-101.*
50. Rajaram SK, Konkimalla S, Sarisa M, Gollangi HK, Madhavaram CR, Reddy, et al. (2023) AI/ML-Powered Phishing Detection: Building an Impenetrable Email Security System. *ISAR Journal of Science and Technology* 1: 10-19.
51. Eswar Prasad G, Hemanth Kumar G, Venkata Nagesh B, Manikanth S, Kiran P, et al. (2023) Prediction of Financial Stock Market Based on Machine Learning Technique. *J Contemp Edu Theo Artific Intel: JCETAI-102*, Available at SSRN.
52. Kalla D, Samiuddin V (2020) Chatbot for medical treatment using NLTK Lib. *IOSR J. Comput. Eng.*, 22: 50-56.
53. Kalla D, Chandrasekaran A(2023)Heart disease prediction using machine learning and deep learning, *International Journal of Data Mining and Knowledge Management Process* 13: 1-14.
54. Chandrasekaran A, Kalla D (2023) Heart disease prediction using chi-square test and linear regression. *Computer Science & Information Technology* 13: 135-146.
55. Kalla D, Smith N, Samaah F, Polimetla K (2022) Enhancing Early Diagnosis: Machine Learning Applications in Diabetes Prediction. *Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-205191:* 2-7.
56. Kuraku DS, Kalla D (2023) Phishing Website URL's Detection Using NLP and Machine Learning Techniques. *Journal on Artificial Intelligence-Tech Science.*
57. Sarma D, Rali AS, Jentzer JC (2025) Key Concepts in Machine Learning and Clinical Applications in the Cardiac Intensive Care Unit. *Curr Cardiol Rep* 27: 30.
58. Jaltotage B, Dwivedi G (2024) Essentials for AI Research in Cardiology: Challenges and Mitigations. *CJC Open* 6: 1334-1341.
59. Mehrabi M, Salek N (2024) Enhancing diagnostic accuracy in breast cancer: integrating novel machine learning approaches with enhanced image preprocessing for improved mammography analysis. *Pol J Radiol* 89: e573-e583.
60. Bota P, Thambiraj G, Bollepalli SC, Arroundas AA (2024) Artificial Intelligence Algorithms in Cardiovascular Medicine: An Attainable Promise to Improve Patient Outcomes or an Inaccessible Investment? *Curr Cardiol Rep* 26: 1477-1485.
61. Javanmard Z, Zarean Shahraki S, Safari K, Omid A, Raoufi S, et al. (2025) Artificial intelligence in breast cancer survival prediction: a comprehensive systematic review and meta-analysis. *Front Oncol* 14: 1420328.