Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Classification and Prediction of Breast Cancer using Bagging and ANN

[*]Sampoornamma sudarsa [1], and R.Pradeep Kumar Reddy [2]

[1] *Department of Computer Science and Engineering, Y.S.R.Engineering college of Yogi vemana University, Proddatur , India.*

Abstract. Artificial Intelligence is the emerging field to make many works almost in all diciplane automation. The machine learning and deep learning are the subset of artificial intelligence and plays a crucial role in making applications automation. Deep learning is a subset of machine learning and showing the effective results and will work as the human neural network works. In the healthcare field, particularly in the classification and prediction of breast cancer, deep learning algorithms such as bagging and artificial neural networks have been employed. These algorithms have been evaluated based on metrics like precision, recall, F1 score, and support. The results indicate that artificial neural networks outperform the bagging algorithm in terms of accuracy, showcasing their effectiveness in healthcare applications.The current work is focused on the healthcare field for classification and the prediction of the breast cancer. For this the deep learning algorithms the bagging and artificial neural networks were used and generated results in terms of the precision, recall, F1 score and support. İn terms of accuracy the artificaial neural network has shown effecive results when campared with the bagging algorithm.

Keywords: *Bagging, Artificial Neural Network, Classification, prediction, Breast cancer, and Deep Learning.*

1 Introduction

Breast cancer is a common cause of cancer-related deaths in women. Detecting it early is crucial for better treatment and survival. Tumors can be non-cancerous (benign) or cancerous (malignant). About 12% of US women may get breast cancer, and it's the most common cancer in women. Detecting it involves methods like mammography and biopsies. Researchers use machine learning to improve detection accuracy. This study proposes using deep learning with pre-trained models to identify affected areas in breast images, enhancing classification accuracy, and reducing training time.[1] Detecting cancer that has spread (metastasized) is crucial. Using CNN-based methods, such as machine learning, can help, but detecting metastases in large pathology images is tough due to image size and lack of labelled data. This study focusses on breast cancer metastases in lymph node images. Pathologists used to diagnose under microscopes, but this is slow and prone to errors. Machine learning offers faster, accurate solutions, but needs of labeled data. This study proposes a method that uses cell counting as a supporting task to improve cancer detection with limited labelled data.[2]

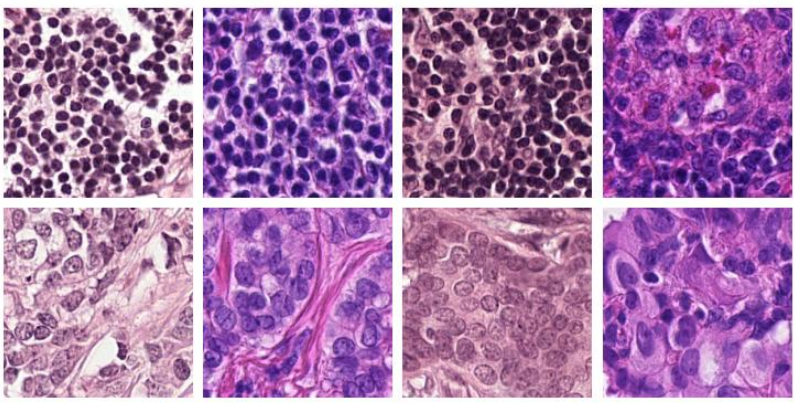

Fig.1 Some normal (top) and tumor (bottom) patches cropped from the whole-slide pathological tissue in the biggest tile.

Detecting and classifying breast cancer is crucial but complex. Deep learning shown promise in medical imaging. Breast cancer affects many women and is a leading cause of death. Early detection is key, and methods like mammography and biopsies are used. Manual analysis by specialists is slow and can have errors. Lack od specialists and the complexity of histopathological images can lead to missed diagnoses. Automated methods are needed for accurate and efficient breast cancer detection.[3] Breast cancer is a serious type of cancer that starts in the breast cells. Even though there have been improvements in treating the main form of this cancer in recent years, we really need a better way to predict its outcome. Breast cancer happens when cells in the breast grow out of control, forming a mass called a tumor. There are two types of tumors: malignant (harmful) and benign (not harmful). Malignant tumors can spread and some that stay in one place. This type of cancer is a significant cause of death, particularly among women.[4], those normal and tumor patches shown in Fig 1.

Scientists have come up with a new way to help doctors diagnose breast cancer using pictures of tissue samples. They made a special computer program that has two parts: one to find odd parts in the images, and another to figure out if the tissue is cancerous. The first part helps the second part work better by finding mistakes in the picture. This program is smarter and quicker than what doctors usually do. Breast cancer is a big problem for women.[5]

Breast cancer is a common cause of death among women worldwide. Detecting it early is crucial for treatment and survival. Doctors often use microscope images to diagnose it. A special type of computer program, based on advanced techniques, is being used to analyze these images. It's a challenge because the images are complex and have small details. This study aims to make the program better by combining different approaches, improving the images, and using smart techniques. The goal is to help doctors make quicker and more accurate diagnoses.[6]

Many newsletters have highlighted the serious problem of invasive cancerous tumors affecting women's lives since 2010. These tumors spread and create new tumors, causing a lot of harm. CT scans are used to detect cancer early, but sometimes they give wrong results, especially for women with large breast fat. This leads to unnecessary worry, especially in poorer countries where people can't afford more tests. About 10-15% of women above 30 are affected by invasive breast cancer, and researchers are using advanced AI and ML technology to find better ways to detect it. Some places have high mortality rates due to lack of medical care and income, while others focus on regular mammography screenings to reduce costs.[7]

In computer vision, there are properties called rotation equivariance and translation invariance that help recognize transformed images accurately. These properties are achieved in Convolutional Neural Networks (CNNs) through a process called data augmentation. CNNs are widely used in tasks like medical image analysis, including detecting diseases like Covid-19 and breast cancer. The goal of this research is to improve breast cancer detection in mammograms and make CNN models more efficient in processing different images.[8]

Artificial intelligence (AI) technology has developed significantly, and breast cancer diagnosis is one area where AI is being applied. Most research has been done in centralized learning, which can risk privacy breaches. This study aimed to improve breast cancer classification by using FL and involved multiple hospitals and medical centre's collaborating without sharing patient data. The goal was to enhance recall performance for diagnosing breast cancer, focusing on real-world effectiveness and overcoming challenges like data heterogeneity, data processing, privacy, and efficiency in distributed learning algorithms.[9]

Breast cancer mortality among women has risen due to late detection. Advanced Artificial Intelligence (AI) can improve medical detection, but data privacy is a challenge. Federated learning (FL) is gaining traction for secure data analysis. This study uses FL to secure breast cancer data and proposes a new deep learning model for accurate disease classification.[10] Discussion about the various regression algorithms has been discussed [11] and its uses were given clearly, Various Machine learning algorithms and internet of things technologies [12-15] were used for various applications such as live bidding, smart home automation, and crop dieases predictions.

Based on the survey, the current work is to apply the machine learning and deep learning algorithms for classification and prediction of the breast cancer. Major seach research has been done as per the survey related to the same, here particularly to do the same work the enseamblng ans the parameter hasbeen done and the bagging

and artificial neural network algorithms were used and the clasification report as been generated and comparitive analysis of the same has been done.

The remaining paper is organized as following, in the section II discussed about the related work, in section III and IV about methodology and implementation, in V discussed about the results and discussion and finally in section VI discussed about the conclusion and future work.

2 Related Work

In this [1], the authors study, a new deep learning model using transfer learning is developed to help detect and diagnose suspected breast cancer areas. Different techniques are used to assess the model's performance. The model uses pre-trained neural network architecture like Inception V3, ResNet50, VGG-19, VGG-16, and Inception-V2 ResNet to analyze mammogram images. The results show that the VGG16 model is effective, achieving high accuracy, sensitivity, specificity, and other measures for breast cancer diagnosis. This approach could be applied to other medical diagnoses in the future.

In this [2], the authors focuses on detecting breast cancer spread in pathology images at a patch level. To overcome challenges, a few-shot learning method is introduced. It involves ranking cell counts within patches and uses limited labelled data. By leveraging unlabelled images, the method enhances accuracy compared to regular supervised approaches. This approach could help tasks with hard-to-get labelled data.

In this [3], the authors introduces a new method called Pa-DBN-BC for detecting and classifying breast cancer in histopathology images. It uses a Deep Belief network (DBN) to extract features from images patches. The model is trained with images from different datasets and achieves an 86% accuracy. The proposed method automatically learns important features, outperforming traditional approaches and improving cancer classification accuracy.

In this [4], the authors working together to find ways to save people from breast cancer. They want to create a helpful model that can predict how this disease will affect people. This prediction can be really useful because it can find cancer early, avoid unnecessary treatment, and save money. They are using advanced methods involving deep learning and combining different types of data, like genes and medical information.

In this [5], the authors proposed a new approach for diagnosing breast cancer was tested on two sets of pictures: BreaKHis and BACH 2018. They checked if the program could correctly say if the tissue was cancerous or not. At times when the picture was bigger, the program was really accurate, like 98.83% at 200x magnification. It also worked well on another set of pictures, BACH 2018, where it got 99.25% for certain cases. This new method is faster and better than some other methods used before, which is really helpful for doctors making diagnoses.

In this [6], the authors created a special computer program to help identify breast cancer using microscope images. The program combines different techniques and models to extract important information from the images. It's designed to overcome challenges like disappearing small details in the images. We trained and tested the program on a dataset of cancer images and achieved high accuracy.

In this [7], the authors created to better identify breast cancer in mammogram images and reduce false alarms. This method uses a special computer program called WOMT that has been trained with certain rules to make better predictions. It's important because invasive tumors can spread and cause serious problems. CT scans are often used to find cancer, but they can give wrong results, especially for women with large breast fat. This new method helps reduce mistakes.

In this [8], the authors proposed a new deep neural network architecture, combining a group convolutional neural network (G-CNN) using a special Euclidean motion group (SE2) and a discreate cosine transform (DCT). The G-CNN with SE2 ensures consistent handling of 2D image rotations, while the DCT helps encode the model space. The model is tested on rotated MNIST datasets, showing promise, and applied to mammography images, achieving 94.84% accuracy.

In this [9], the authors introduces new ideas: 1. Using transfer learning to extract important data features from specific image areas (ROI) for better data training. 2. Applying synthetic minority oversampling technique (SMOTE) to process data and improve disease prediction. 3. Employing FeAvg-CNN + MobileNet in an FL

framework for privacy and security in healthcare. The results show that out solution outperforms other approaches in classifying mammography images for breast cancer detection.

In this [10], the authors aims to create an automatic diagnosis system using federated and deep learning to speed up the process. It involves steps like image collection, encryption, optimal key generation, secure storage, and disease classification. The proposed method achieves high performance in accuracy, precision, recall, and other metrics, for secure breast cancer detection.

Based on the survey the various machine learning algorithms were used for detection, prediction and classification of human and plants dieases. The current work has used the brest cancer datasets, the dataset papmeters were fine tuned, and the ensemble techniques were used. The artificial neural network and bagging algorithms has been used on the fine tuned dataset and te classification report has been generated. Based on that the inference hasbeen drawn.

3 Methodology

The methodology of the proposed system, along with its working principles, is shown in Fig 2, there are a few key steps. The architecture of the model and its working principles are shown in Fig. 2.

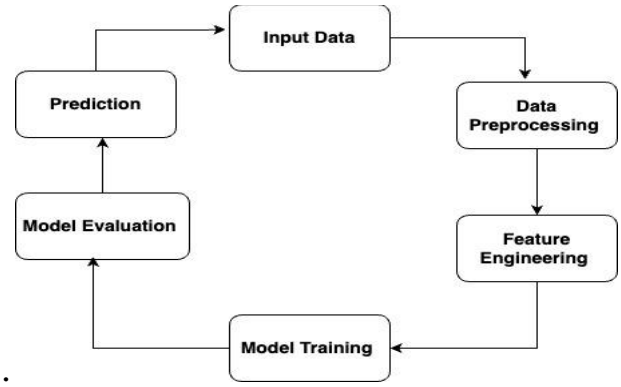

Fig 2: Working Model of this project.

Step 1: Input Data

Here Breast Cancer Dataset taken from UCI repository, and it has 11 columns and 686 rows.

Step 2: Data Preprocessing

clean and prepare the data by identifying the outliers and missing data. In our dataset there is no missing and identified 43 outliers and removed using few outlier removal techniques like IQR,Z-SCORE.

Step 3: Feature Engineering

The main features were identified and selected from the dataset and designated as the independent variable (X), while the dependent variable (Y) represented the status. I used a min-max scaler variable to increase the consistency of the variables.

Step 4: Model Training

Machine learning algorithms such as Artificial Neural Network and Bagging Classifier are used here and this data is used to train and evaluate the model. Use 80% of the data for training and 20% of the data for testing.

Step 5: Model Evaluation

You can evaluate the performance of the model in terms of accuracy, precision, recall, f1 score, etc. We will use the Scikit-learn library to check with measurements. Precision measures the accuracy of prediction quality and recall, measures the accuracy of quality and F1 scores, measures precision and recall.

Step 6: Prediction

Based on the above metrics we will finalize the model performance and result of the train and test model.

4 Implementation

Here breast cancer dataset taken from UCI repository.It contains 686 rows and 12 columns and this dataset has essential features like patient ID (pid), age, menopausal status (meno), tumor size, grade, regional lymph node status (node), progesterone receptor status, hormonal status, recurrence-free survival time(rfstime), and overall patient status. These attributes show details about patient information, tumor characteristics and survival time. Tumor size and grade help understand the stage and aggressiveness of the cancer, while the location of the tumor (nodes) indicates how far the tumor has spread. Molecular concepts such as estrogen and progesterone receptors help understand hormonal factors. rfstime tell us about the patient's last date. status help us to identify breast cancer.

4.1 Trainig and Setup

The training and testing setup for the proposed system utilized a Jupyter Notebook, It is a software which is used for understanding the dataset and evaluating model performance. Specifically, the accuracy between the Artificial Neural Network and Bagging Classifier . 80% data taken for training and 20% data taken for testing.

5 Results and Discussion

Artificial Neural Networks performing precision, measuring the accuracy of positive predictions, were around 96% for Class 0 and 95% for Class 1. Recall, measuring the accuracy of positive instances, was around 96% for Class 0 and 95% for Class 1. The F1-score, balancing precision and recall, was around 96% for Class 0 and 95% for Class 1. Support values indicate 179 instances for Class 0 and 142 instances for Class 1.These results are shown in Table 1. Using a horizontal bar graph explaining the Artificial Neural Network performance metrics of two classes. The graph shows Precision, recall, F1-score,Support with different colors of each metrics. Class is mentioned in y axis and Percentage and count is mentioned in x-axis. In class 0 Precision is mentioned as 96% and Recall is mentioned as 96% and F1-score is mentioned as 96% and support as 179.In class 1 Precision is mentioned 95% and Recall is mentioned as 95% and F1-score is mentioned as 95% and support as 142.These results are shown in Fig 2. Bagging Classifiers performing precision, measuring the accuracy of positive predictions, were around 81% for Class 0 and 79% for Class 1. Recall, measuring the accuracy of positive instances, was around 87% for Class 0 and 71% for Class 1. The F1-score, balancing precision, and recall, was around 84% for Class 0 and 75% for Class 1. Support values indicate 187 instances for Class 0 and 135 instances for Class 1. These results are shown in Table 2.

Table 1. Classification report for Artificial Neural Network

Prediction	Precision	Recall	F1 Score	Support
0	96%	96%	96%	179
1	95%	95%	95%	142

Table 2. Classification report for Gradient Bagging Classifier

Prediction	Precision	Recall	F1 Score	Support
0	81%	87%	84%	187
1	79%	71%	75%	135

Using a horizontal bar graph explaining the Bagging classifier performance metrics of two classes. The graph shows Precision, recall, F1-score,Support with different colors of each metrics. Class is mentioned in y axis and Percentage and count is mentioned in x-axis. In class 0 Precision is mentioned as 81% and Recall is mentioned as 87% and F1-score is mentioned as 84% and support as 187..In class 1 Precision is mentioned 79% and Recall is mentioned as 71% and F1-score is mentioned as 75% and support as 135.These results are shown in Fig 3.

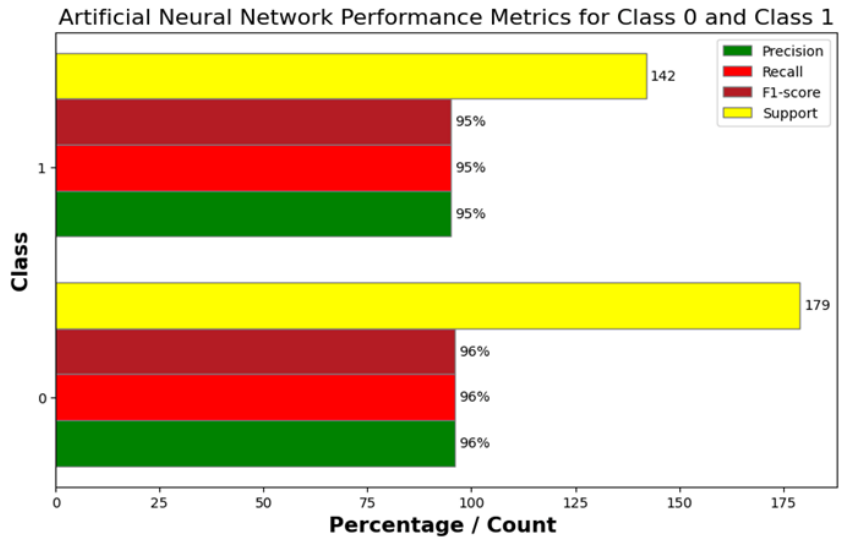

Fig 3. Artificial Neural Network Performance Metrics for class 0 and class1.

Fig 4. Bagging Classifier Performance Metrics for class 0 and class1.

After implementing the proposed methodology and obtaining results from both algorithms, the Artificial Neural Network got a prediction accuracy of 96%, which appears to be highest compared to Bagging Classifier,shown in Fig 4, which achieved an accuracy of 80%. These results are shown in Table 3.

Table 3. Comparison of Artificial Neural Network vs Bagging Classifier

S.No	**Accuracy**	
	Artificial Neural Network	Bagging Classifier
1	96%	80%

The accuracy of two machine learning algorithms such as Artificial Neural Network and Bagging Classifier where Artificial Neural Network is represented in orange and Bagging Classifier in green. Accuracy value is represented in Y axis and Algorithm is mentioned in X axis. Artificial Neural Network got 96% accuracy whereas Bagging Classifier got a Accuracy as 80 %. Artificial Neural Network got the highest accuracy compared to Bagging Classifier as shown in fig 5.

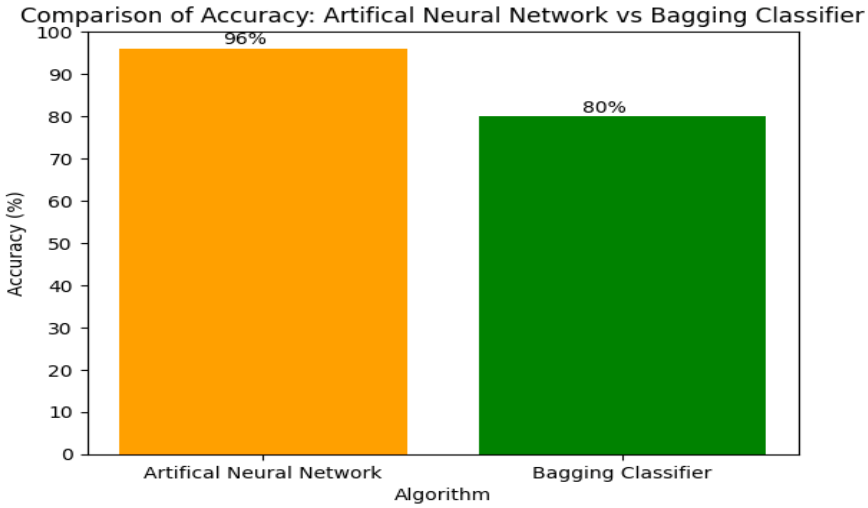

Fig 5. Comparison between Artificial Neural Network vs Bagging Classifier.

6 Conclusion and Future Work

In our dataset, after applying both the algorithms Artificial Neural Network performs well compared to Bagging Classifier based on Accuracy, Precision, and recall. So Artificial Neural Network is suitable for this dataset. Hence Artificial Neural Network is the more efficient algorithm for identifying the type of Breast Cancer. In future work different machine learning and deep learning algorithms will be applied on the same dataset.

7 References

[1] Saber, Abeer, et al. "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique." IEEE Access 9 (2021): 71194-71209.

[2] Chen, Jiaojiao, et al. "Few-shot breast cancer metastases classification via unsupervised cell ranking." IEEE/ACM transactions on computational biology and bioinformatics 18.5 (2019): 1914-1923.

[3] Hirra, Irum, et al. "Breast cancer classification from histopathological images using patch-based deep learning modeling." IEEE Access 9 (2021): 24273-24287.

[4] Arya, Nikhilanand, and Sriparna Saha. "Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model." IEEE/ACM transactions on computational biology and bioinformatics 19.2 (2020): 1032-1041.

[5] Zhou, Yiping, Can Zhang, and Shaoshuai Gao. "Breast cancer classification from histopathological images using resolution adaptive network." IEEE Access 10 (2022): 35977-35991.

[6] Khan, Hameed Ullah, et al. "MSF-Model: Multi-Scale Feature Fusion-Based Domain Adaptive Model for Breast Cancer Classification of Histopathology Images." IEEE Access 10 (2022): 122530-122547.

[7] Lakshmi, L., et al. "WOMT: Wasserstein Distribution based minimization of False Positives in Breast Tumor classification using Deep Learning." IEEE Access (2023).

[8] Sani, Zaharaddeen, Rajesh Prasad, and Ezzeddin Kamil Mohamed Hashim. "Breast Cancer Classification Using Equivariance Transition in Group Convolutional Neural Networks." IEEE Access 11 (2023): 28454-28465.

[9] Tan, Y. Nguyen, et al. "A Transfer Learning Approach to Breast Cancer Classification in a Federated Learning Framework." IEEE Access 11 (2023): 27462-27476.

[10] Peta, Jyothi, and Srinivas Koppu. "Breast Cancer Classification In Histopathological Images Using Federated Learning Framework." IEEE Access (2023).

[11] Saravanakumar, S., & Thangaraj, P. (2019). A computer aided diagnosis system for identifying Alzheimer's from MRI scan using improved Adaboost. Journal of medical systems, 43(3), 76.

[12] Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. Cluster Computing, 22(Suppl 1), 33-46.

[13] Saravanakumar, S., & Saravanan, T. (2023). Secure personal authentication in fog devices via multimodal rank-level fusion. Concurrency and Computation: Practice and Experience, 35(10), e7673.

[14] Thangavel, S., & Selvaraj, S. (2023). Machine Learning Model and Cuckoo Search in a modular system to identify Alzheimer's disease from MRI scan images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(5), 1753-1761.

[15] Saravanakumar, S. (2020). Certain analysis of authentic user behavioral and opinion pattern mining using classification techniques. Solid State Technology, 63(6), 9220-9234.

[16] Baseer, K. K., et al. "Analysing various regression models for data processing." International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8.8 (2019): 731-736.

[17] Neerugatti, Vikram, and A. Rama Mohan Reddy. "Secured Architecture for Internet of Things-Enabled Personalized." Internet of Things and Personalized Healthcare Systems (2018).

[18] Maji, Supriti, et al. "Cotton Crop Certainty Identification Using Deep Learning Techniques." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.

[19] Mudlapur, Chetan, et al. "Live Bidding Application: Predicting Shill Bidding Using Machine Learning." International Conference on Multi-disciplinary Trends in Artificial Intelligence. Cham: Springer Nature Switzerland, 2023.

[20] Vasu, Sreenivasulu, Vikram Neerugatti, and C. Naga Swaroopa. "A Machine Learning Based Decision Support System for Improvement of Smart Watering Equipment in Agricultural Fields."