A | R | J
Amrit Research Journal December 2025, Vol. 6, Issue 1, 91-97

ISSN: 2738-9553
RESEARCH ARTICLE

Comparative Evaluation of AI models for Early Detection of Breast Cancer

***Chetna Sharma*[1]*, Binod Kumar Adhikari*[2*]*, Vijay Kumar Jha*[3]**

[1]*St. Xavier's College, Kathmandu, Nepal*

[2]*Central Department of Computer Science and Information Technology, Tribhuvan University, Kathmandu, Nepal*

[3]*Department of Physics, Amrit Campus, Tribhuvan University, Kathmandu, Nepal*

**Email: bkadhikari@cdcsit.tu.edu.np*

(Received: November 3, 2025, Received in revised form: November 29, 2025, Accepted: December 9, 2025, Available online: December 19, 2025)

DOI: https://doi.org/10.3126/arj.v6i1.87536

Highlights

- Artificial Intelligence has revolutionized the detection of cancers
- Six models are trained, tested, and compared
- The result highlights Logistic Regression for the early detection of breast cancer

Abstract

Most of the deaths caused by breast cancer are due to lack of early detection of tumors. Mammography, biopsies, and ultrasounds are commonly used for the detection of breast cancer, which have major limitations like high false positives, subjective interpretation, and the variability in radiologists' skills. However, advancement in the Artificial Intelligence (AI) has significantly revolutionized the detection of cancers using image classification and analysis of historical mammography and MRI data. In this paper, six different models were evaluated on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The performance of traditional machine learning models, particularly logistic regression (LR), random forest (RF), support vector machine (SVM), multi-layer perceptron (MLP), and deep learning approaches, particularly 1D convolutional neural networks (CNN) and deep neural networks, was compared using multiple evaluation metrics, including accuracy, precision, recall, and F1-score. The result highlights Logistic Regression as the most effective model for early detection of breast cancer, offering a reliable prediction on whether a tumor is malignant or benign. The research contributes invaluable insights into the performance of AI models and the implementation of such models for breast cancer detection.

Keywords: *breast cancer, diagnosis, machine learning, artificial intelligence, deep learning*

Introduction

One of the most life-threatening diseases in the world is cancer. Among different types of cancer, Breast cancer is one of the leading causes of death in women. WHO's World Cancer report shows that Breast cancer is the most commonly diagnosed cancer in women (2.1 million new cases in 2018) and the leading cause of cancer death in women globally (627,000 deaths in 2018). Most of the deaths are caused by a diagnosis in the advanced stage of cancer. This is why early detection is very crucial. Research shows that if breast cancer is detected early, it can increase the survival rate by more than 70%. Breast cancer is influenced by

[*] *Corresponding author*

several factors, including breast density, women's age, type of lesion, medical history, breastfeeding, and others. Some have a high influence, whereas others have a low influence. The most common approaches to identifying breast cancer are mammography, biopsies, and physical examinations. Among these, mammograms are most widely used, which have an accuracy rate of around 70-80%, according to research. However, mammography highly depends on breast density, which differs in different age groups. This becomes the reason for high false positives, causing some patients to go through unnecessary biopsies, which are expensive as well as invasive. Physical examinations are conducted by radiologists, and the reports of the examinations are interpreted differently by different radiologists. This inconsistency results in a lower probability of early detection of breast cancers. This is why there is a need for an automated, accurate, and efficient approach to detecting breast cancer.

Artificial Intelligence is advancing medical sector. Deep Learning and Machine learning approaches are used to classify images, analyze historical medical data, and make predictions. In the field of early detection of breast cancer, AI is used to effectively detect tumors early, accurately, and efficiently. In this paper, traditional machine learning models, particularly logistic regression (LR), random forest (RF), support vector machine (SVM), Multi-Layer Perceptron (MLP), and deep learning models, particularly 1D convolutional neural networks (CNN) and deep neural networks, were evaluated on the WDBC datasets. The dataset taken had 569 samples (212 malignant, 357 Benign) with 30 features, including radius, texture, perimeter, area, and others. Among the six models, Logistic Regression turned out to be the most accurate and precise model to classify malignant and benign breast cancer.

This approach for early detection of cancer doesn't need the suggestion of any expert, reduces the space for different interpretations, and doesn't depend on factors like breast density or women's age. It is purely the data collection, pattern recognition, training, analysis, and prediction. This is the very reason AI has significant potential in contributing to the early detection of tumors in breast cancer and enhancing doctors' confidence in identifying malignant tumors. This is why AI has been rapidly integrated into the field of the health sector to improve accuracy and early detection of breast cancer. (Alashban et al., 2025; Alom et al., 2025; Arravalli et al., 2025; Chen et al., 2025; Chen et al., 2025; Choe et al., 2023; Das et al., 2023; Elkorany & Elsharkawy, 2023; Guo et al., 2024; Liu et al., 2025; Ramamoorthy et al., 2024; Rafiq et al., 2025) results show that deep learning algorithms perform well in feature extraction from medical images and provide high accuracy for image based breast cancer detection. (Alom, et al., 2025; Gurcan, F., 2025; Raja, et al., 2025) results show that hybrid and ensemble methods improved prediction robustness for complex breast cancer detection tasks. (Kallah-Dagadu et al., 2025; Latha, 2024) concluded about the gap between high-performing deep learning models and clinical trust in breast cancer detection. Authors of (Alzahrani,2025; Chang,2025; Eisemann, 2025; Frazer, 2024; Islam, 2024; Liu,2025) concluded that population level AI systems are critical for large scale screening for early detection of breast cancer in woman.

In most of the studies, the authors have compared and evaluated deep learning models in image-based detection of breast cancer and they have not done comparative studies combining classical ML and simple deep learning models on structured tabular datasets. Moreover, 1D CNN architecture applied to tabular features are also underexplored. To address this research gap, the present study makes a comparison and evaluation of six AI models, including classical ML and deep learning architecture. This paper has also explored the potential of 1-D CNN to detect breast cancer in woman

Methodology

This research aims to evaluate the role of Artificial Intelligence in the early detection of breast cancer. The dataset used for this study was obtained from the Wisconsin Diagnostic Breast Cancer (WDBC) repository, consisting of 569 samples in total, with 212 malignant and 357 benign cases. To conduct the training, essential libraries such as NumPy, Pandas, Matplotlib, Seaborn, and the required machine learning modules and evaluation metrics were imported. The dataset was loaded from sklearn.datasets and organized into a Pandas DataFrame for efficient inspection and preprocessing. The data was then split into training and testing sets with an 80:20 ratio. Before model training, standardization was applied so that each feature achieved a mean of 0 and a standard deviation of 1, ensuring equal contribution of all variables during the learning process. To assess performance, each model was evaluated using five key metrics: Accuracy, Precision, Recall, F1-score, and AUC-ROC. A confusion matrix was also generated to visualize the number of true positives, true negatives, false positives, and false negatives for each model.

Evaluation Metrics

True Positives (TP)= Both observations' class and model suggest that the tumor is benign.

True Negatives(TN)=Both observations' class and model suggest that the tumor is malignant.

False positive (FP)=The observations' class is malignant, but the model predicts it as benign.

False Negatives (FN) =The observations' class is benign, but the model predicts it as malignant.

I. Accuracy = (TP + TN) / (TP + TN + FP + FN)

II. Precision = TP / (TP + FP)

III. Recall = TP / (TP + FN)

IV. F1-score = 2 × (Precision × Recall) / (Precision + Recall)

V. AUC-ROC = Area under the ROC curve

The following sections provide a detailed description of the six machine learning models applied in this study.

Logistic Regression:

Logistic Regression is a type of supervised machine learning classification algorithm used to predict the probability of a target variable. The nature of the target is dichotomous, which means there would be only two possible cases. We used Logistic Regression to predict whether a tumor is malignant or benign. We trained our model on a scaled training dataset and evaluated its performance using evaluation metrics as shown in Fig. 1.

```
1. LOGISTIC REGRESSION
==================================================
Accuracy: 0.9825
Precision: 0.9861
Recall: 0.9861
F1-Score: 0.9861
AUC-ROC: 0.9954

Confusion Matrix:
[[41  1]
 [ 1 71]]
```

Fig. 1. Logistic Regression

Random Forest

Random Forest works under an ensemble machine learning method that merge many decision trees to generate a model by reducing the overfitting tendencies of decision trees. We trained the random classifier on scaled data to predict whether a tumor is malignant or benign. The model was evaluated using evaluation metrics (Fig. 2). Lastly, we analyzed feature importance, which helps in understanding which features are most important for the detection of breast cancer as shown in Fig. 3.

```
2. RANDOM FOREST
==================================================
Accuracy: 0.9561
Precision: 0.9589
Recall: 0.9722
F1-Score: 0.9655
AUC-ROC: 0.9939

Confusion Matrix:
[[39  3]
 [ 2 70]]
```

Fig. 2. Random Forest

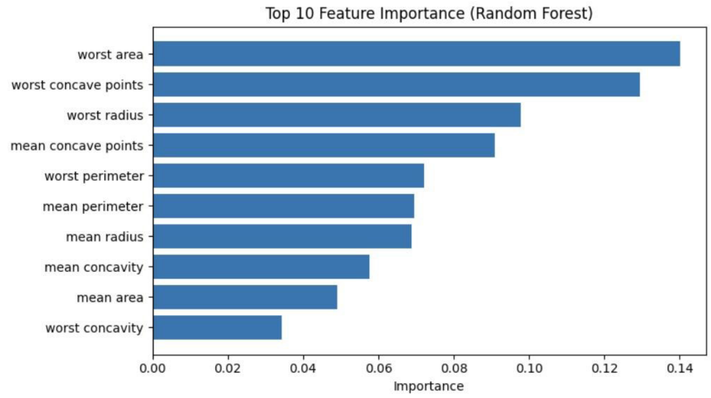

Fig. 3. Top 10 Feature Importance (Random Forest)

Support Vector Machine:

Support Vector Machines work by finding the best boundary(also known as a hyperplane) that best separates data points of different classes. We used an SVM with a radial basis function (RBF) kernel, which allows the algorithm to handle non-linear relationships between features and the target variable. We evaluated the model using evaluation metrics as shown in Fig. 4.

```
3. SUPPORT VECTOR MACHINE
==================================================
Accuracy: 0.9825
Precision: 0.9861
Recall: 0.9861
F1-Score: 0.9861
AUC-ROC: 0.9950

Confusion Matrix:
[[41  1]
 [ 1 71]]
```

Fig. 4. Support Vector Machine

Neural Network:

The multilayer perceptron is an artificial neural network used for classification tasks. It consists of an input, a hidden, and an output layer. The MLP model had two hidden layers with 100 and 50 neurons, respectively. The model was trained on scaled data, and performance was evaluated on evaluation metrics. A confusion matrix was also generated to show the distribution of true positives, true negatives, false positives, and false negatives as shown in Fig. 5.

```
4. MULTI-LAYER PERCEPTRON (NEURAL NETWORK)
==================================================
Accuracy: 0.9474
Precision: 0.9853
Recall: 0.9306
F1-Score: 0.9571
AUC-ROC: 0.9944

Confusion Matrix:
[[41  1]
 [ 5 67]]
```

Fig. 5. Neural Network

Deep Neural Network (DNN) :

A DNN is a deep learning model that has an input layer, output layer, and multiple hidden layers. We used a DNN model, having three hidden layers with 128,64, and 32 neurons, respectively. We used ReLU activation function to learn complex patterns. Adam optimizer, algorithm choice for model convergence, and binary cross-entropy loss function compiled the model. DNN was trained on the scaled training set for 100 epochs with a batch size of 32, and 20% of the training data was used for validation. The model was evaluated using evaluation metrics as shown in Fig. 6.

Fig. 6. Deep Neural Network

1-D CNN:

Convolutional neural networks are a deep learning algorithm that takes in an image, assigns learnable weights and biases to various aspects in the picture, and distinguishes one from another. 1-D CNN can be adapted for tabular data by treating each feature as a sequential input. Convolution is performed on the input data by a CNN using a kernel to produce a feature map. The network architecture included two Conv1D layers with 32 and 64 filters (kernel size 3) and ReLU activation. A GlobalMaxPooling1D layer was used to condense feature maps. A sigmoid output layer was used to predict the probability of breast cancer. The model was trained using the Adam optimizer and binary cross-entropy loss over 100 epochs with a batch size of 32, using 20% of training data for validation. The model was then evaluated using evaluation metrics as shown in Fig. 7.

Fig. 7. Convolutional Neural Network

Results and Discussion

Results

This section looks at the outcomes of different models that we trained on the breast cancer dataset. The WDBS dataset was used to assess how well the AI model performed.

```
================================================================================
MODEL COMPARISON RESULTS
================================================================================
                 Model  Accuracy  Precision  Recall  F1-Score  AUC-ROC
0  Logistic Regression    0.9825     0.9861  0.9861    0.9861   0.9954
1        Random Forest    0.9561     0.9589  0.9722    0.9655   0.9939
2                  SVM    0.9825     0.9861  0.9861    0.9861   0.9950
3                  MLP    0.9474     0.9853  0.9306    0.9571   0.9944
4              Deep NN    0.9561     0.9855  0.9444    0.9645   0.9917
5               1D CNN    0.9035     0.9420  0.9028    0.9220   0.9749

BEST PERFORMING MODELS:
------------------------------
Accuracy: Logistic Regression (0.9825)
Precision: Logistic Regression (0.9861)
Recall: Logistic Regression (0.9861)
F1-Score: Logistic Regression (0.9861)
AUC-ROC: Logistic Regression (0.9954)
```

Fig. 8. Model Comparison Results

The following picture shows the comparison among the six AI models that were trained and evaluated using metrics such as Accuracy, Precision, Recall, F1 score, and AUC-ROC. Logistic Regression achieves 98.25% Accuracy, 98.61% Precision, 98.61% Recall, and 99.54% AUC-ROC, demonstrating exceptional performance characteristics. The F1-score of 98.61% indicates a strong balance between precision and recall, highlighting the model's effectiveness in accurately identifying cases while minimizing both false positives and false negatives. Logistic Regression was considered to be the most effective for the accurate prediction of malignant and benign tumors, offering potential for early detection of breast cancer.

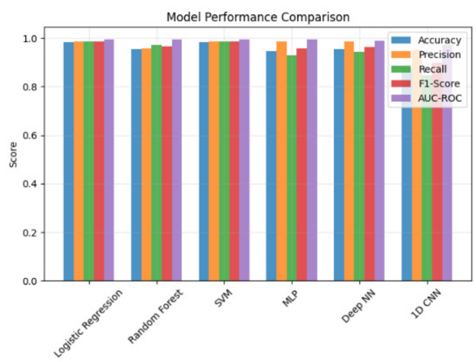

Fig. 9. Model Performance Comparison

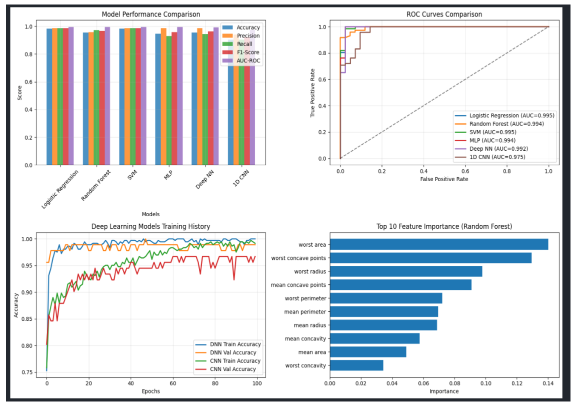

Fig. 10. ROC Curves Comparison

The above figure represents a bar graph of the comparison of model performance on different metrics and the ROC Curves Comparison. The ROC curve shows the performance of binary classification at different threshold values. By comparing these curves, we can evaluate how well each model distinguishes between malignant and benign tumors. A model with a curve closer to the top-left corner better separates the binary classes. Among the models evaluated, Logistic Regression demonstrated the best overall performance.

Discussion

Artificial Intelligence has the power to transform the medical sector, including diagnosing lethal diseases, such as breast cancer. However, despite its various applications and high accuracy, there are many obstacles to integrating it into the real world. In the context of Nepal, the expenses are high to integrate AI into the medical field. As a result, people either move across borders, especially to India, in the hope of quality treatment. It is therefore essential to find effective ways to integrate AI into the healthcare system to enable early cancer detection and improve patient survival rates.

Conclusions

This study demonstrates that Logistic Regression is the most effective model for the early detection of breast cancer, achieving the highest accuracy of 98.25%. The superior performance of Logistic Regression can be attributed to the small size and tabular nature of the dataset, where simpler, interpretable models often outperform more complex deep learning architectures.

References

Alashban, B., Hendawi, Y., et al. (2025). Feature selection and deep learning architecture of mammography and tomosynthesis images for breast cancer classification. *Diagnostic and Interventional Radiology, 31*(2), 95–109.

Alom, M. Z., Taha, T. M., et al. (2025). DNBCD: An explainable CNN+BiLSTM model for breast cancer diagnosis. *Scientific Reports, 15*(1), 14510. https://doi.org/10.1038/s41598-025-24458-2

Alzahrani, N. A., Zidan, A., et al. (2025). Early breast cancer detection via infrared thermography using a CNN enhanced with particle swarm optimization. *Sensors, 25*(3), 1050. https://doi.org/10.3390/s25031050

Arravalli, D., Ugwoke, H., Olayinka, E., & Afolabi, I. (2025). Detection of breast cancer using machine learning and explainable AI. *Scientific Reports, 15*(1), 26931. https://doi.org/10.1038/s41598-025-21919-9

Chang, Y.–W., Ryu, J. K., et al. (2025). AI-STREAM: AI for breast cancer screening in mammography. *Nature Communications, 16*(1), 2248. https://doi.org/10.1038/s41467-025-57469-3

Chen, J., Pan, T., Zhu, Z., et al. (2025). A deep learning–based multimodal medical imaging model for breast cancer screening. *Scientific Reports, 15*(1), 12496. https://doi.org/10.1038/s41598-025-99535-2

Chen, X., Fang, X., et al. (2025). MFF-HistoNet: A multi-modal feature-fusion network for histopathology image classification. *Scientific Reports, 15*(1), 7113. https://doi.org/10.1038/s41598-025-10387-4

Choe, S. J., Arif, A., et al. (2023). Vision-Transformer-based transfer learning for accurate breast cancer classification on mammogram images. *Diagnostics, 13*(9), 1572. https://doi.org/10.3390/diagnostics13091572

Das, S. K., Santra, S., & Zhang, Y. (2023). Breast cancer detection: Comparing shallow CNN and fine-tuned pretrained models. *Frontiers in Genetics, 13*, 890653. https://doi.org/10.3389/fgene.2022.890653

Eisemann, N., Bunk, S., et al. (2025). Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nature Medicine, 31*, 917–924.

Elkorany, S. E., & Elsharkawy, M. (2023). Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance. *European Physical Journal Plus, 138*(1), 981.

Frazer, H. M. L., Peña-Solórzano, C. A., et al. (2024). Comparison of AI-integrated pathways with human–AI interaction in population mammographic screening. *Nature Communications, 15*(1), 7525. https://doi.org/10.1038/s41467-024-51725-8

Guo, Z., Gao, M., et al. (2024). An improved MobileNetV3 for fine-grained breast histology classification. *Applied Sciences, 14*(17), 11932. https://doi.org/10.3390/app141711932

Gurcan, F. (2025). Enhancing breast cancer prediction through stacking ensemble and deep learning integration. *PeerJ Computer Science, 11*, e1648. https://doi.org/10.7717/peerj-cs.1648

Islam, M. Z., Tanvir, S., & Hossain, M. (2024). Predictive modeling for breast cancer classification in a low-resource setting. *Informatics in Medicine Unlocked, 39*, 101207. https://doi.org/10.1016/j.imu.2023.101207

Kallah-Dagadu, S., Ogunleye, A. A., Ransome-Kuti, O., & Bogere, M. (2025). Breast cancer prediction based on gene expression data using interpretable ML techniques. *Scientific Reports, 15*(1), 6608.

Latha, M., Kumar, P. S., et al. (2024). Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and explainable AI. *BMC Medical Imaging, 24*(1), 230. https://doi.org/10.1186/s12880-024-01404-3

Liu, B., Liu, S., Cao, Z., et al. (2025). Accurate classification of benign and malignant breast tumors in ultrasound imaging with an enhanced deep learning model. *Frontiers in Bioengineering and Biotechnology, 13*, 1526260.

Liu, B., Zhang, J., et al. (2025). BreastXploreAI: TransBreastNet for multimodal mammography classification. *Scientific Reports, 15*(1), 7534. https://doi.org/10.1038/s41598-025-10387-4

Ramamoorthy, P., Ramakantha Reddy, B. R., Askar, S. S., & Abouhawwash, M. (2024). Histopathology-based breast cancer prediction using deep learning methods for healthcare applications. *Frontiers in Oncology, 14*, 1300997. https://doi.org/10.3389/fonc.2024.1300997

Raja, C. P., Raju, J., et al. (2025). DenseCapsNet: A hybrid deep learning model optimized by a metaheuristic for breast cancer classification. *Scientific Reports, 15*(1), 21887. https://doi.org/10.1038/s41598-025-37090-5

Rafiq, A., Jaffar, A., Latif, G., Masood, S., & Abdelhamid, S. E. (2025). Enhanced multi-class breast cancer classification from whole-slide histopathology images using a proposed deep learning model. *Diagnostics, 15*(5), 582. https://doi.org/10.3390/diagnostics15050582

Xiang, M., Lin, J., & Wang, K. (2024). Quantum convolutional neural network for breast cancer prediction. *Scientific Reports, 14*(1), 21610. https://doi.org/10.1038/s41598-024-53774-8