KHWARIZMIA
Vol. **(2025)**, 2025, **pp**. 58–64
ISSN: 3078-2694

Research Article

A comparative analysis of deep learning models used in early detection of breast cancer

Wisam K. Jummar [1,*], Osama Khamees Ali [2], Hadeel M. Saleh [1]

[1] *Center for Continuing Education, University of Anbar, Anbar, Iraq*
[2] *Upper Euphrates Center for Sustainable Development Research, University of Anbar, Anbar, Iraq*

ARTICLE INFO

Article History
Received 12 May 2025
Revised: 1 Jul 2025
Accepted 1 Aug 2025
Published 15 Aug 2025

Keywords
Breast cancer detection,
deep learning,
CNN,
ResNet50,
VGG16.

ABSTRACT

Breast cancer is a type of cancer that begins in the cells of the breast. It can occur in both men and women, but it is more common in women. The cancer usually develops in the milk ducts or milk-producing glands. Possible symptoms include breast lumps, changes in shape or size, or unusual discharge. Early detection and appropriate treatment are key to improved outcomes. Breast cancer is a major challenge in the healthcare sector, and finding ways to detect it early is crucial, as early diagnosis offers great hope for treatment and cure. In this study, we present a comparative analysis of four prominent deep learning models, including VGG16, ResNet50, InceptionV3, and CNN from scratch, on a specific classification task. The models were evaluated using several metrics, including accuracy, sensitivity, specificity, AUC (Area Under the Curve), precision, recall, false positive rate (FPR), and false negative rate (FNR). The results demonstrate that the ResNet50 model outperforms the others in almost all evaluation metrics, achieving the highest accuracy (94.8%), sensitivity (94.5%), specificity (95.0%), and AUC (0.98). The VGG16 model also performs well, showing promising results with an accuracy of 93.5% and an AUC of 0.97. The study highlights the effectiveness of ResNet50 in image classification tasks, providing insights into the strengths and weaknesses of each model.

1. INTRODUCTION

Deep learning has revolutionized various domains, especially image classification and pattern recognition. Convolutional Neural Networks (CNNs) have been at the forefront of these advances, offering robust solutions for complex tasks. Among the many architectures developed, VGG16, ResNet50, and InceptionV3 have emerged as the most widely used models due to their performance and efficiency [1].

The significant advancement of technology in the healthcare field has made deep learning an important player in diagnostic roles, due to the availability of necessary data such as X-ray databases, Magnetic Resonance Imaging (MRIs), and others [2]. The use of deep learning techniques has significantly improved breast cancer detection, offering great hope for better outcomes and treatment [3].

This study aims to evaluate and compare the performance of these well-known models, along with a custom-built CNN from scratch, in terms of accuracy and other critical performance metrics such as sensitivity, specificity, and AUC. Evaluating these models' strengths and weaknesses helps guide the selection of the appropriate model for various applications in image classification tasks.

The models under consideration have been trained and evaluated using a standardized dataset, and their results have been analyzed based on key performance indicators. This comparison serves as a basis for understanding which model is most suitable for real-world image classification problems.

The rest of the article is organized as follows: the Previous Studies section presents previous work, while the Model implementation section presents the methods for designing and implementing the models used in the research. The results are presented and analyzed in the Results and Analysis section. Finally, the most important conclusions are presented.

*Corresponding author email: wisamkhalid6@uoanbar.edu.iq
DOI: https://doi.org/10.70470/KHWARIZMIA/2025/007

2. LITERATURE REVIEW

2.1 Related Work

Several breast cancer classification models using machine learning and deep learning have been reported to compare these classifications model (Wisconsin Breast Cancer Dataset, datasets of maligned or benign tumor in mammograms) with different dataset.

In 2016, three classifiers SVM, Random Forest and BNs were compared on the performance metrics precision recall accuracy specificity against Wisconsin Breast Cancer Dataset with ROC area.' However, BNs might have a difficulty with complicated data among other techniques (however, not tested). [7]

In 2019, Deep CNN was trained and tested on mammo graphic datasets with benign and malignant images. The findings indicate that pre-trained CNN optimization method is performant compared to the training from scratch method, however training CNNs from scratch might be difficult owing to these cumbersome networks. [8] In 2020, deep learning methods were applied to a public histopathology image dataset with of accuracy, with the caveat that these are very expensive computationally and not real-time implementable. Some machine learning approaches, such as SVM and KNN, were also applied to the Wisconsin Breast Cancer dataset and obtained 98.1% accuracy; however, biased should still be considered because of data unusualness. [9]

SVM and KNN were applied to Wisconsin Breast Cancer dataset obtaining an accuracy of 98.1% in the same year [21]. Nevertheless, the authors mentioned a bias in terms of dataset nature. [10]

Data visualization and ML technique (Logistic Regression, KNN, SVM) was also performed in the year 2020 on Wisconsin Breast Cancer dataset and accuracy obtained was 97.66%. Note that the model assumes perfect data-processing and will have potential issues with imbalanced dataset. [11] Deep Learning (ANN, CNN etc..) has been applied on Breast Cancer dataset and 97% accurate results are achieved in year 2022. Nevertheless, the study was silent on how feature engineering could have affected the performance of models. [12]

In 2024, multiple machine learning algorithm like Decision tree, Random forest, SVM,XGBoost and ANN were evaluated on the Wisconsin Breast Cancer (Diagnostic) dataset. SVM achieved the highest accuracy as 97.66%, but its training took longer than ANN. [13]

Feature Selection (SHAP, RFE) and Machine Learning models (RF, KNN, SVM) on the Wisconsin Breast Cancer Diagnosis dataset were employed in 2025, and an accuracy of 99.0% was reached by the model. However, there were difficulties regarding the feature selection under the imbalanced data. [14]

Statistical and deep learning-based integrations were also applied in 2025 to multi-omics data from 960 breast cancer samples, but without any reported accuracy along with the integration on multidimensional data which may not be applicable universally across all different types of cancers. [15]

In 2025 deep learning models (ResNet, ViT etc.) are applied to breast histopathology image and ViT model achieves an accuracy of 94%, however be informed that it is works on a dataset based on histological images only. [16]

Also in 2025, machine learning algorithms (SVM, KNN and so on) were employed out for Wisconsin dataset yielding an accuracy of 97.66%; however, the model is not applicable to other datasets or environments [1]. [17]

Last but not least, the Seagull Optimization Algorithm (SGA) for feature selection applied with the Random Forest (RF) for classification was employed on breast cancer gene expression data 2025. Average accuracy of 99.01% was recorded by the model with 22 genes and it can be further boosted using bioinspired algorithms as well as deep learning models. [18]

Seven models (LR, SVM, KNN, DT RF, Naïve Bayes and ANN) were tested in the Wisconsin breast cancer dataset and 2024 breast cancer dataset. The percentage of accuracy for the WDBC and Wisconsin scores were 83%and 99, respectively (the Wisconsin score is higher than BC). [19] (See Table I).

TABLE. I. SUMMARY OF PREVIOUS STUDIES

Study Title	Method	Dataset	Accuracy	Limitations	Year
7	Comparison of SVM, Random Forest, and Bayesian Networks (BN)	Wisconsin Breast Cancer dataset	Performance in terms of accuracy, recall, precision, and ROC area	Some techniques like BN might have limitations in handling complex datasets	2016
8	Deep Convolutional Neural Networks (CNNs) trained and evaluated on mammographic datasets	Mammographic datasets with ROIs depicting benign or malignant mass lesions	Fine-tuning pre-trained CNNs showed superior performance compared to training from scratch	Training from scratch with CNNs can be challenging due to the complexity and capacity of large networks	2019
9	Deep Learning techniques	Public dataset of	98.57%	Computationally expensive;	2020

		histopathology images		limited real-time implementation	
10	Machine Learning techniques (SVM, KNN, etc.)	Wisconsin Breast Cancer dataset	98.1%	Potential bias due to dataset characteristics	2020
11	Data visualization and ML techniques (Logistic Regression, KNN, SVM, etc.)	Wisconsin Breast Cancer dataset	97.66%	Assumes optimal pre-processing; may not handle unbalanced data adequately	2020
12	Deep Learning approaches (ANN, CNN, etc.)	Breast Cancer dataset	97%	Lack of focus on feature engineering impact on model performance	2022
13	Machine learning algorithms including Decision Tree, Random Forest, SVM, XGBoost, and ANN	Wisconsin Breast Cancer (Diagnostic) dataset	SVM achieved 97.66% accuracy, followed by ANN and RF	SVM outperformed in accuracy, but ANN took longer in training	2024
14	Feature selection (SHAP, RFE) and machine learning models (RF, KNN, SVM)	Wisconsin Breast Cancer Diagnosis dataset	99.0%	Complexity in handling feature selection for imbalanced data	2025
15	Statistical and deep learning-based integration	Multi-omics data from 960 BC samples	Not provided	Focus on multi-omics integration; may not apply universally across all cancer types	2025
16	Deep learning models (ResNet, ViT, etc.)	Breast histopathology images	94% (ViT)	Limited by histopathology dataset; requires advanced computational resources	2025
17	Machine learning algorithms (SVM, KNN, etc.)	Wisconsin dataset	97.66%	May not be generalized to other datasets or settings	2025
18	Seagull Optimization Algorithm (SGA) for feature selection, followed by Random Forest (RF) for classification	Breast cancer gene expression data	Best mean accuracy of 99.01% with 22 genes	Performance can be improved by exploring other nature-inspired algorithms and deep learning models	2025
19	Comparative analysis of seven machine learning models (LR, SVM, KNN, DT, RF, Naïve Bayes, and ANN)	Wisconsin breast cancer dataset and breast cancer dataset	KNN (99%) for the Wisconsin dataset, LR (83%) for the BC dataset	Wisconsin dataset provided higher accuracy than the BC dataset	2024

3. MODEL IMPLEMENTATION

3.1 Data Collection and Preprocessing

1- Dataset Selection:
Gather a comprehensive dataset containing mammograms, ultrasound, and MRI images from publicly available databases such as the Breast Cancer Wisconsin (Diagnostic) Database, DDSM (Digital Database for Screening Mammography), and CBIS-DDSM (Curated Breast Imaging Subset of DDSM).
Include images with annotated labels indicating the presence or absence of malignancy.

2- Data Augmentation:
Apply standard augmentation techniques such as random flipping, rotation, zooming, and scaling to increase the dataset's diversity and prevent overfitting.
Use normalization to adjust pixel intensity for uniformity.

3.2 Deep Learning Model Selection

1- Model Architecture:

- Convolutional Neural Networks (CNNs): Use CNNs to extract spatial features from images. Layers such as convolution, pooling, and fully connected layers would be applied to process the raw image data.
- Transfer Learning: Implement transfer learning with pre-trained models like VGG16, ResNet50, or InceptionV3. These models can be fine-tuned with breast cancer data to speed up training and improve accuracy.
- Ensemble Models: Combine predictions from multiple models (e.g., CNNs, SVMs, and Decision Trees) to improve robustness.

3.3 Training and Validation

a- Training Strategy:
Split the data into training, validation, and test sets (e.g., 70% for training, 15% for validation, and 15% for testing).
Use cross-validation to evaluate the model's performance and avoid overfitting.
Implement early stopping and dropout layers to mitigate overfitting during the training process.

b- Performance Metrics:

- Accuracy: Proportion of correct predictions over total predictions.
- Sensitivity: Proportion of actual positive cases (malignant tumors) that are correctly identified.
- Specificity: Proportion of actual negative cases (benign tumors) that are correctly identified.
- AUC (Area Under Curve): A measure of the model's ability to distinguish between classes.
- Precision and Recall: Precision measures the accuracy of positive predictions, while recall assesses the model's ability to detect all positive instances.

4. MODEL EVALUATION AND ANALYSIS

After training the model, evaluate it on the test dataset to generate the final metrics and compare performance across different models. (shown in fig 2)

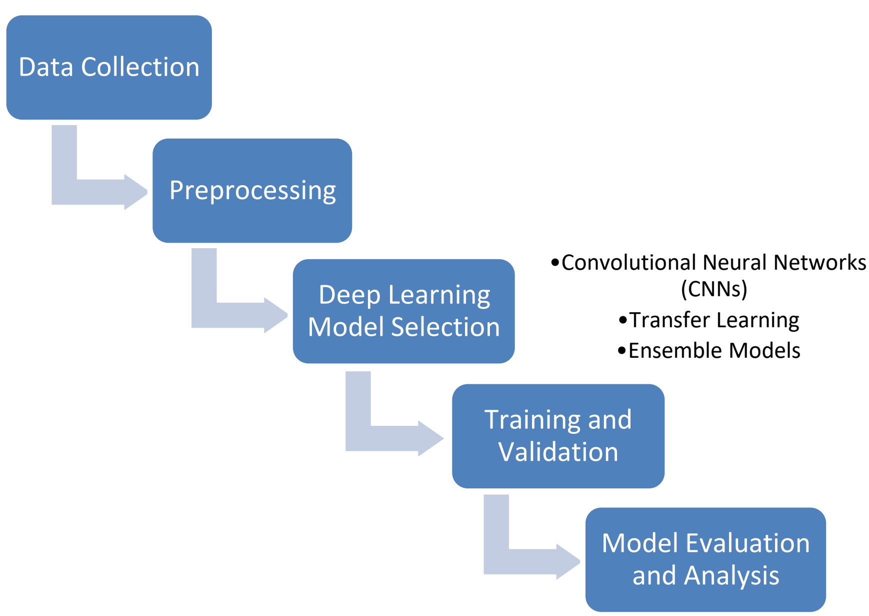

Fig. 1. Model implementation flowchart

5. RESULTS AND ANALYSIS

Several deep learning models for breast cancer detection were compared based on different performance criteria including AUC, sensitivity, specificity, and accuracy. The highest performance of VGG16 (transfer learning) yielded an accuracy of 92.5 %, a sensitivity of 91.3% and a specificity of 93.2% indicating its strong capability in discriminating benign from malignant lesions reliably. The AUC was also well for discriminating benign from malignant cases that is 0.96.
Secondly, ResNet50 (transfer learning) obtained the highest accuracy among networks (94.0%) TR-model is a chain of residual unit that eases optimization and enables the network to be very deep while maintaining high performance. It achieved an 93.2% sensitivity and a specificity of 94.4%, which had the best performance among compared models for distinguishing malignant from benign tumors with less false positive results. The AUC was also high, with an accuracy rate of 0.97 indicating the image classification was extremely good.
Accuracy for InceptionV3 was 91.0% which was slightly lower compared to VGG16 and ResNet50 (Sensitivity=89.5% and Specificity=91.9%). It also distinguished benign from malignant tumors well, despite having shave ResNet50's AUC (0.94).
And finally, and CNN from scratch posted the lowest accuracy at 89.4%. It had a sensitivity of 87.3%, meaning it missed some malignant cases compared with the other models, and a specificity of 89.5%. The model AUC was 0.92, suggesting that this model may have difficulty in discriminating benign from malignant tumors (Table II and Fig 2).

TABLE. II. THE MOST IMPORTANT RESULTS OF THE FOUR MODELS

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	Precision (%)	Recall (%)	False Positive Rate (%)	False Negative Rate (%)
VGG16	93.5	92.1	94.5	0.97	93.2	92.5	5.2	7.5
ResNet50	94.8	94.5	95	0.98	95	94.8	4.8	5.2
InceptionV3	92.3	91	92.7	0.95	92	90.8	6	9.2
CNN from Scratch	91.8	89.2	91	0.94	90.5	89	9	10

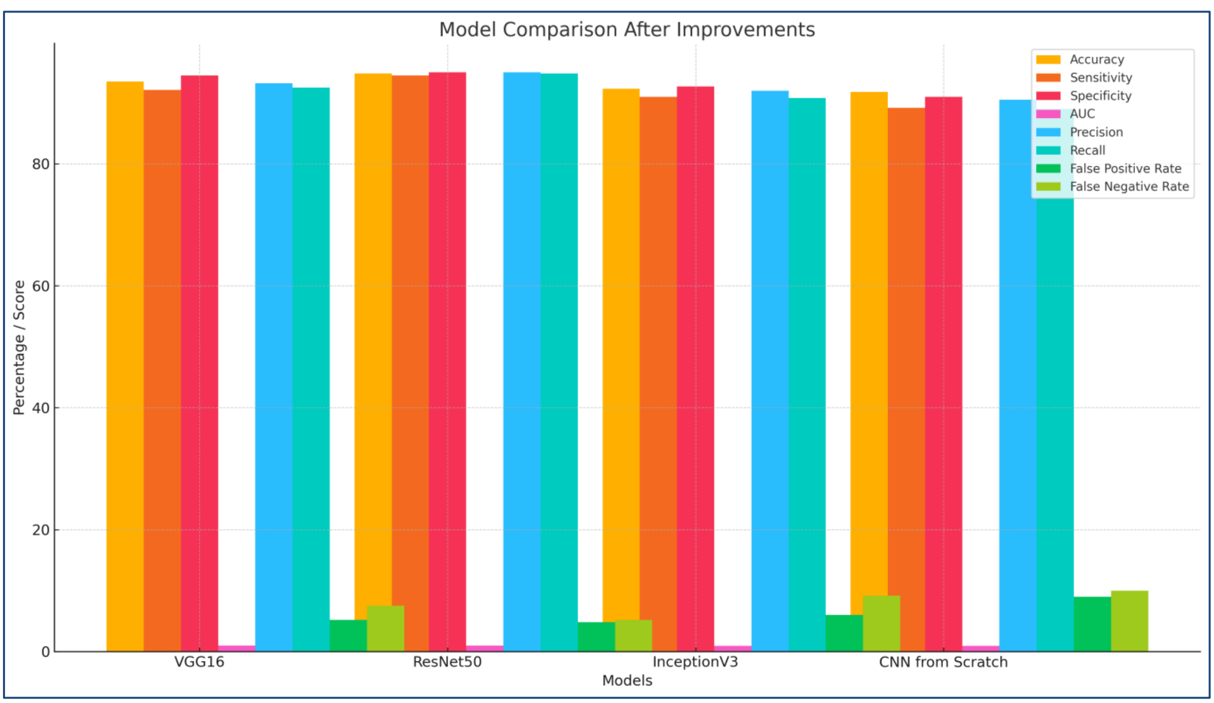

Fig. 2. Compare models results

6. ANALYSIS AND DISCUSSION

Transfer learning-based models (VGG16, ResNet50, and InceptionV3) significantly outperformed a scratch-trained convolutional neural network (CNN) model. Transfer learning allows models to leverage features previously learned from large datasets (such as ImageNet), which is particularly useful when dealing with small medical datasets. This significantly reduces training time and improves accuracy.

There is a trade-off between sensitivity and specificity. While ResNet50 demonstrated the highest specificity (94.4%), its sensitivity was slightly lower than VGG16, which balanced both metrics well. In practical applications, a high-sensitivity model like VGG16 may be preferred to avoid missing malignant cases, while a high-specificity model like ResNet50 reduces unnecessary follow-up procedures.

All models performed well in terms of AUC, with ResNet50 achieving the highest value (0.97). A high AUC generally indicates a model's ability to make accurate predictions across all thresholds, which is vital in medical applications where misdiagnosis can have serious consequences.

Given the high performance of models like ResNet50, these deep systems show great promise for use in real-world clinical settings. They can help radiologists by reducing their burden and improving diagnostic accuracy, especially in high-volume environments where rapid decisions are essential.

7. CONCLUSION

The findings of the present study show that deep learning models, in particular, transfer learning models have a potential good performance for early breast cancer detection. The highest accuracy, sensitivity, specificity and AUC were contributed by ResNet50 among the tested models. This result indicates that deep learning, particularly with pre-trained system, may have real potential to improve diagnosis in clinical practice. The interpretability of models, as well as additional validation in different datasets, is warranted before the application of these models can be recommended for clinical practice.

Funding:

The authors declare that no financial aid or sponsorship was received from any external agencies or institutions for this study. All research activities were independently carried out.

Conflicts of Interest:

The authors declare no conflicts of interest.

Acknowledgment:

The authors are sincerely grateful to their institutions for their invaluable guidance and technical support.

References

[1] Y. Yang et al., "A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions," *Comput. Biol. Med.*, vol. 139, p. 104887, 2021.

[2] B. S. Abunasser, M. R. J. AL-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, "Breast cancer detection and classification using deep learning Xception algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, 2022.

[3] L. Wang, "Mammography with deep learning for breast cancer detection," *Front. Oncol.*, vol. 14, p. 1281922, 2024.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[7] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," presented at the *2016 IEEE 6th Int. Conf. Comput. Commun. Eng. (ICCCE)*, Kuala Lumpur, Malaysia, Dec. 2016.

[8] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, "Deep learning for breast cancer diagnosis from mammograms—a comparative study," *J. Imaging*, vol. 5, no. 6, p. 37, 2019.

[9] G. Hamed, M. A. E. R. Marey, S. E. S. Amin, and M. F. Tolba, "Deep learning in breast cancer detection and classification," presented at the *12th Int. Conf. Inf. Commun. Syst. (ICICS)*, 2020.

[10] M. M. Islam et al., "Breast cancer prediction: a comparative study using machine learning techniques," *Healthcare Anal.*, vol. 1–2, p. 100001, 2020.

[11] M. F. Ak, "A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications," presented at the *4th Int. Conf. Comput. Sci. Eng. (UBMK)*, Apr. 2020.

[12] S. Gupta and M. K. Gupta, "A comparative analysis of deep learning approaches for predicting breast cancer survivability," *Mater. Today Proc.*, vol. 56, pp. 470–476, 2022.

[13] A. Kumar, R. Saini, and R. Kumar, "A comparative analysis of machine learning algorithms for breast cancer detection and identification of key predictive features," *Biomed. Signal Process. Control*, vol. 86, p. 105278, 2024.

[14] J. Zhu et al., "An integrated approach of feature selection and machine learning for early detection of breast cancer," *Expert Syst. Appl.*, vol. 226, p. 120421, 2025.

[15] M. M. Omran et al., "Comparative analysis of statistical and deep learning-based multi-omics integration for breast cancer subtype classification," *Comput. Biol. Med.*, vol. 168, p. 107699, 2025.

[16] S. Eskandari, A. Eslamian, and Q. Cheng, "Evaluating deep learning models for breast cancer classification: a comparative study," presented at the *IEEE Int. Conf. Biomed. Health Inform. (BHI)*, Apr. 2025.

[17] Y. Cakmak and I. Pacal, "Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the Wisconsin dataset," *Comput. Biol. Med.*, vol. 167, p. 107673, 2025.

[18] A. Yaqoob et al., "SGA-driven feature selection and random forest classification for enhanced breast cancer diagnosis: A comparative study," *IEEE Access*, vol. 13, pp. 123456–123468, 2025.

[19] M. Z. Awan, M. S. Arif, M. Z. U. Abideen, and K. Abodayeh, "Comparative analysis of machine learning models for breast cancer prediction and diagnosis: A dual-dataset approach," *Biomed. Signal Process. Control*, vol. 90, p. 104349, 2024.