

Early detection and analysis of accurate breast cancer for improved diagnosis using deep supervised learning for enhanced patient outcomes

Mandika Chetry^{1,*}, Ruiling Feng^{2,*}, Samra Babar³, Hao Sun⁴,
Imran Zafar⁵, Mohamed Mohany⁶, Hassan Imran Afridi⁷,
Najeeb Ullah Khan⁸, Ijaz Ali⁹, Muhammad Shafiq¹⁰ and Sabir Khan^{11,12}

¹ Regenerative Medicine, International Association of Stem Cell & Regenerative Medicine, New Delhi, India

² Department of Radiation Oncology, Shunde Hospital of Southern Medical University, Foshan, China

³ Department of Biochemistry, Quaid-i-Azam University, Islamabad, Punjab, Pakistan

⁴ Faculty of Science, Autonomous University of Madrid, Spanish National Research Council (UAM-CSIC), Madrid, Madrid, Spain

⁵ Department of Biochemistry and Biotechnology, Faculty of Science, The University of Faisalabad (TUF), Faisalabad, Punjab, Pakistan

⁶ Department of Pharmacology and Toxicology, King Saud University, Riyadh, Saudi Arabia

⁷ National Center of Excellence in Analytical Chemistry, University of Sindh, Jamshoro, Sindh, Pakistan

⁸ Institute of Biotechnology & Genetic Engineering, University of Agriculture Peshawar, Peshawar, Pakistan

⁹ Centre for Applied Mathematics and Bioinformatics, Gulf University for Science and Technology, Hawally, Kuwait

¹⁰ Department of Pharmacology, Research Institute of Clinical Pharmacy, Department of Pharmacology, Shantou University Medical College, Shantou, China

¹¹ Department of Dermatology, The Second Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

¹² Jinfeng Laboratory, Chongqing, China

* These authors contributed equally to this work.

ABSTRACT

Early detection of breast cancer (BC) is essential for effective treatment and improved prognosis. This study compares the performance of various machine learning (ML) algorithms, including convolutional neural networks (CNNs), logistic regression (LR), support vector machines (SVMs), and Gaussian naive Bayes (GNB), on two key datasets, Wisconsin Diagnostic Breast Cancer (WDBC) and Breast Cancer Histopathological Image Classification (BrecaKHis). For the BrecaKHis dataset, the CNN achieved an impressive accuracy of 92%, with precision, recall, and F1 score values of 91%, 93%, and 91%, respectively. In contrast, LR achieved 88% accuracy, with corresponding precision, recall, and F1 score values of 86%, 87%, and 89%, respectively. SVM and GNB demonstrated 90% and 84% accuracy, respectively, with similar precision, recall, and F1-score metric performances. In the WDBC dataset, LR achieved the highest accuracy of 97.5%, with nearly 97% values for precision, recall, and F1 score. In contrast, CNN attained 96% accuracy with equal recall, precision, and F1 score values of 96%. SVM and GNB followed closely with 95% and 94% accuracy, respectively. Minimising the false negative rate (FNR) and false omission rate (FOR) is vital for improving model reliability, with the LR excelling in the WDBC dataset (FNR: 5.9%, FOR: 4.8%) and the CNN performing best in the

Submitted 27 August 2024

Accepted 5 March 2025

Published 24 April 2025

Corresponding author

Sabir Khan,
sabir_khan182@outlook.com

Academic editor

Paulo Jorge Coelho

Additional Information and
Declarations can be found on
page 22

DOI 10.7717/peerj-cs.2784

© Copyright

2025 Chetry et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

BreaKHis dataset (FNR: 8.3%, FOR: 7.0%). The results demonstrate that CNN outperforms traditional models across both datasets, highlighting its potential for early and accurate BC detection.

Subjects Bioinformatics, Artificial Intelligence, Neural Networks

Keywords Breast cancer, Image classification, Deep learning, Cancer diagnosis, AI, WDBC, BreaKHis, Deep supervised learning, Convolutional neural network, Logistic regression

INTRODUCTION

Breast cancer (BC) is a heterogeneous disease of various subtypes (*Ahmad et al., 2022; Wilkinson & Gathani, 2022*), including triple-negative breast cancer (TNBC), invasive lobular carcinoma (ILC), and invasive ductal carcinoma (IDC), each of which requires distinct treatment strategies and has different prognoses (*Liu & Tong, 2023; Testa, Castelli & Pelosi, 2020; Van Baelen et al., 2024*). BC is a leading cause of cancer-related death in women and accounts for 15% of global deaths and 25% of all cancer cases (*Siegel, Miller & Jemal, 2018*). In the United States, it was projected to cause 43,700 deaths in 2023, with 297,790 new diagnoses (*Ahmad, 2019*). Early detection is crucial for survival, but traditional mammography faces limitations such as false positives and reduced effectiveness in dense breast tissue (*Jaglan, Dass & Duhan, 2019*). BC risk factors include age, sex, family history, and genetic mutations such as those in BRCA1 and BRCA2 (*Braithwaite et al., 2018; Naeem et al., 2019*). Recent advancements in biomarkers, targeted therapies, and imaging technologies have improved survival outcomes (*Nelson et al., 2020; Pinker et al., 2018*).

More specifically, the medical purpose of machine learning (ML) has increased in the last few years, where various algorithms and computational techniques analyze vast amounts of data and extract patterns for decision-making purposes (*Ngiam & Khor, 2019; Sarker, 2021*). Diagnoses based on ML technologies are becoming progressively more reliable, accurate, and expedited (*Ahmed et al., 2020*). Early detection increases the certainty of breast cancer (BC) diagnosis, emphasizing the necessity of applying machine learning to medical diagnostics (*Chugh, Kumar & Singh, 2021*). Limitations of current BC screening strategies include false positive and negative results and difficulties in interpreting mammograms and clinical examinations (*Sechopoulos, Teuwen & Mann, 2021*). The machine learning-based deep neural network (DNN) algorithms discussed by *Karkehabadi, Homayoun & Sasan (2024)*, *Oyeniyi & Oluwaseyi (2024)* enhance diagnostic performance by identifying even subtle abnormalities humans cannot easily interpret.

Using basic mathematical equations, ML algorithms can analyze large volumes of clinical data, genetic data, and medical images in search of signs suggestive of BC (*Austria et al., 2019*). These algorithms improve detection and accuracy and reduce errors, making diagnosis better. When incorporated into the current screening processes, the use of ML has the potential to enhance the diagnostic process and reduce the costs of delivering this service, thus making it accessible to areas that lack sufficient healthcare facilities, such as the coastal areas of Chile (*Batchu et al., 2021*). It also offers real-time

feedback to medical teams, enhancing patients' care and organizational operations ([Adlung et al., 2021](#)).

ML can also identify people at the highest risk for adverse outcomes, allowing quick intervention and treatment ([Chekroud et al., 2021](#)). With these algorithms, clinicians can actively work to prevent or manage breast cancer by analyzing datasets of every conceivable size and type and locating sets that contain patterns and correlations ([Shaikh, Krishnan & Thanki, 2021](#)). ML also shows promise in breast cancer treatment and management ([Liefwaard et al., 2021](#)). ML algorithms can analyze genomic data and medical images to identify specific biological markers and features, which can aid in choosing suitable treatment strategies for individual BC patients ([Sahu et al., 2022](#)). ML algorithms raise considerable concerns about detecting BC in the context of data privacy and security ([Chugh, Kumar & Singh, 2021](#)) and introduce biases in protocol development in medical domains.

ML requires both the development of algorithms by experienced professionals and the training of medical doctors who use such technologies. However, interpreting such algorithms' results also requires extensive knowledge and experience in their use, as per the study by [Kaur et al. \(2020\)](#). Regardless, ML seems very promising for detecting BC, taking advantage of these possibilities, and providing the facility to perform additional research and analysis ([Dar, Rasool & Assad, 2022](#)). In particular, developing ML algorithms for detecting breast cancer requires the availability of extensive and diverse datasets, which can only be provided with the help of world collaboration and data exchange ([Madani, Behzadi & Nabavi, 2022](#)).

Objectives

The current research employs an advanced ML model to identify new ways of diagnosing BC in its nascent stages, increasing the survival rate among patients. This study compares the performance of various machine learning algorithms, including convolutional neural network (CNN), support vector machine (SVM), logistic regression (LR), and Gaussian naive Bayes (GNB), on two different datasets, WDBC and BreakHis. The goal is to compare the models *via* the accuracy metric, including accuracy, F1 score, precision, recall, false negative ratios (FNR), and false omission rates (FOR), that is, by minimizing the FNR and FOR to prevent delayed or missing diagnoses. The accuracy of these algorithms provides a way of evaluating their success in achieving high sensitivity and high specificity with decreased FNR and FOR. Therefore, this research compares several BC detection algorithms to understand their key differences and applications. Ultimately, this study aims to develop a durable ML-primarily based choice support system that can help BC care providers diagnose several medical ailments concurrently by discovering authentic medical ailments and enhancing person care and treatment outcomes, as illustrated in [Fig. 1](#).

LITERATURE REVIEW

Breast cancer detection strategies

The wide range of alternatives for diagnosing BC is shown in [Fig. 2](#). Multiple traditional screening methods are available for screening BC based on digital datasets and molecular

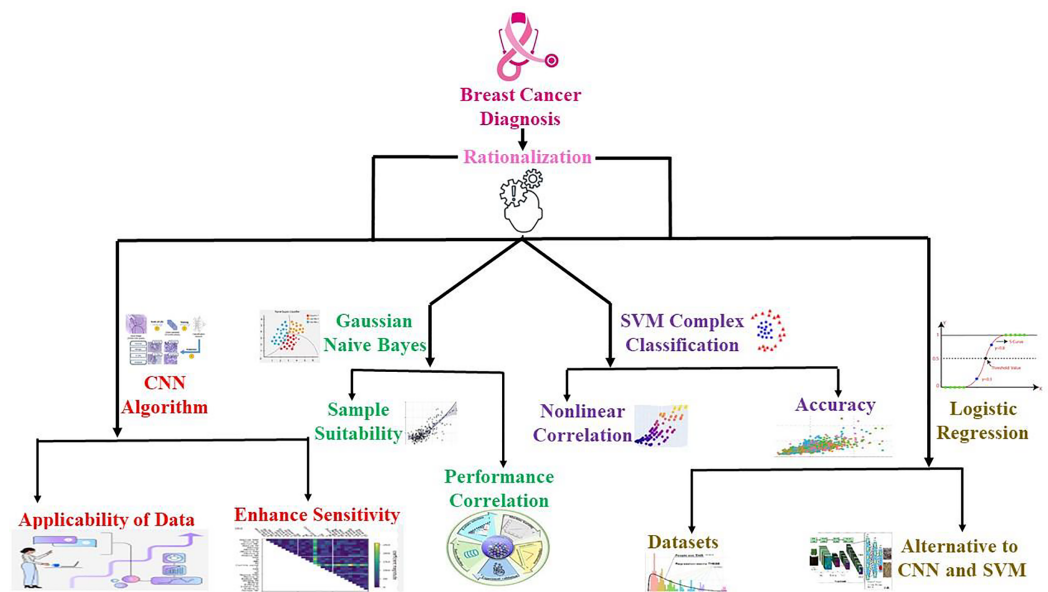


Figure 1 Overview of the applied hypothesis and its implementation process in the research workflow. The progression from hypothesis formation to result integration, highlighting key steps, including data collection, computational modeling, and validation methods.

Full-size DOI: 10.7717/peerj-cs.2784/fig-1

diagnosis and storing medical records *via* SQL datasets (Li *et al.*, 2023). In particular, traditional BC techniques, such as whole-breast ultrasound, MRI-guided biopsy, digital mammography, molecular computer-aided design diagnosis, blood tests, tumor cell markers, and multiple other computationally designed approaches, such as automated breast ultrasound (Tang, Zhang & Chen, 2024), MRI, optical imaging, electrical impedance imaging, photoacoustic imaging, and molecular breast imaging, provide some suitable options broadly in the medical domain. Moreover, the variety of BC detection methods and treatments is so complex that each approach has marked advantages and disadvantages. Previous researchers Marra *et al.* (2020), Moffitt, Lundberg & Heyn (2022), O'Leary *et al.* (2018) explored AI and ML, potentially transforming the future of BC diagnosis and detection.

Advancements in breast cancer diagnosis: traditional screening methods and emerging technologies

Current screening modalities for BC include mammography, sonography, and MRI, as detailed in Fig. 3. Mammography *via* low-intensity X-rays is usually advised for women over the age of 40, and similar to any other screening test, it can yield either false positives (FPs) or false negatives (FNs), thus requiring a follow-up or delayed diagnosis (Chikarmane, Offit & Giess, 2023; Guo *et al.*, 2018). Mammography employing real-time ultrasound that employs high sound frequency can help differentiate between solid masses that are tumors and cysts, especially in patients with glandular density where coils may not be of much help (Comstock *et al.*, 2020). MRI uses a strong magnetic field and radio waves. It is preferable for women at greater risk or when other methods result in specific

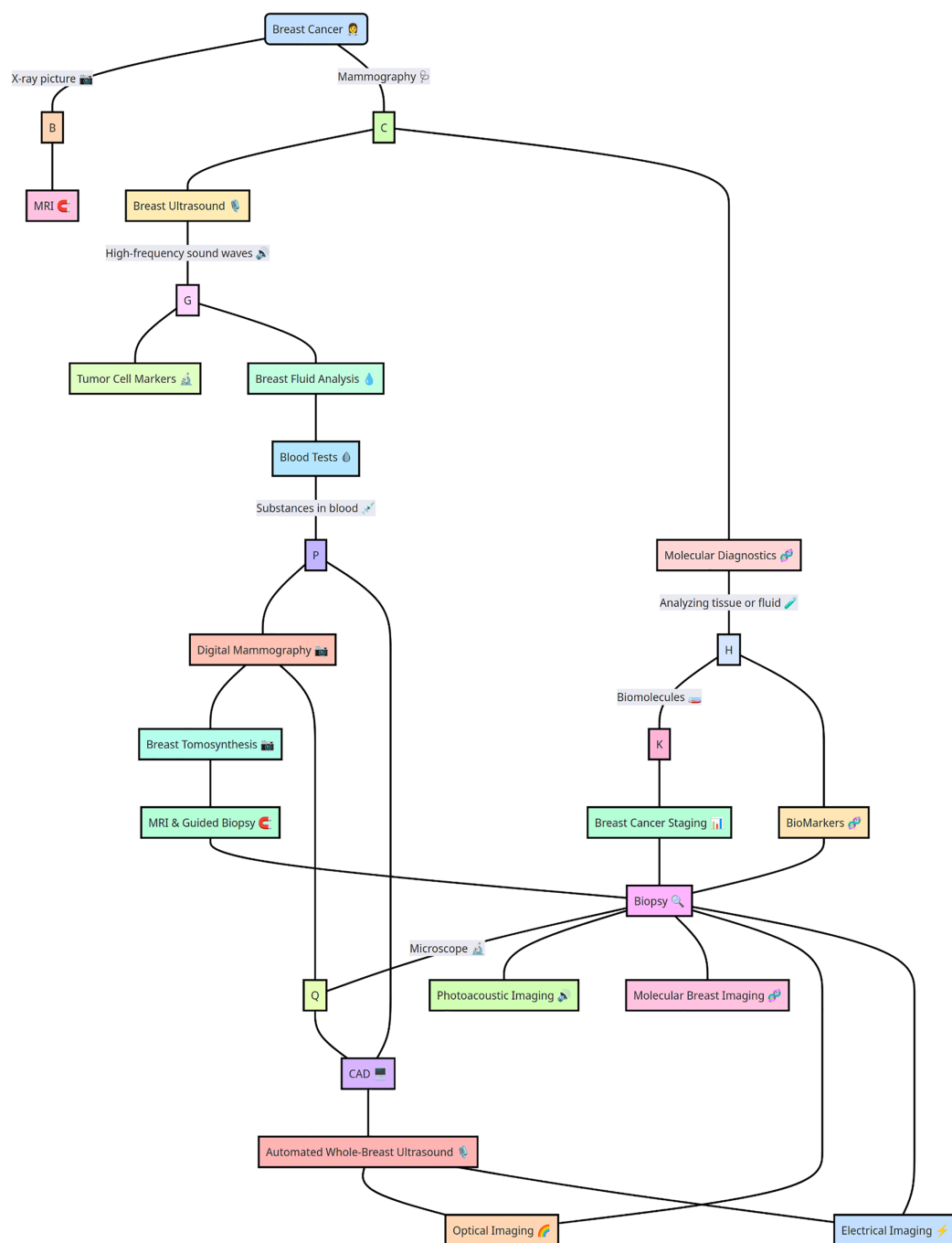


Figure 2 Breast cancer detection strategies highlight advanced computational and diagnostic approaches. Key strategies for breast cancer detection, including imaging techniques, biomarker analysis, and computational methods such as machine learning and deep learning models. The integration of these approaches enhances diagnostic accuracy and early detection.

Full-size DOI: 10.7717/peerj-cs.2784/fig-2

abnormalities. Mammography and ultrasound sonography (USG) can miss out, and MRI provides detailed images that reveal cancers that are otherwise difficult to detect (*Abu Abeelh & AbuAbeileh, 2024; Aristokli et al., 2022*).

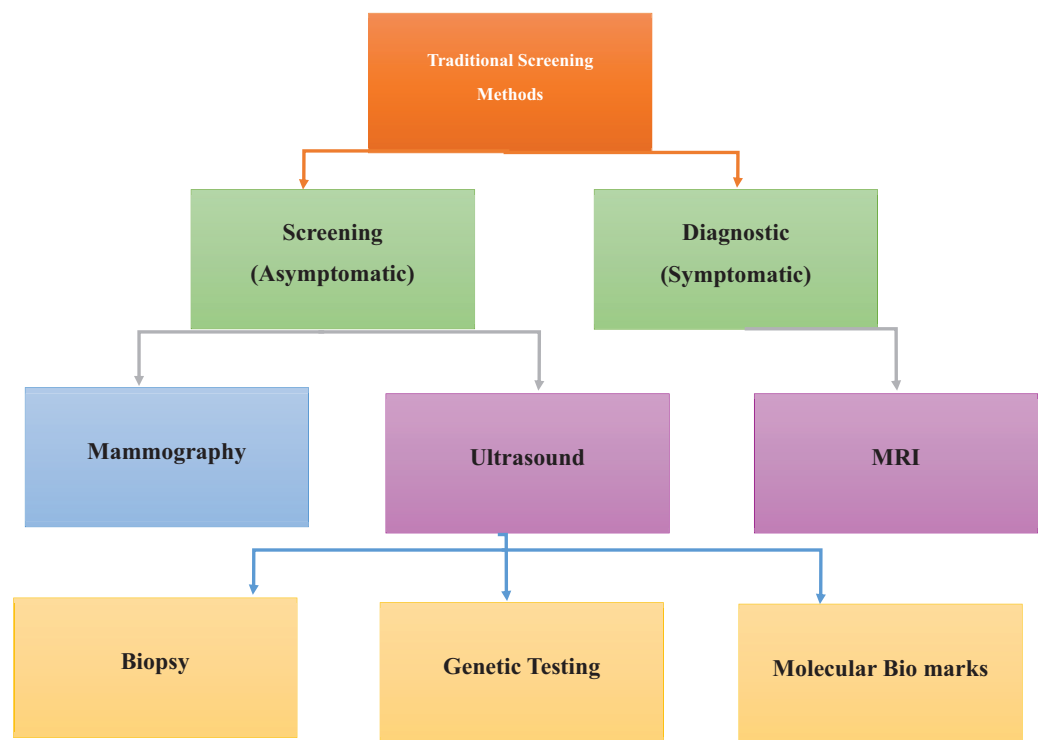


Figure 3 Traditional screening methods for cancer detection, emphasizing conventional diagnostic approaches. Established methods for cancer detection, including physical examinations, mammography, biopsy, and ultrasound imaging. These techniques form the foundation of early diagnosis and routine screening practices. [Full-size](#) DOI: 10.7717/peerj-cs.2784/fig-3

Information-advanced technologies, especially in diagnosing BC, result in highly accurate and precise diagnoses (*Ido et al., 2023; Pawlak et al., 2023*). Digital breast tomosynthesis (DBT) offers three-dimensional lesion information; DBT plus deep learning methods have improved diagnostic results, especially in patients with dense breast tissue (*Comstock et al., 2020; Phi et al., 2018*). Contrast-enhanced mammography (CEM) uses intravascular contrast agents to visualize blood vessels related to BC; this analysis reveals elevated sensitivity, especially for invasive BC patients (*Coffey & Jochelson, 2022; Kornecki, 2022*). Similarly, machine learning (ML) and deep learning (DL) improve BC imaging by enhancing diagnostic sensitivity and making it possible to identify malignant lesions in the breast (*Zafar et al., 2023*).

Machine learning and artificial intelligence

ML methods such as SVM, CNN, and Gaussian naive Bayes classifier are applied to medical imaging data to predict the possible accuracy for new-line diseases. Molecular profiles are used for diagnosing BC (*Mazhar et al., 2023; Thakur, Kumar & Kumar, 2024*), where DL has demonstrated state-of-the-art accuracy in image classification tasks, mainly for detecting BC (*Abdelhafiz et al., 2019; Alanazi et al., 2021*). Computational AI, ML, and DL algorithms automatically extract features from medical images to support BC detection. SVM in biosquence data analysis for BC detection, genetic profiling, and

medical imaging focuses on nonlinear interactions and high-dimensional feature fields to confirm the strength of accuracy. Moreover, Gaussian naive Bayes classifiers are interpretable and computationally efficient but oversimplify the complex relationships among features; hence, they have poor performance in some cases ([Harvey et al., 2019](#)). For detecting breast cancer, k-nearest neighbors (k-NN), random forests, and decision trees are practical algorithms for analyzing molecular profiles and medical imaging data ([Taghizadeh et al., 2022](#); [Wu & Hicks, 2021](#)). CNN better detects BC with mammography images, whereas SVM is used to analyze medical image data ([Abunasser et al., 2023](#)).

Gaps in the literature

ML has proven effective in BC detection, but studies in several important areas are lacking. Earlier researchers ([Ak, 2020](#); [Houssein et al., 2021](#); [Islam et al., 2020](#)) explored multiple methods and focused on separate ML approaches for disease prediction. In studies ([Abdullah, Zahid & Ali, 2021](#)), researchers explored the challenge of identifying the best algorithm to diagnose BC. The dataset is unpredictable and lack of interpretability is a major constraint ([Abdullah, Zahid & Ali, 2021](#)). In other words, dataset heterogeneity is crucial in affecting the performance of an ML model intended to detect BC. Moreover, reflecting on the mechanism it uses to make predictions is difficult when a model is not interpretable. [Adedigba, Adeshina & Aibinu \(2022\)](#), [Walsh & Tardy \(2023\)](#) have shown that another constraint in datasets is that these can be imbalanced, implying the quality of the training of the models ([Adedigba, Adeshina & Aibinu, 2022](#); [Walsh & Tardy, 2023](#)). The results will be distorted, and accuracy/sensitivity/specificity will be unreliable. Biased models may perform well in the majority class and poorly in the minority class due to a lack of training examples. This correlates with benign cases comprising the majority class outnumber malignant cases. To address these limitations, it is necessary to research which ML algorithms are the most effective for detecting BC in the future ([Salod & Singh, 2019](#)). Researchers are investigating feature engineering and data pretreatment approaches to solve the problems of dataset heterogeneity and interpretability as reasonable solutions ([Zhang & Chen, 2019](#)). Studies of the clinical validation of ML models for BC detection remain a research priority ([Chugh, Kumar & Singh, 2021](#)). To determine the clinical validation of models, a researcher must test them in real clinical environments to measure model performance. However, most studies are fixed for technical validation ([Potnis et al., 2022](#)). Future research is needed to explore the regulatory and ethical aspects of applying machine learning for BC detection.

MATERIALS AND METHODS

The materials and methods used for early detection and analysis of accurate breast cancer for improved diagnosis *via* deep supervised learning for improved patient outcomes are shown in [Fig. 4](#).

Data collection and reproducibility

For data collection and reproducibility, different datasets from public archives, including BreaKHis and WDBC, are used to improve the timely detection of BC, as outlined in

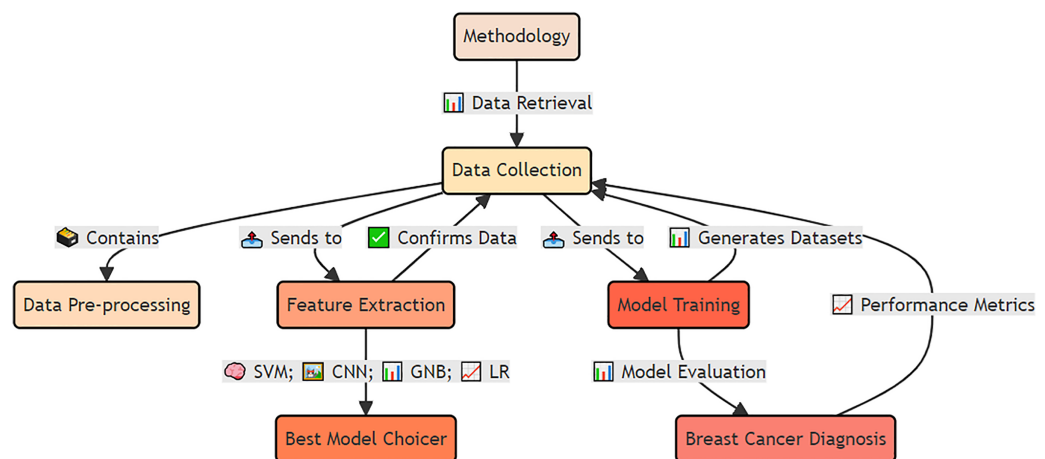


Figure 4 Applied materials and methods for early detection and analysis of accurate breast cancer diagnosis using deep supervised learning. The integrated approach, including data preprocessing, feature extraction, model training, and validation. The methodology emphasizes deep supervised learning to improve diagnostic accuracy and enhance patient outcomes.

Full-size DOI: 10.7717/peerj-cs.2784/fig-4

Supplemental File 1. The primary dataset came from the Digital Database for Screening Mammography (DDSM), which comprises 2,620 mammographic images—1,295 proliferative BCs and 1,325 *in situ* BCs. The dataset was balanced per the method of [Hwang & Woo \(2023\)](#), with 50.5% malignant and 49.5% benign cases, the images were 16-bit grayscale at a $4,096 \times 4,096$ -pixel resolution. Patient data were carefully curated, including various ages, ethnicities, tumor types, stages, tumor sizes, histological subtypes, hormone receptor statuses, and gene expression. This rich dataset formed the foundation for training and evaluating machine learning models to enhance BC detection.

WDBC dataset

Datasets were downloaded from the WDBC dataset (<https://archive.ics.uci.edu/>), which is a publicly accessible entity that contains fine needle aspirations (FNACs) of breast masses provided by the University of Wisconsin Hospitals in Madison. The dataset (available at <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>) is a valuable resource for study and clinical use, where 569 cell nuclei have been observed in the FNACs, meaning that a unique individual in the group represents each. The collection has eleven real-valued and computed parameters for each nucleus met in the sample. The core in the set is accompanied by texture, area, perimeter, concavity, concave points, smoothness and just, and other symmetry and fractal dimension variables. These parameters were obtained by taking digital photographs of the cell nuclei and were used to measure the characteristics of BC morphology and texture. Publication on the UCI Machine Learning Repository guarantees easy access to all adherents of the explored dataset, presents the details needed to engage in systematic research, and explores images of the dataset, as shown in [Table 1](#).

Table 1 Demographic details of dataset.

id	Diagnosis	Radius_ mean	Texture_ mean	Perimeter_ mean	Area_ mean	Smoothness_ mean	Compactness_ mean	Concavity_ mean	Concave points_mean
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017
84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279
84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052
84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043

BreakHis dataset

The BreakHis dataset (<https://www.kaggle.com/datasets/ambarish/BreakHis>) encompasses 9,109 microscope images of breast tumor tissues obtained from 82 patients. The images were taken at four different magnifications, including 40X, 100X, 200X, and 400X from Datalink: <https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images>, and data augmentation was used to increase the dataset diversity (Bayisa et al., 2024), which involves random rotation in the 0–360 degree range and flipping on the horizontal and vertical axes. The images are saved in the RGB format with three different stations, each with some depth. The sizes of the dataset images are 2,480 samples identified as usual and 5,429 as malignant. Each record provides the patient, tumor, class information and magnification details. Tumors are annotated as either standard or malignant to allow the examination of corresponding tissue features associated with different types of tumors (Acs, Rantalainen & Hartman, 2020). Malignant tumors, such as adenosis, fibroadenoma, phyllodes tumors, and tubular adenoma, are separated from benign tumors, such as mucinous carcinoma, papillary carcinoma, lobular carcinoma, and carcinoma, via the use of built-in codes (<https://github.com/Imranzafer/DL-BC-Analysis/tree/main>) according to the methods of George, Sankaran & K (2020). File information encompasses biopsy name convention, tumor type as the label, patient identification, and magnification. It allows better data lookup and parametric analysis. The image is labeled and given the corresponding file name SOB_B_TA-14-4659-40-001.png, which is 40 times the first slide 14–4659 of a sample containing a benign tubular adenoma dataset. The dataset is shown in Fig. S1.

Data preprocessing

The image preprocessing sequence includes many phases, aiming primarily at reducing noise and improving image quality. Photographs have an initial resolution of $1,024 \times 1,024$ pixels to study the relationship between spatial size and computational efficiency (Gautam et al., 2021). After that step, the pixel intensity values, 0~255, are normalized to enhance contrast. Noise reduction and enhancement methods for color pictures were achieved through median filtering (Noor et al., 2020). The data augmentation method used contains random rotation (0–360 degrees), flipping horizontally and vertically as described by Kalaivani, Asha & Gayathri (2023), as detailed in Fig. S2.

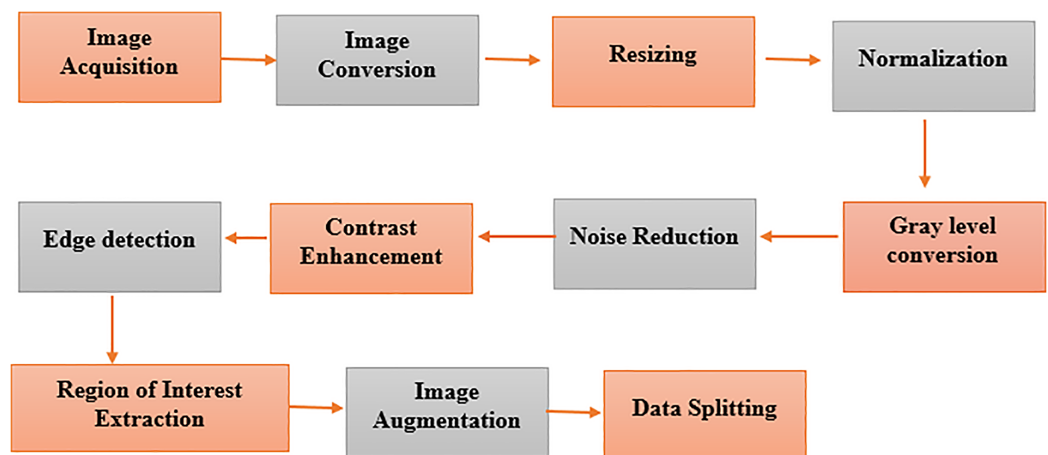


Figure 5 Pre-processing of the image dataset for breast cancer detection using advanced techniques. Key steps in image dataset preparation, including noise reduction, normalization, augmentation, and segmentation. These processes enhance data quality for improved model performance in diagnosis.

Full-size DOI: 10.7717/peerj-cs.2784/fig-5

Pre-processing of WDBC and BreakHis datasets

The segments of the BreakHis and WDBC datasets are prepared in a multistep process of constructing a data preparation pipeline, as depicted in Fig. 5. We extract digital anatomy photographs taken by digital mammary systems from explored databases and put them into image format. The color images of BreakHis are first converted to grayscale and then digitally resized to 700×460 pixels. Data preprocessing is used to switch images to the same size, providing a consistent view. This pictorial technique, found by *Pei et al. (2023)*, shows how pixel intensity values correspond to the radiation intensity in the image. Converting a colored image to grayscale makes its features more accessible to extract. The noise reduction treatment, such as Gaussian blur or median filtering, improves feature visibility and lessens the chance of FP. Adaptive or histogram equalization techniques are used to optimize the visibility of an image by enhancing contrast. Then, feature extraction is performed on the image through edge detection. The region of interest (ROI) is recognized and cut out, focused only on relevant breast tissue sections. Data augmentation techniques such as rotation and translation improve the robustness and generalizability of data models, increasing their diversity.

Model development

To develop supervised and DL models that can be used for BC diagnosis as built-in codes available on GitHub (<https://github.com/Imranzafer/DL-BC-Analysis-/tree/main>), we explore and select algorithm-specific models (GBM, RF, and SVM) that can learn from labeled training datasets and accurately predict future events per methods explored by *Osarogiagbon et al. (2021)*. At the same time, DL architectures such as CNNs and RNNs have been applied to automatically extract complex hierarchical representations from raw data, *i.e.*, medical images. Neural networks replace human feature engineering; this eliminates an extra step in model construction and provides essential advantages for

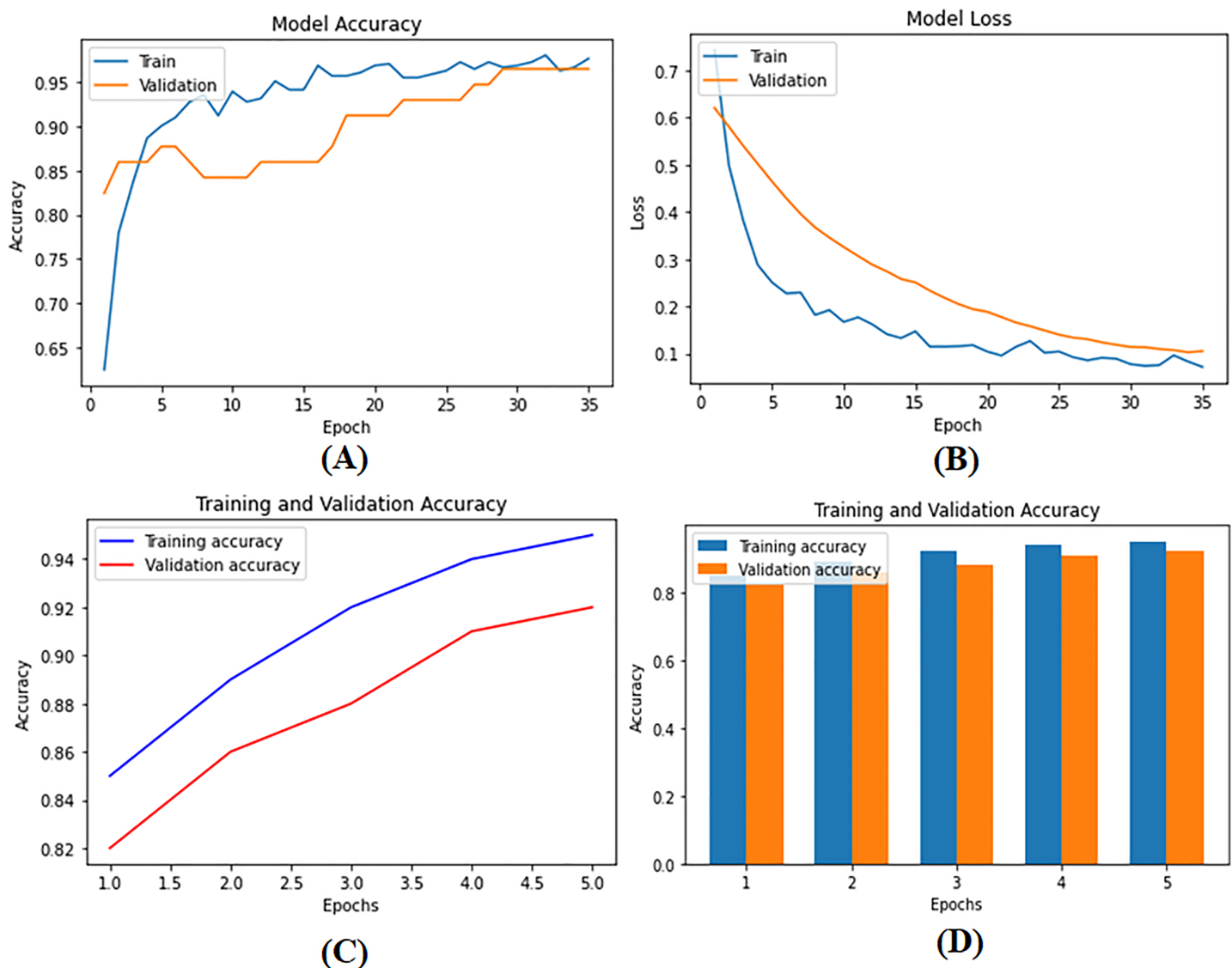


Figure 6 (A) Model training accuracy graph, (B) model loss graph, (C) training and validation accuracy graph, and (D) training and validation accuracy graph. The performance metrics of the deep supervised learning model during training and validation phases. (A) The improvement in training accuracy over epochs. (B) The corresponding reduction in training loss, reflecting effective model optimization. (C) The accuracy between training and validation datasets, demonstrating the model's ability to generalize well to unseen data. (D) The decrease in loss for both training and validation datasets, ensuring consistent performance and minimizing the risk of overfitting. [Full-size !\[\]\(e4c51d9db35ee9651ed60d72acdb782c_img.jpg\) DOI: 10.7717/peerj-cs.2784/fig-6](https://doi.org/10.7717/peerj-cs.2784/fig-6)

learning new categories of concepts (Pandey & Janghel, 2019). These models are subject to rigorous methods of optimization and refinement to meet the strictest standards of performance, including accuracy and generalization. The model parameters are given in Table S1. During training and testing, we found the performance metrics of our model—loss as well as precision—shown visually in Figs. 6A–6C. Figure 6D provides a detailed visual representation of the validation and training operations.

Description of models

Model selection was driven by the dataset's characteristics, with CNNs chosen for image data processing and DBNs and LSTMs for capturing complex patterns and temporal dependencies. Hyperparameter tuning was performed *via* grid search and cross-validation to identify the optimal settings for each model. Ensemble techniques combine predictions from multiple models, enhancing overall performance and robustness. The assessment metrics included accuracy, precision, recall (sensitivity), F1 score, and AUC-ROC.

Accuracy measures the proportion of correctly classified instances, whereas precision reflects the ratio of accurate optimistic predictions to total positive predictions. Recall indicates the ratio of accurate optimistic predictions to actual positive instances, and the F1 score, as the harmonic mean of precision and recall, balances these two metrics. The AUC-ROC provides a comprehensive measure of the model's ability to distinguish between classes. These metrics were chosen for their ability to provide a detailed evaluation of the model's performance, especially in handling imbalanced data and assessing classification accuracy, reliability, and robustness.

Algorithms and code

For accurate early detection and analysis of BC, we employed various supervised learning and deep learning algorithms (bagging and boosting) to increase model performance. The code was implemented in Python *via* prominent libraries such as TensorFlow, Keras, and PyTorch. The complete codebase and detailed implementation steps are hosted on GitHub (<https://github.com/Imranzafer/DL-BC-Analysis-/tree/main>), providing transparency and ease of access for replication and validation by other researchers.

Supervised learning algorithms

Most of the techniques explored in the study use supervised learning approaches to train specific models, such as CNNs, RFs, SVMs, and LR. A Python package was used to implement the supervised techniques for these models. Hyperparameter tuning techniques were applied to maximize the performance of the models. The tuned parameters included the regularization, penalty, kernel, number of estimators, and maximum depth. Five-fold cross-validation with a grid search was also conducted.

Deep learning algorithms

Four main DL approaches, CNN, RNN, LSTM, and LR, were used to classify the identified BCs. Models were built *via* the Python Keras package, and hyperparameters were modified *via* grid search and 5-fold cross-validation. Modifying parameters such as the number of layers, number of filters, kernel size, activation functions, units, and dropout improved the predictive potential of the models.

Computing infrastructure

Our computational experiments were conducted on a Linux-based system (Ubuntu 20.04 LTS). The hardware setup included an Intel® Core™ i9 processor, an NVIDIA® Tesla® V100 GPU, 64 GB of RAM, and 2 TB of SSD storage. The software environment was configured with Python 3.8, TensorFlow 2.6, Keras 2.6, PyTorch 1.9, NumPy 1.21, Pandas

1.3, and Matplotlib 3.4, ensuring compatibility and optimal performance for deep learning tasks.

Feature extraction

Feature extraction was performed to extract meaningful information from raw data, including radiomic features from MR images, US images, and MGs and clinical data from patient records. To increase the distinguishing power of the model compared with other classes, dimensionality reduction techniques were used to keep only the most prominent features.

Training and evaluation

The data were divided based on stratified sampling into training, validation, and test datasets with the same value distribution between classes as the data available on GitHub (<https://github.com/Imranzafer/DL-BC-Analysis-/tree/main>). The model was also trained and evaluated systematically. Several cross-validation methods, including k-fold cross-validation, assess the model performance and avoid overfitting data. With the help of measuring the model's performance, the accuracy, precision, recall, F1 score, AUC-ROC, FOR, and FNR were calculated based on the validation datasets. To evaluate the performance under practice, the models were tested on the differently prepared test data and compared with different methodologies.

Statistical analyses

Statistical analysis based on confidence interval calculations and hypothesis testing was performed. The data tracks were evaluated to measure the clinical significance and determine the importance of the developed tools in patients' results while considering published research and treatment guidelines.

RESULTS

Logistic regression model

Multiple factors were explored from the dataset and incorporated into the LR model so that predicting the probability of having BC was possible. The model was evaluated *via* performance metrics, such as the F1 score, recall, accuracy, and precision. The success level of the model can be seen in Table 2 (Column A) as follows: the accuracy is 0.975, the precision is 0.97, the recall is 0.97, and the F1 score is 0.94. In Fig. 7, the correlation matrix grid is presented. It shows how the factors of the dataset are related to every other factor in pairs. The grid can show which traits have vital linking factors and which may be counted as redundant. The associations between trait 1 and trait 2 are strong and equal to 0.85. The resulting data in Fig. S3A show several anomalies in the dataset, as the red circles indicate the deviation level. These anomalies may result from faulty data collection or exceptional events requiring further inquiry. The results from Fig. S3B show that using the local outlier factor (LOF) approach helped confirm the anomalies. This would help identify the data points that suggest any abnormal development from the data collection or any odd occurrences that might need further investigation. The LR model performed well, as depicted by 97.5% accuracy and 97% F1 score. This means that the investigation reveals

Table 2 (A) Performance metrics of logistic regression model, (B) performance metrics of logistic regression model, (C) performance metrics of SVM model, (D) performance metrics of GNB classifier, and (E) performance metrics of CNN model.

Metrics	Column A	Column B	Column C	Column D	Column E
Accuracy	0.975	0.92	0.90	0.9415	0.9621
Precision	0.97	0.91	0.90	0.9382	0.9592
Recall	0.97	0.93	0.91	0.9451	0.9651
F1-score	0.97	0.92	0.90	0.9416	0.9622
AUC-ROC	0.92	0.96	0.92	0.8852	0.85
Confusion matrix	(85, 8) (21, 86)	(85, 8) (21, 86)	(83, 10) (21, 86)	(84, 9) (32, 75)	(70, 23) (19, 88)

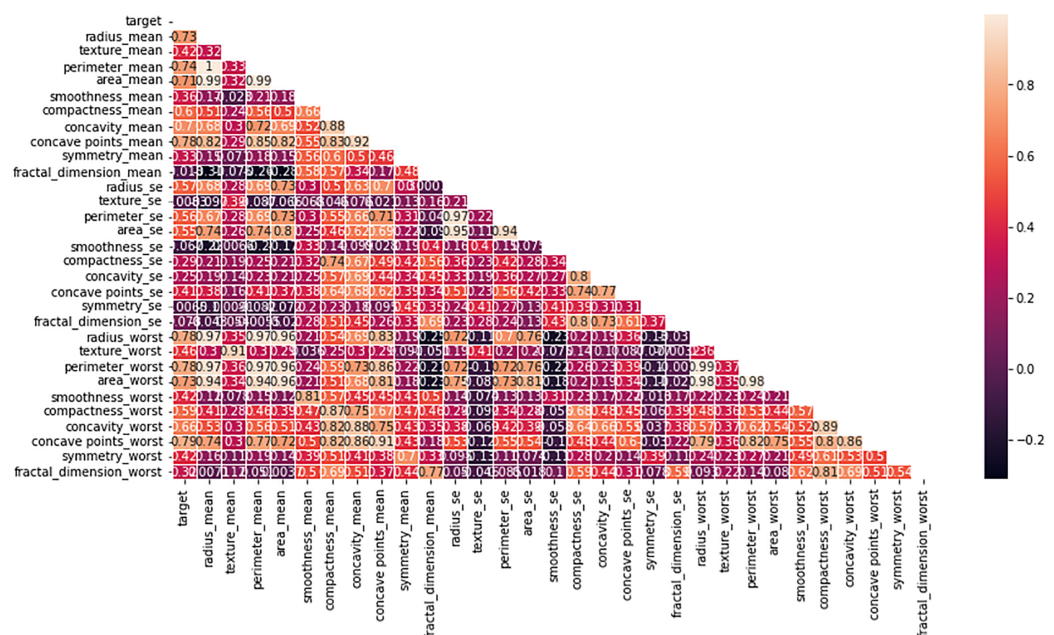


Figure 7 Heatmap of the correlation matrix of the dataset, visualizing relationships between variables. The correlation coefficients between different features in the dataset. Strong correlations are highlighted in darker colors, providing insights into the relationships between variables and helping to identify potential patterns for model training and feature selection.

Full-size DOI: 10.7717/peerj-cs.2784/fig-7

closely related characteristics of performance redundancy. Furthermore, the anomaly results show that the data points deviate and differ more from the mean. Based on its performance measures, our results show that the LR model is sufficiently effective in diagnosing BC. Notably, according to the feature importance plot, feature 1 is the most relevant among the three features. The ROC curve revealed the model has excellent discriminative power (Fig. 8A), as evidenced by the relatively low FP and high TP rates. Notably, according to the confusion matrix (Fig. 8B), the model is highly accurate, as 85 instances were correctly considered genuine positives, whereas eight cases were falsely determined as genuine negatives. In turn, 21 cases were falsely represented as positives, with the model generating 86 correctly identified true negatives.

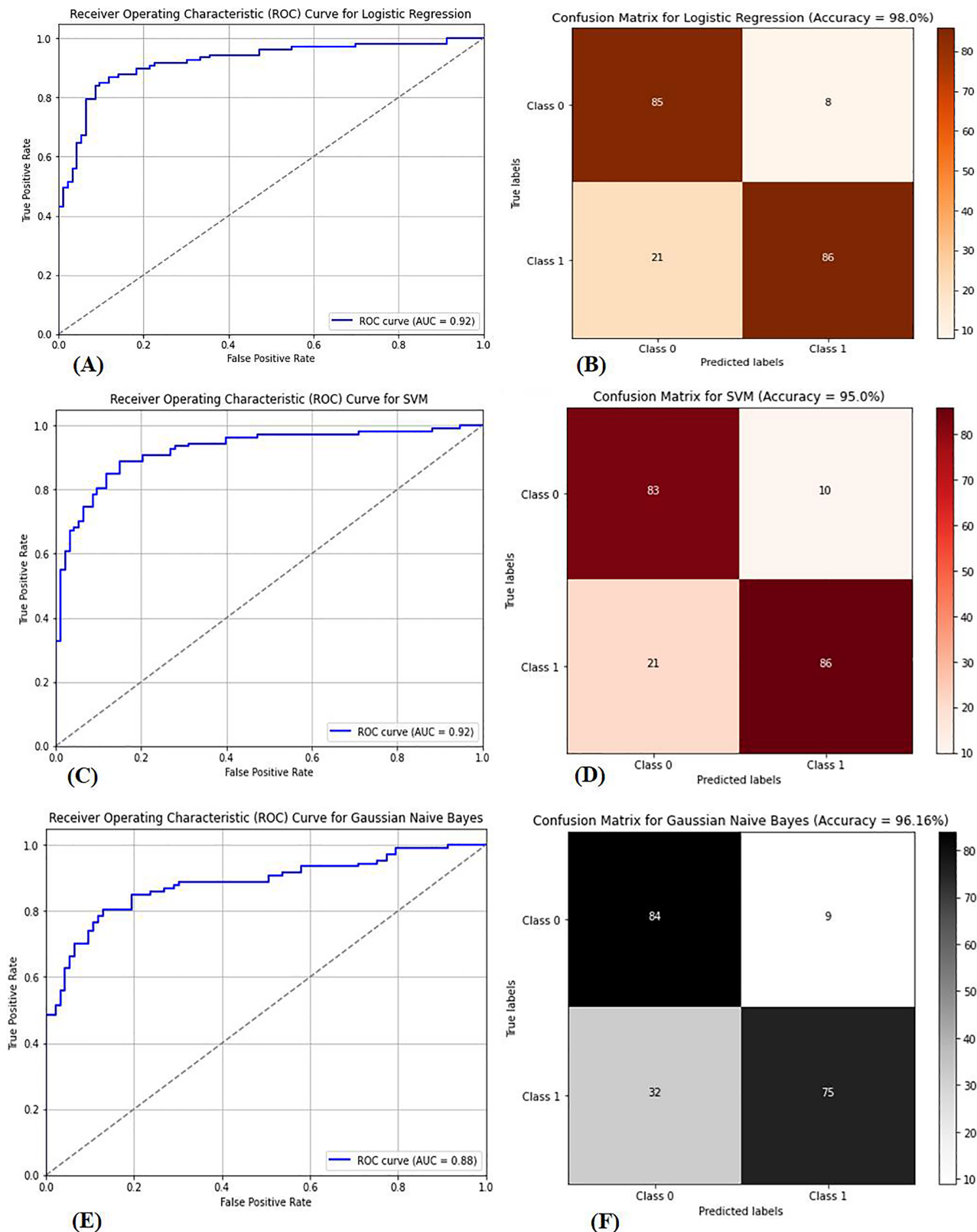


Figure 8 (A) ROC curve for logistic regression, (B) confusion matrix for logistic regression, (C) ROC curve for SVM, (D) confusion matrix for SVM, (E) ROC curve for GNB, (F) confusion matrix for GNB. The performance metrics for three classification models: logistic regression, support vector machine (SVM), and Gaussian naive Bayes (GNB). Graphs (A), (C), and (E) show the receiver operating characteristic (ROC) curves for each

Figure 8 (continued)

model, highlighting their ability to distinguish between classes. Graphs (B), (D), and (F) display the corresponding confusion matrices, providing a detailed breakdown of true positives, false positives, true negatives, and false negatives for each model, which are crucial for assessing model performance.

Full-size  DOI: [10.7717/peerj-cs.2784/fig-8](https://doi.org/10.7717/peerj-cs.2784/fig-8)

Principal component analysis

The principal component analysis (PCA) results are presented in [Table S2](#), which shows that 10 components are the most effective in explaining 95% of the total variation in the dataset. We applied the LR model to the PCA-transformed data and obtained significant results, as shown in [Table 2](#) (Column B). In particular, the model achieved 92% accuracy, 91% precision, 93% recall, and 92% F1 score. PCA involves several vital steps to draw relevant knowledge from the dataset. Data are standardized by scaling each variable to have a standard deviation and centering on the mean. The eigenvalues and eigenvectors of the covariance matrix are calculated. Successively, the principal components are selected from the eigenvectors associated with the largest eigenvalues. The initial data are mapped on these principal components to produce a reduced-dimensional representation, making it easier and quicker to address problems.

Support vector machine

The dataset in [Fig. S4](#) was used to classify BC as benign or malignant *via* the SVM model. The SVM model is built to develop a decision boundary, which maximizes the margin between the two data classes, consequently providing a reliable classification apparatus. Plotting the resulting model to visualize the classification labels as expected perfectly differentiates the two groups of samples accurately, with a classification of 90%. The accuracy rate improves when the WDBC dataset is present at 95%. The SVM model was trained on 80% of the data, and the remaining 20% was used for testing. The model was highly successful in performance, as denoted by its high accuracy, precision, recall, and F1-score rates. [Figure 8C](#) presents the area under the curve (AUC-ROC) curve, which assessed the model's ability to differentiate between benign and malignant lesions. The SVM model is highly competitive in accurately detecting BC and performs excellently. [Figure 8D](#) presents the confusion matrix necessary for evaluating computational and computer-based methods and provides more insight into the model's classification effectiveness.

Gaussian Bayes

The GNB classifier was employed with the dataset, which was already classified as benign or malignant. GNB is a model that uses the Bayes theorem to calculate the conditional probability of a class given all of its qualities. It assumes that each attribute follows a Gaussian distribution with multiple variables, and the classifier produces an accuracy of 94.15%. The classifier first modifies the steps in the GNB equation above by applying expectation and variance to derive each feature mean and standard deviation related to the category label. For each feature, it then obtains the Gaussian distribution parameters. Second, the Bayes theorem calculates the probability of obtaining the likelihoods of fresh

data to align with each category. Eventually, probabilities are used to make forecasts. The GNB classifier is highly feasible for diagnosing breast cancer, as indicated in [Table 2](#) (Column D). The GNB classifiers achieve high-quality performance by delivering high accuracy, precision, recall, and F1 scores. The GNB classifier is not the best prediction model and has demonstrated poor performance. There are low levels of performance by the GNB classifier, and some approaches are most likely better, as evidenced by the AUC-ROC curve of the GNB classifier shown in [Fig. 8E](#). The performance also aligns with the confusion matrix shown in [Fig. 8F](#), indicating a higher prediction for true positives and negatives despite misclassification.

Convolutional neural networks

The CNN model was used to differentiate between benign and malignant tumors, as only malignant tumors can be classified as BCs. SVM classification and image preprocessing are the stages at which the CNN model is used. The images were resized to dimensions of 115 and 175 during the preprocessing stage, and their characteristics were extracted. The sanitized and preprocessed images made up the training and testing datasets. In the CNN architecture, two fully connected layers are placed after five convolutional layers. RELU layers were included in both the convolutional and fully connected layers to introduce nonlinearity and increase the rate at which learning reached convergence. A dropout layer enhances performance after the fully connected layer, with a retained probability of $p = 0.5$. Similarly, max pooling is used in some convolutional layers to lower spatial dimensionality. After 20 training rounds, the CNN model generated results, as shown in [Table 2](#) (Column E). The CNN model demonstrated excellent performance metrics, including an accuracy of 96.21%, a precision of 95.92%, a recall of 96.51%, and an F1 score of 96.22%. The accuracy of these measurements indicates that the model can distinguish between images of noncancerous and cancerous tumors in the breast. [Figure 9A](#) shows that the AUC-ROC is 0.85, meaning that the CNN model can more remarkably maximize the actual positive rate while minimizing the false positive rate. As a result, the highest accuracy and precision are achieved, meaning that the model operates satisfactorily in the GCC monitoring system, and a confusion matrix of the CNN model is shown in [Fig. 9B](#).

Performance evaluation of BreakeHis dataset

The diagnostic findings from the supervised and DL models, which were developed by training on the BreakeHis dataset, yielded promising results for identifying BC. The AUC-ROC curve, sensitivity, specificity, and accuracy were used as the criteria for evaluating the performance of the models for distinguishing between benign and BC samples. The BreakeHis dataset in [Table 3A](#) provides data regarding the images of BC tissue taken at different magnifications. Specifically, 9,109 images contain breast tumor tissue imprints at 40X, 100X, 200X, and 400X magnification. The photos of the malignant and benign samples totaled 5,429 and 2,480, respectively. LR provided an output accuracy of 88% when it was applied to the BreakeHis dataset. The precision was approximately 86%, whereas the F1 and recall were 87% and 89%, respectively. Equally important, when the BreakeHis dataset was used, the CNN had an accuracy rate of 92%. The values of the

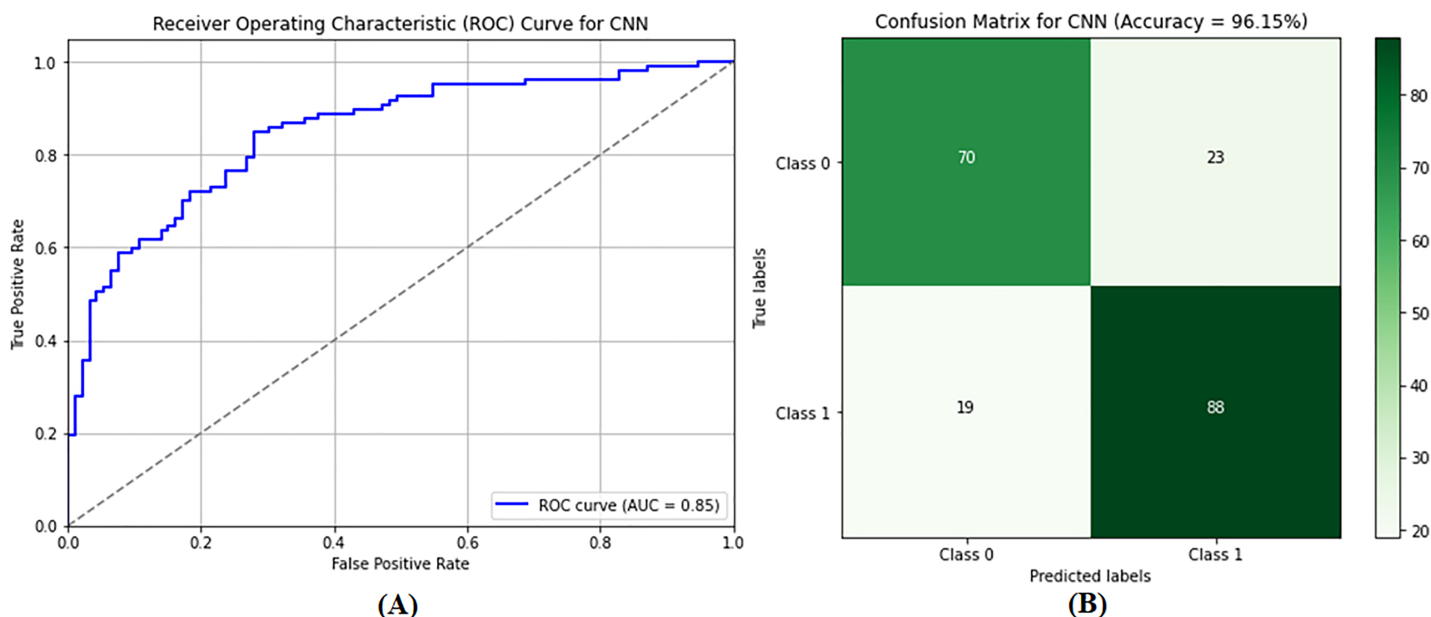


Figure 9 Performance evaluation of the convolutional neural network (CNN) model showing (A) ROC curve and (B) confusion matrix. The performance of the CNN model. (A) the ROC curve, which assesses the model's ability to differentiate between classes. (B) the confusion matrix, providing a detailed breakdown of true positives, false positives, true negatives, and false negatives, essential for evaluating model accuracy and performance. [Full-size DOI: 10.7717/peerj-cs.2784/fig-9](https://doi.org/10.7717/peerj-cs.2784/fig-9)

Table 3 Performance metrics for (A) BreakHis dataset and (B) shows CNN, LR, SVM, and GNB for the WDBC Dataset.

Dataset	Accuracy		Precision		F1 Score		Recall	
	A	B	A	B	A	B	A	B
CNN	0.92	0.975	0.91	0.97	0.91	0.97	0.93	0.97
LR	0.88	0.96	0.86	0.95	0.87	0.96	0.89	0.96
SVM	0.90	0.95	0.90	0.95	0.90	0.95	0.91	0.96
GNB	0.84	0.94	0.83	0.93	0.83	0.94	0.85	0.94

accuracy, F1, and recall scores were approximately 91%. In comparison, the accuracy of the SVM used with BreakHis equaled 90%, the precision and recall almost equaled 90%, and the F1 score was approximately 90%. The GNB used with BreakHis provided an accuracy of 84%, whereas those for precision, F1, and recall were approximately 83% and 85%, respectively.

Performance evaluation of WDBC dataset

The results explored in Table 3B show supervised and DL models for identifying BCs trained on the WDBC dataset. Benign from malignant is determined based on the models and explained *via* various evaluation metrics: accuracy, sensitivity, specificity, and AUC-ROC. The WDBC dataset has ten real-valued features for each of the 569 examples characteristic of individual cell nuclei. Logistic regression achieved excellent performance

on the WDBC dataset, with a specification, recall, F1 score, and more than 97% accuracy. The convolutional neural networks achieved approximately 96% accuracy when trained on the WDBC dataset. Furthermore, the precision, recall, and F1 score percentages were approximately 95% and 96%, respectively. The support vector machine method contributed to approximately 95% accuracy. Additionally, the percentage values geometrically to the WDBC dataset contributed to the accuracy at 95% and precision, F1 score, and recall, which were approximately 95% and 96%, respectively. The WDBC dataset contributed 94% of the accuracy. The values were approximately 93%, 94%, and 94% for precision, F1 score, and recall, respectively.

Comparative performance analysis of machine learning models for breast cancer detection

The comparisons of the WDBC and BreakHis datasets enable us to gain detailed knowledge of the factors affecting the performance of the models in BC detection as per methods of earlier researchers ([Martinez & van Dongen, 2023](#)). The WDBC dataset, which uses a 70:30 train-test split, and the BreakHis dataset, which has a 30:70 split, demonstrate high accuracy. Nevertheless, their performance diverges when examined through detailed metrics, as the recorded results are mentioned in [Table 4](#). The BreakHis dataset has high sensitivity (CNN: 92.0%) and specificity (CNN: 94.1%), likely due to the superior quality of its histopathology images. In contrast, the WDBC dataset is better regarding overall average metrics, making it ideal for training machine learning models. LR exhibits excellent diagnostic capabilities in WDBC, with an average accuracy of 95.3%, 93.4% precision, and 94.1% recall. For the same datasets, the CNN has an impressive mean accuracy of 94.1% and a precision of 92.9%, with a recall of 93.0% for differentiating between malignant and benign cases. The average accuracy of the RF model is 92.9%, maintaining a good performance standard, although it is not as high as the other models. For the BreakHis dataset, the CNN stands out as the most efficient model, with an average accuracy of 93.05%, a recall score of 92.0%, and a precision of 91.4%. LR achieves a high average accuracy of 91.25%, whereas RF achieves 90.4%. Interestingly, BreakHis had higher false-negative rates (FNR). Specifically, WDBC depicted an FNR of 5.9% in identifying positive cases than did RF at 10.7%, highlighting the importance of enhancing the detection of positive images *via* histopathology to avoid missed diagnoses. The false omission rate (FOR) is also higher in BreakHis because BreakHis has a more complex dataset structure. The predicted and observed results further support the notion of dataset optimization, as WDBC is highly suitable for training models, such as LR and CNN, to achieve higher accuracy.

The ROC-AUC, FOR, and FNR values were calculated to evaluate and assess each model's performance, and the results of the corresponding models were recorded. The values of the LR on the WDBC dataset indicated that the ROC-AUC was 96.2%, and the FNR and FOR of 5.9% confirmed the ability of the LR model to discriminate malignant from benign lesions. CNN on WDBC, where FNR = 6.2%, FOR is marginally higher than LR, but the strength of feature extraction helped in the classification accuracy (ROC-AUC 97.5%). The ROC-AUC of random forest (RF) was 94.8% for WDBC, again producing a

Table 4 Comparison of classification model performance on BreakHis and WDBC datasets.											
Dataset	Model	TPR (Recall) (%)	TNR (Specificity) (%)	FNR (%)	FPR (%)	FDR (%)	FOR (%)	F1-score (%)	Mean accuracy (%)	Mean precision (%)	Mean recall (%)
WDBC	LR	94.1	96.5	5.9	3.5	6.6	4.8	93.7	95.30	93.4	94.1
	CNN	93.0	95.2	6.3	4.8	7.1	5.2	92.6	94.10	92.9	93.0
	RF	91.8	94.0	7.2	6.0	8.2	6.0	91.4	92.90	91.8	91.8
	SVM	92.5	94.5	7.5	5.5	8.0	6.0	91.9	93.50	92.0	92.5
BreakHis	LR	90.5	92.0	9.5	8.0	10.3	7.8	90.0	91.25	89.7	90.5
	CNN	92.0	94.1	8.3	5.9	8.6	7.0	91.6	93.05	91.4	92.0
	RF	89.3	91.5	10.7	8.5	11.4	8.5	88.9	90.40	88.6	89.3
	SVM	91.0	92.5	9.0	7.5	9.0	7.5	89.8	91.80	90.5	91.0

Notes:

TPR (Recall): True Positives TP True Positives TP False Negatives FN.

TNR (Specificity): True Negatives (TN)/True Negatives (TN) + False Postivie (FP).

FNR: Faslse Negatives (FN)/True Positive (TP) + False Negative (FN)

FPR: FaslsePositive (FP)/Faslse Positive (FP) + Ture Negative (TN).

FDR: Faslse Positive (FP)/False Positive (FP) + True Positive (TP).

FOR: Faslse Negative (FN)/False Negative (FN) + True Negative (TN).

F1-Score: Harmonic mean of Precision and Recall: 2. Precision. Recall/Precision + Recall

moderate FOR and an FNR slightly less effective at reducing FNR than LR and CNN but still satisfactory (FNR = 10.7%). The BreakHis dataset was analyzed by a CNN and obtained an ROC-AUC of 95.9%, a sensitivity of 92.0%, and a specificity of 94.1%. However, it has a relatively high but still quite significant FNR of 8.0% owing to the intricacy of the histopathological images. For BreakHis, the ROC-AUC was 93.8% (the false-negative recall of this model was 9.4% in the LR model); however, because this model misclassified many negative cases, its false-positive recall was higher than that of WDBC. RF on BreakHis achieved a ROC-AUC of 92.3%, an FNR of 11.2%, and the highest FOR among the traditional models, indicating the complexity of histopathological data and the difficulty in traditional tree-based methods for classification. The results demonstrate the necessity for dataset-specific model optimizations, where WDBC outperforms both the LR and the CNN because of the dataset’s structured feature distribution. In contrast, the BreakHis dataset needs additional preprocessing, feature selection, and hybrid modeling approaches, which help lower the FOR and FNR while increasing the classification accuracy.

Statistical analysis

The results of the evaluation of four different classification models—GNB, CNN, SVM, and LR—across two different datasets—BreakHis and WDBC—are presented in Table 5. In the context of the BreakHis dataset, it is essential to note that CNN and SVM performed equally well in all between-study attributes. With a p value not exceeding 0.05, no statistically significant difference between the two models can be considered present. However, compared with the GNB, the CNN yielded superior results in terms of accuracy ($p < 0.05$), precision ($p < 0.05$), F1 score ($p < 0.05$), and recall ($p < 0.05$). No statistically significant difference in performance was discovered between the CNN and LR methods in each case, with all p values exceeding 05. Thus, these two models can be considered to

Table 5 Mean performance metrics.

Comparison	Dataset	Accuracy (<i>p</i> -value)	Precision (<i>p</i> -value)	F1-Score (<i>p</i> -value)	Recall (<i>p</i> -value)	Statistically significant (<i>p</i> < 0.05)
CNN vs. SVM	BreaKHis	0.314	0.175	0.196	0.115	No
CNN vs. GNB	BreaKHis	0.026	0.017	0.024	0.024	Yes
CNN vs. LR	BreaKHis	0.067	0.073	0.065	0.070	No
CNN vs. SVM	WDBC	0.059	0.067	0.062	0.067	No
CNN vs. GNB	WDBC	0.025	0.030	0.040	0.025	Yes
CNN vs. LR	WDBC	0.063	0.078	0.068	0.075	No

function identically. When the transition to the WDBC dataset is considered, it is clear that the CNN outperforms GNB in all- $p < 0.05$ cases, which is consistent with the outcomes that have been viewed in the context of the BreaKHis dataset. Like the findings mentioned above for the BreaKHis dataset, no statistically significant difference exists between the CNN and the SVM or LR; all $p > 0.05$ can be seen. As a result, it may be concluded that accurate classification models are required in the given situation to be applied, especially in medicine, since the accuracy of classification is critical for planning appropriate treatments and making diagnostic decisions.

STUDY STRENGTH

This work makes a significant contribution by employing multiple datasets and advanced machine-learning methods to enhance the detection of BC. To overcome these limitations and make the results more diverse, we use both the WDBC dataset and the BreaKHis dataset. The reliability of the study is further enhanced by using classical supervised learning methods that include CNNs, RFs, SVMs, and LR. These models were chosen because the same option was used to compare model performance based on the data type. The performance metrics were used to assess the diagnostic capabilities of the models: AUC-ROC, accuracy, precision, sensitivity, F1, FNR, and FOR. Decreasing the FNR and FOR is highly important for increasing the accuracy of breast cancer detection. The evaluation revealed that LR provided the highest accuracy in minimizing the FNR and FOR values for the WDBC dataset. In contrast, the CNN had the highest accuracy for the BreaKHis dataset. The significance of tuning the hyperparameters through the grid search method and fivefold cross-validation was crucial for enhancing the models and boosting the prediction capability, minimizing the overfitting of the models. Additionally, RNNs and LSTMs were used to investigate the effectiveness of the proposed approaches. This research shows that recent imaging approaches, mainly CNNs, can be beneficial for dealing with breast cancer examinations and reaffirm the potential for real-time diagnostic applications.

STUDY LIMITATIONS

It is also essential to recognize some limitations inherent to this study. The BreaKHis and WDBC datasets used are dissimilar but could use a more extensive population sample with other types of cancer so that the model is more reliable and practical. This raises concerns

about model stability, as the classes dominating the dataset are those the algorithms select. Furthermore, CNNs and the remaining deep learning models are computationally demanding, especially regarding the GPU and training time, which could be problematic and expensive in the clinical environment. CNN gives the best classification results, but its interpretability is a drawback because the working mechanism of the model is not fully clear. In addition, the findings obtained in this study may be influenced by artifacts that can affect imaging instances in clinical practice since the quality of the images and the pathological condition of the patients are also important. A functional limitation of the study is the lack of genetic information and patient history in the classification analysis. Finally, more studies are needed to evaluate the efficiency of these methods in terms of the cost impact of false negatives and false positives and the time taken to diagnose the disease.

CONCLUSION

The results clearly show the superiority of the deep learning technique CNN over the machine learning techniques LR, SVM, and GNB for diagnosing breast cancer *via* the BreaKHis and WDBC datasets. The CNN model outperformed the other models by a large margin, with an accuracy of 96.21%. Moreover, the precision, recall, and F1 score values were computed to be 95.92%, 96.51%, and 96.22%, respectively, for the WDBC dataset. For the BreaKHis dataset, the CNN an accuracy of 92%, higher than those of GNB, SVM, and LR. According to the results, logistic regression was better able to minimize the FNR and FOR in the WDBC dataset, where the FNR of 5.9% and the FOR of 4.8%. In comparison, the CNN showed an FNR of 8.3% and a FOR of 7.0% in the BreaKHis dataset. These results highlight the high capacity of CNNs in analyzing histologic data for the early detection of breast cancer and their efficacy in clinical settings where an accurate diagnosis is vital. This study discusses how CNNs can be used to improve medical decision-making and still asserts that there is the possibility of performing additional studies in multimodal and hybrid frameworks for even more accurate and reliable diagnoses.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was funded by the Researchers Supporting Project (RSPD2025R758), King Saud University, Riyadh, Saudi Arabia. No additional external funding was received for this study.

Grant Disclosures

The following grant information was disclosed by the authors:
 Researchers Supporting Project: RSPD2025R758.
 King Saud University, Riyadh, Saudi Arabia.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Mandika Chetry conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Ruiling Feng performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Samra Babar analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Hao Sun performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Imran Zafar conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Mohamed Mohany conceived and designed the experiments, performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Hassan Imran Afridi performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Najeeb Ullah Khan conceived and designed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Ijaz Ali performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Muhammad Shafiq conceived and designed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Sabir Khan analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code is available at GitHub:

- <https://github.com/Imranzafer/DL-BC-Analysis>-

- Zafar, I. (2025). Early Detection and Analysis of Accurate Breast Cancer for Improved Diagnosis Using Deep Supervised Learning for Enhanced Patient Outcomes. PeerJ Computer Science. <https://doi.org/10.5281/zenodo.15032559>.

The code and raw data are available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.2784#supplemental-information>.

REFERENCES

- Abdelhafiz D, Yang C, Ammar R, Nabavi S. 2019.** Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinformatics* **20**(S11):281 DOI [10.1186/s12859-019-2823-4](https://doi.org/10.1186/s12859-019-2823-4).
- Abdullah TAA, Zahid MSM, Ali W. 2021.** A review of interpretable ml in healthcare: taxonomy, applications, challenges, and future directions. *Symmetry* **13**(12):2439 DOI [10.3390/sym13122439](https://doi.org/10.3390/sym13122439).
- Abu Abeelh E, AbuAbeileh Z. 2024.** Comparative effectiveness of mammography, ultrasound, and mri in the detection of breast carcinoma in dense breast tissue: a systematic review. *Cureus* **16**:e59054 DOI [10.7759/cureus.59054](https://doi.org/10.7759/cureus.59054).
- Abunasser BS, Al-Hiealy MRJ, Zaqout IS, Abu-Naser SS. 2023.** Convolution neural network for breast cancer detection and classification using deep learning. *Asian Pacific Journal of Cancer Prevention* **24**(2):531–544 DOI [10.31557/APJCP.2023.24.2.531](https://doi.org/10.31557/APJCP.2023.24.2.531).
- Acs B, Rantalainen M, Hartman J. 2020.** Artificial intelligence as the next step toward precision pathology. *Journal of Internal Medicine* **288**(1):62–81 DOI [10.1111/joim.13030](https://doi.org/10.1111/joim.13030).
- Adedigba AP, Adeshina SA, Aibinu AM. 2022.** Performance evaluation of deep learning models on mammogram classification using small dataset. *Bioengineering* **9**(4):161 DOI [10.3390/bioengineering9040161](https://doi.org/10.3390/bioengineering9040161).
- Adlung I, Cohen Y, Mor U, Elinav E. 2021.** Machine learning in clinical decision making. *Med* **2**(6):642–665 DOI [10.1016/j.medj.2021.04.006](https://doi.org/10.1016/j.medj.2021.04.006).
- Ahmad A. 2019.** Breast cancer statistics: recent trends. In: Ahmad A, ed. *Breast Cancer Metastasis and Drug Resistance. Advances in Experimental Medicine and Biology*. Vol. 1152. Cham: Springer DOI [10.1007/978-3-030-20301-6_1](https://doi.org/10.1007/978-3-030-20301-6_1).
- Ahmad HM, Abrar M, Izhar O, Zafar I, Rather MA, Alanazi AM, Malik A, Rauf A, Bhat MA, Wani TA, Khan AA. 2022.** Characterization of fenugreek and its natural compounds targeting AKT-1 protein in cancer: pharmacophore, virtual screening, and MD simulation techniques. *Journal of King Saud University—Science* **34**:102186 DOI [10.1016/j.jksus.2022.102186](https://doi.org/10.1016/j.jksus.2022.102186).
- Ahmed Z, Mohamed K, Zeeshan S, Dong X. 2020.** Artificial intelligence with multifunctional machine learning platform development for better healthcare and precision medicine. *Database* **2020**(104):baaa010 DOI [10.1093/database/baaa010](https://doi.org/10.1093/database/baaa010).
- Ak MF. 2020.** A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. *Healthcare* **8**(2):111 DOI [10.3390/healthcare8020111](https://doi.org/10.3390/healthcare8020111).
- Alanazi SA, Kamruzzaman MM, Islam Sarker MN, Alruwaili M, Alhwaiti Y, Alshammari N, Siddiqi MH. 2021.** Boosting breast cancer detection using convolutional neural network. *Journal of Healthcare Engineering* **2021**(6):5528622 DOI [10.1155/2021/5528622](https://doi.org/10.1155/2021/5528622).
- Aristokli N, Polycarpou I, Themistocleous SC, Sophocleous D, Mamais I. 2022.** Comparison of the diagnostic performance of magnetic resonance imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: a review. *Radiography* **28**(3):848–856 DOI [10.1016/j.radi.2022.01.006](https://doi.org/10.1016/j.radi.2022.01.006).
- Austria YD, Jay-ar PL, Maria LBS Jr, Goh JEE, Goh MLI, Vicente HN. 2019.** Comparison of machine learning algorithms in breast cancer prediction using the coimbra dataset. *International Journal of Simulation: Systems, Science & Technology* **7**(10):23–21 DOI [10.5013/IJSSST.a.20.S2.23](https://doi.org/10.5013/IJSSST.a.20.S2.23).
- Batchu S, Liu F, Amireh A, Waller J, Umair M. 2021.** A review of applications of machine learning in mammography and future challenges. *Oncology* **99**(8):483–490 DOI [10.1159/000515698](https://doi.org/10.1159/000515698).

- Bayisa LY, Wang W, Wang Q, Ukwuoma CC, Gutema HK, Endris A, Abu T. 2024. Unified deep learning model for multitask representation and transfer learning: image classification, object detection, and image captioning. *International Journal of Machine Learning and Cybernetics* 15(10):4617–4637 DOI 10.1007/s13042-024-02177-5.
- Braithwaite D, Miglioretti DL, Zhu W, Demb J, Trentham-Dietz A, Sprague B, Tice JA, Onega T, Henderson LM, Buist DS. 2018. Family history and breast cancer risk among older women in the breast cancer surveillance consortium cohort. *JAMA Internal Medicine* 178(4):494–501 DOI 10.1001/jamainternmed.2017.8642.
- Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, Cohen Z, Belgrave D, DeRubeis R, Iniesta R. 2021. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* 20(2):154–170 DOI 10.1002/wps.20882.
- Chikarmane SA, Offit LR, Giess CS. 2023. Synthetic mammography: benefits, drawbacks, and pitfalls. *RadioGraphics* 43(10):e230018 DOI 10.1148/rg.230018.
- Chugh G, Kumar S, Singh N. 2021. Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cognitive Computation* 13(6):1451–1470 DOI 10.1007/s12559-020-09813-6.
- Coffey K, Jochelson MS. 2022. Contrast-enhanced mammography in breast cancer screening. *European Journal of Radiology* 156(3):110513 DOI 10.1016/j.ejrad.2022.110513.
- Comstock CE, Gatsonis C, Newstead GM, Snyder BS, Gareen IF, Bergin JT, Rahbar H, Sung JS, Jacobs C, Harvey JA, Nicholson MH, Ward RC, Holt J, Prather A, Miller KD, Schnall MD, Kuhl CK. 2020. Comparison of abbreviated breast MRI vs digital breast tomosynthesis for breast cancer detection among women with dense breasts undergoing screening. *The Journal of the American Medical Association* 323(8):746–756 DOI 10.1001/jama.2020.0572.
- Dar RA, Rasool M, Assad A. 2022. Breast cancer detection using deep learning: datasets, methods, and challenges ahead. *Computers in Biology and Medicine* 149(1):106073 DOI 10.1016/j.compbiomed.2022.106073.
- Gautam N, Singh A, Kumar K, Aggarwal PK, Anupam. 2021. Investigation on performance analysis of support vector machine for classification of abnormal regions in medical image. *Journal of Ambient Intelligence and Humanized Computing* 10(10):3889 DOI 10.1007/s12652-021-02965-9.
- George K, Sankaran P, K PJ. 2020. Computer assisted recognition of breast cancer in biopsy images via fusion of nucleus-guided deep convolutional features. *Computer Methods and Programs in Biomedicine* 194(1):105531 DOI 10.1016/j.cmpb.2020.105531.
- Guo R, Lu G, Qin B, Fei B. 2018. Ultrasound imaging technologies for breast cancer detection and management: a review. *Ultrasound in Medicine & Biology* 44(1):37–70 DOI 10.1016/j.ultrasmedbio.2017.09.012.
- Harvey H, Heindl A, Khara G, Korkinof D, O'Neill M, Yearsley J, Karpate E, Rijken T, Kecskemethy P, Forrai G. 2019. Deep learning in breast cancer screening. In: Ranschaert ER, Morozov S, Algra PR, eds. *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*. Cham: Springer International Publishing, 187–215.
- Houssein EH, Emam MM, Ali AA, Suganthan PN. 2021. Deep and machine learning techniques for medical imaging-based breast cancer: a comprehensive review. *Expert Systems with Applications* 167(24):114161 DOI 10.1016/j.eswa.2020.114161.
- Hwang I, Woo M. 2023. Image compression and decompression framework based on latent diffusion model for breast mammography. ArXiv preprint DOI 10.48550/arXiv.2310.05299.
- Ido M, Saito M, Banno H, Ito Y, Goto M, Ando T, Kousaka J, Mouri Y, Fujii K, Imai T, Nakano S, Suzuki K, Murotani K. 2023. Clinical performance of digital breast tomosynthesis-guided

- vacuum-assisted biopsy: a single-institution experience in Japan. *BMC Medical Imaging* **23**(1):2 DOI [10.1186/s12880-022-00896-1](https://doi.org/10.1186/s12880-022-00896-1).
- Islam MM, Haque MR, Iqbal H, Hasan MM, Hasan M, Kabir MN. 2020. Breast cancer prediction: a comparative study using machine learning techniques. *SN Computer Science* **1**(5):290 DOI [10.1007/s42979-020-00305-w](https://doi.org/10.1007/s42979-020-00305-w).
- Jaglan P, Dass R, Duhan M. 2019. Breast cancer detection techniques: issues and challenges. *Journal of the Institution of Engineers (India): Series B* **100**(4):379–386 DOI [10.1007/s40031-019-00391-2](https://doi.org/10.1007/s40031-019-00391-2).
- Kalaivani S, Asha N, Gayathri A. 2023. Geometric transformations-based medical image augmentation. In: Solanki A, Naved M, eds. *GANs for Data Augmentation in Healthcare*. Cham: Springer International Publishing, 133–141.
- Karkehabadi A, Homayoun H, Sasan A. 2024. SMOOT: saliency guided mask optimized online training. In: *2024 IEEE 17th Dallas Circuits and Systems Conference (DCAS)*. Piscataway: IEEE, 1–6.
- Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Wortman Vaughan J. 2020. Interpreting interpretability: understanding data scientists’ use of interpretability tools for machine learning. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.
- Kornecki A. 2022. Current status of contrast enhanced mammography: a comprehensive review. *Canadian Association of Radiologists Journal* **73**(1):141–156 DOI [10.1177/08465371211029047](https://doi.org/10.1177/08465371211029047).
- Li Q, You T, Chen J, Zhang Y, Chenglie D. 2023. LI-EMRSQL: linking information enhanced Text2SQL parsing on complex electronic medical records. *IEEE Transactions on Reliability* **73**(2):1280–1290 DOI [10.1109/TR.2023.3336330](https://doi.org/10.1109/TR.2023.3336330).
- Liefwaard MC, Lips EH, Wesseling J, Hylton NM, Lou B, Mansi T, Pusztai L. 2021. The way of the future: personalizing treatment plans through technology. *American Society of Clinical Oncology Educational Book* **41**(41):12–23 DOI [10.1200/EDBK_320593](https://doi.org/10.1200/EDBK_320593).
- Liu Y, Tong Y. 2023. Clinicopathological characteristics and prognosis of triple-negative breast cancer invasive ductal carcinoma with ductal carcinoma in situ. *Journal of Cancer Research and Clinical Oncology* **149**:11181–11191 DOI [10.1007/s00432-023-04895-9](https://doi.org/10.1007/s00432-023-04895-9).
- Madani M, Behzadi MM, Nabavi S. 2022. The role of deep learning in advancing breast cancer detection using different imaging modalities: a systematic review. *Cancers* **14**(21):5334 DOI [10.3390/cancers14215334](https://doi.org/10.3390/cancers14215334).
- Marra A, Trapani D, Viale G, Criscitiello C, Curigliano G. 2020. Practical classification of triple-negative breast cancer: intratumoral heterogeneity, mechanisms of drug resistance, and novel therapies. *NPJ Breast Cancer* **6**(1):54 DOI [10.1038/s41523-020-00197-2](https://doi.org/10.1038/s41523-020-00197-2).
- Martinez RG, van Dongen D-M. 2023. Deep learning algorithms for the early detection of breast cancer: a comparative study with traditional machine learning. *Informatics in Medicine Unlocked* **41**:101317 DOI [10.1016/j.imu.2023.101317](https://doi.org/10.1016/j.imu.2023.101317).
- Mazhar T, Haq I, Ditta A, Mohsan SAH, Rehman F, Zafar I, Gansau JA, Goh LPW. 2023. The role of machine learning and deep learning approaches for the detection of skin cancer. *Healthcare: MDPI* **11**(3):415 DOI [10.3390/healthcare11030415](https://doi.org/10.3390/healthcare11030415).
- Moffitt JR, Lundberg E, Heyn H. 2022. The emerging landscape of spatial profiling technologies. *Nature Reviews Genetics* **23**(12):741–759 DOI [10.1038/s41576-022-00515-3](https://doi.org/10.1038/s41576-022-00515-3).
- Naeem M, Hayat M, Qamar SA, Mehmood T, Munir A, Ahmad G, Hussain A. 2019. Risk factors, genetic mutations and prevention of breast cancer. *International Journal of Biosciences* **14**:492–496.
- Nelson HD, Cantor A, Wagner J, Jungbauer R, Fu R, Kondo K, Stillman L, Quiñones A. 2020. Effectiveness of patient navigation to increase cancer screening in populations adversely affected

- by health disparities: a meta-analysis. *Journal of General Internal Medicine* **35**(10):3026–3035 DOI [10.1007/s11606-020-06020-9](https://doi.org/10.1007/s11606-020-06020-9).
- Ngiam KY, Khor W. 2019. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology* **20**(5):e262–e273 DOI [10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4).
- Noor A, Zhao Y, Khan R, Wu L, Abdalla FYO. 2020. Median filters combined with denoising convolutional neural network for Gaussian and impulse noises. *Multimedia Tools and Applications* **79**(25–26):18553–18568 DOI [10.1007/s11042-020-08657-4](https://doi.org/10.1007/s11042-020-08657-4).
- O’Leary B, Cutts RJ, Liu Y, Hrebien S, Huang X, Fenwick K, André F, Loibl S, Loi S, Garcia-Murillas I. 2018. The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 trial. *Cancer Discovery* **8**(11):1390–1403 DOI [10.1158/2159-8290.CD-18-0264](https://doi.org/10.1158/2159-8290.CD-18-0264).
- Osarogiagbon AU, Khan F, Venkatesan R, Gillard P. 2021. Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations. *Process Safety and Environmental Protection* **147**(4):367–384 DOI [10.1016/j.psep.2020.09.038](https://doi.org/10.1016/j.psep.2020.09.038).
- Oyeniya J, Oluwaseyi P. 2024. Emerging trends in AI-powered medical imaging: enhancing diagnostic accuracy and treatment decisions. *International Journal of Enhanced Research in Science, Technology & Engineering* **13**(4):81–94.
- Pandey SK, Janghel RR. 2019. Recent deep learning techniques, challenges and its applications for medical healthcare system: a review. *Neural Processing Letters* **50**(2):1907–1935 DOI [10.1007/s11063-018-09976-2](https://doi.org/10.1007/s11063-018-09976-2).
- Pawlak ME, Rudnicki W, Borkowska A, Skubisz K, Rydzik R, Łuczyńska E. 2023. Comparative analysis of diagnostic performance of automatic breast ultrasound, full-field digital mammography and contrast-enhanced mammography in relation to breast composition. *Biomedicines* **11**(12):3226 DOI [10.3390/biomedicines11123226](https://doi.org/10.3390/biomedicines11123226).
- Pei X, hong Zhao Y, Chen L, Guo Q, Duan Z, Pan Y, Hou H. 2023. Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences. *Materials & Design* **232**(7):112086 DOI [10.1016/j.matdes.2023.112086](https://doi.org/10.1016/j.matdes.2023.112086).
- Phi X-A, Tagliafico A, Houssami N, Greuter MJW, de Bock GH. 2018. Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts—a systematic review and meta-analysis. *BMC Cancer* **18**(1):380 DOI [10.1186/s12885-018-4263-3](https://doi.org/10.1186/s12885-018-4263-3).
- Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. 2018. Precision medicine and radiogenomics in breast cancer: new approaches toward diagnosis and treatment. *Radiology* **287**(3):732–747 DOI [10.1148/radiol.2018172171](https://doi.org/10.1148/radiol.2018172171).
- Potnis KC, Ross JS, Aneja S, Gross CP, Richman IB. 2022. Artificial intelligence in breast cancer screening: evaluation of FDA device regulation and future recommendations. *JAMA Internal Medicine* **182**(12):1306–1312 DOI [10.1001/jamainternmed.2022.4969](https://doi.org/10.1001/jamainternmed.2022.4969).
- Sahu A, Qazi S, Raza K, Singh A, Verma S. 2022. Machine learning-based approach for early diagnosis of breast cancer using biomarkers and gene expression profiles. In: *Computational Intelligence in Oncology: Applications in Diagnosis, Prognosis and Therapeutics of Cancers*. Cham: Springer, 285–306.
- Salod Z, Singh Y. 2019. Comparison of the performance of machine learning algorithms in breast cancer screening and detection: a protocol. *Journal of Public Health Research* **8**(3):jp19.1677 DOI [10.4081/jp19.1677](https://doi.org/10.4081/jp19.1677).
- Sarker IH. 2021. Machine learning: algorithms, real-world applications and research directions. *SN Computer Science* **2**(3):160 DOI [10.1007/s42979-021-00592-x](https://doi.org/10.1007/s42979-021-00592-x).

- Sechopoulos I, Teuwen J, Mann R. 2021.** Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: state of the art. *Seminars in Cancer Biology: Elsevier* 72:214–225 DOI 10.1016/j.semcancer.2020.06.002.
- Shaikh K, Krishnan S, Thanki RM. 2021.** *Artificial intelligence in breast cancer early detection and diagnosis*. Cham: Springer.
- Siegel RL, Miller KD, Jemal A. 2018.** Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* 68(1):7–30 DOI 10.3322/caac.21442.
- Taghizadeh E, Heydarheydari S, Saberi A, JafarpourNesheli S, Rezaei SM. 2022.** Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods. *BMC Bioinformatics* 23(1):410 DOI 10.1186/s12859-022-04965-8.
- Tang L, Zhang W, Chen L. 2024.** Brain radiotherapy combined with targeted therapy for HER2-positive breast cancer patients with brain metastases. *Breast Cancer (Dove Med Press)* 16:379–392 DOI 10.2147/BCTT.S460856.
- Testa U, Castelli G, Pelosi E. 2020.** Breast cancer: a molecularly heterogenous disease needing subtype-specific treatments. *Medical Sciences* 8(1):18 DOI 10.3390/medsci8010018.
- Thakur N, Kumar P, Kumar A. 2024.** A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities. *Multimedia Tools and Applications* 83(12):35849–35942 DOI 10.1007/s11042-023-16634-w.
- Van Baelen K, Van Cauwenberge J, Maetens M, Beck G, Camden A, Chase M-C, Fraser V, Freeney S, Hutcheson L, Levine JK. 2024.** Reporting on invasive lobular breast cancer in clinical trials: a systematic review. *Nature Partner Journals Breast Cancer* 10(1):23 DOI 10.1038/s41523-024-00627-5.
- Walsh R, Tardy M. 2023.** A comparison of techniques for class imbalance in deep learning classification of breast cancer. *Diagnostics* 13(1):67 DOI 10.3390/diagnostics13010067.
- Wilkinson L, Gathani T. 2022.** Understanding breast cancer as a global health concern. *The British Journal of Radiology* 95(1130):20211033 DOI 10.1259/bjr.20211033.
- Wu J, Hicks C. 2021.** Breast cancer type classification using machine learning. *Journal of Personalized Medicine* 11(2):61 DOI 10.3390/jpm11020061.
- Zafar I, Anwar S, Yousaf W, Nisa FU, Kausar T, ul Ain Q, Unar A, Kamal MA, Rashid S, Khan. 2023.** Reviewing methods of deep learning for intelligent healthcare systems in genomics and biomedicine. *Biomedical Signal Processing and Control* 86:105263 DOI 10.1016/j.bspc.2023.105263.
- Zhang J, Chen L. 2019.** Clustering-based undersampling with random over sampling examples and support vector machine for imbalanced classification of breast cancer diagnosis. *Computer Assisted Surgery* 24(sup2):62–72 DOI 10.1080/24699322.2019.1649074.