Sampson *et al.*
Journal of Electrical Systems and Information Technology (2026) 13:9
https://doi.org/10.1186/s43067-025-00313-x

Journal of Electrical Systems and Information Technology

RESEARCH **Open Access**

Optimizing breast cancer prediction through stacking ensemble machine learning models: a comparative analysis

Aeron Sampson[1], Akini James[2*] and Vrijesh Tripathi[3]

*Correspondence:
Akini James
akini.james@uwi.edu
[1]Department of Computing and Information Technology, University of the West Indies, St. Augustine, Trinidad and Tobago
[2]Department of Clinical Surgical Sciences, Faculty of Medical Sciences, University of the West Indies, St. Augustine, Trinidad and Tobago
[3]Department of Mathematics and Statistics, University of the West Indies, St. Augustine, Trinidad and Tobago

Abstract

Aim To develop and identify an optimal stacking ensemble model for predicting breast cancer, using combinations of Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbours (KNN) as base models with a rotating meta-classifier.

Method An open-source breast cancer dataset of 569 patients (357 benign, 212 malignant) from UCI Machine Learning Repository was analysed. Ten predictive cell nucleus features were selected, and all other irrelevant variables were excluded. Exploratory Data Analysis included detection of outliers (addressed via Winsorization), assessment of normality (square root transformation applied), and correlation analysis to identify multicollinearity. Independent sample t-tests evaluated differences in features by diagnosis. Multicollinear features were assessed using Binary Logistic Regression, retaining the features with the highest Pseudo for modelling. Three stacking models were constructed using combinations of SVM, Naïve Bayes, KNN as base and meta-classifiers. Models were evaluated using 10-fold cross-validation with performance metrics including accuracy, precision, recall, F1-score, and ROC AUC. All analyses were conducted using Python and R.

Results Significant differences were found between malignant and benign cases for all features ($p < 0.001$), except for fractal dimension ($p = 0.98$), which was then excluded from the analysis. Multicollinearity was observed among five features, and "area" was retained for modelling as it demonstrated the strongest predictive power for diagnosis, with a Pseudo R^2 of 81%. Model 2 with Naïve Bayes and KNN as the base models and SVM as the meta-model achieved the best performance (95% accuracy, 90% recall, and 93% F1-score). The ROC-AUC analysis showed strong predictive ability, with an average AUC of 0.97 across 10-fold cross-validation.

Conclusion The stacking ensemble model, integrating SVM, Naïve Bayes, and KNN, achieved improved accuracy and robustness in breast cancer prediction with Model 2 performing the best. This approach demonstrates potential for enhancing early detection and reducing breast cancer mortality. Its application in broader clinical and diverse healthcare settings may further advance disease prediction efforts.

Keywords Breast cancer, Machine learning, Stacking, Ensemble learning, Early diagnosis, Classification

Introduction

Breast cancer ranks as the primary cause of cancer-related mortality and is the most commonly identified cancer among women, particularly in the Americas [75]. In 2020, Latin America and the Caribbean reported over 210,000 new breast cancer diagnoses and nearly 68,000 fatalities [75]. According to the World Health Organization (WHO), breast cancer had a global impact in 2022, affecting over 157 countries and causing an estimated 670,000 deaths worldwide [96]. Despite breast cancer primarily affecting women, the statistics mentioned encompass both female and male patients. On average, approximately 1 in 726 men will experience breast cancer development during their lifetime in 2024 [9]. However, this remains significantly lower than the average likelihood of breast cancer in women, which stands at approximately 1 in 8 [8]. These statistics highlight the persistent burden of breast cancer and underscore the necessity for improved predictive models and treatment strategies aimed at reducing mortality rates.

Despite advancements in screening and treatment, current methods for risk prediction and treatment personalization often lack the precision needed for optimal patient outcomes. Traditional diagnostic techniques rely on singular data sources, limiting their effectiveness. Emerging research indicates that integrating multiple data types—genetic, clinical, and imaging—can enhance risk assessment and treatment planning. Machine learning (ML), a branch of artificial intelligence (AI) that enables computers to learn patterns from data, has shown substantial promise in improving diagnostic accuracy. In particular, ensemble learning methods, which com combine multiple predictive algorithms, have demonstrated enhanced performance over individual models.

Existing breast cancer prediction models have employed a range of ML techniques such as Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbours (KNN), Multilayer Perceptron (MLP), Gradient Boosting Trees (GBT), and Genetic Algorithms (GA), with reported accuracies typically ranging from 80% to 97% across studies [21, 55, 64]. Rabiei et al. demonstrated that RF outperformed several classifiers achieving an 80% accuracy, 95% sensitivity, 80% specificity and 0.56 area under the curve (AUC) [80]. Although ML models are increasingly being integrated into various aspects of society such as healthcare for decision making processes there is an emerging concern of transparency and interpretability. These models risk overfitting, a situation in which the model learns the noise or irrelevant patterns in the training data, reducing its ability to generalize to new datasets. Therefore, the model may only perform well on the current data but lack generalizability for other data sets. While regularization techniques compensate for overfitting, ensemble methods such as bagging (e.g. Random Forest) and boosting (e.g. AdBoost, XGBoost) often provide more robust and stable predictive performance across diverse datasets.

Furthermore, deep learning models such as Convolutional Neural Networks (CNNs), Deep Neural Networks (DNNs), and Autoencoders have demonstrated state-of-the-art performance in breast cancer classification, particularly in image-based diagnosis. However, due to their high data requirements and tendency to overfit on small-structured datasets, ensemble-based machine learning strategies often outperform deep learning approaches in tabular prediction tasks. The Wisconsin dataset used in this study is relatively small and structured, making deep learning models prone to overfitting and performance instability. In addition, these models often offer limited interpretability which is an important limitation in clinical decision-making where transparency and explainability are essential.

In contrast, stacking ensemble learning, a method that combines the outputs of multiple base models using a secondary model known as a meta-learner, provides a balance between predictive accuracy and interpretable. In this approach, level 0 base models generate predictions, which are then used as input features for level 1 meta-model that learns how to optimally integrate them. This structure enhances robustness, reduces model variance, and maintains insight into model contributions while requiring less computational complexity than deep learning methods.

Building on these advancements, this study focuses on developing an optimal stacking ensemble model for breast cancer prediction as a balanced solution that emphasizes both performance and interpretability. By using combinations of three widely used ML classifiers, Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbours (KNN), with two models serving as base learners (level 0) and the remaining model acting as the meta-learner (level 1). This design aims to improve early detection, classification accuracy and risk stratification.

Leveraging these advanced computational techniques has the potential to significantly improve patient outcomes and reduce mortality rates. The proposed stacking ensemble model for breast cancer prediction can be integrated into existing clinical workflows as a decision-support tool. Within Electronic Health Record systems, it could automatically process routinely captured cytological measurement which offers a non-invasive approach for predicting the malignancy of cancer cells in patients as well as histopathological measurements to generate a malignancy risk score. This would assist clinicians in predicting cases, highlighting those that may require urgent review or additional diagnostic testing. Such an approach is particularly useful in high-volume screening settings or facilities with limited pathology staff, where early case prioritization is essential.

The model could also be deployed through mobile or web-based health platforms, allowing remote access in telepathology or outreach programs. Healthcare providers could enter feature values obtained from laboratory reports and receive an instant assessment to support preliminary decision-making or second-opinion review. With appropriate validation and oversight, this type of integration could contribute to earlier detection, streamlined workflows, and more consistent diagnostic support across varying healthcare settings.

Literature review

To comprehend the worldwide effects of breast cancer, a foundational understanding of the disease is essential. Breast cancer is a disease that develops in the cells of the breast, most often beginning in the ducts or lobules [14, 29, 31]. Numerous factors, including genetic mutations, hormonal influences, lifestyle choices, and environmental exposures, can significantly increase a person's cancer risk. The disease can be benign (non-cancerous) or malignant (cancerous). Malignant tumours pose a risk as they can invade surrounding tissues and metastasize to distant body parts [15]. Breast cancer is categorized into different stages, ranging from Stage 0, characterized by non-invasive cancer, to Stage IV, where the cancer has metastasized beyond the breast and nearby lymph nodes [18]. While most breast lumps are benign and not life-threatening, it is crucial to evaluate any breast changes to determine if they are malignant. Timely and precise diagnosis plays a vital role in enhancing patient outcomes [81].

Early breast cancer detection is much more difficult for men than women due to anatomical and social factors [17]. Men have less breast tissue, making small masses easier to detect and causing the cancer to spread more quickly but the rarity of breast cancer in men leads to less awareness and later diagnoses. Therefore, while routine screening is common for women, it's not typically recommended for men unless they have a high risk, such as a family history or breast cancer gene mutations [17]. Given these complexities and more, understanding the various methods used for cancer prediction and diagnosis becomes key in tailoring early detection strategies to prevent breast cancer fatalities [10].

Breast cancer prediction and diagnosis involve a variety of methods that can be categorized into three primary groups: physical examination, biopsies, and imaging tests [20, 61]. These traditional methods can be enhanced by modern approaches such as omics technologies along with the adoption of artificial intelligence (AI) and machine learning method, which provide deeper insights and enhance accuracy in diagnosis and treatment planning [3, 25, 67, 82].

Physical examination

A physical examination for breast cancer diagnosis involves a comprehensive and manual assessment of the breasts to detect abnormal signs, such as changes in size, skin dimpling, nipple abnormalities, and lumps. A clinical breast exam is a physical assessment conducted by a healthcare professional to detect early signs of breast cancer, examining the breasts, underarms, and collarbone area for lumps, texture changes, or other abnormalities, crucial for early detection and prevention [54].

Biopsies

A breast biopsy is a method that uses a medical procedure, such as surgery or using a needle, to obtain a sample of breast tissue for examination. A doctor extracts a small tissue or tumour sample from the suspicious area, which is then sent to a laboratory to determine if it contains cancer cells [16, 72].

Imaging test

An imaging test creates detailed pictures of the of the interior of your body. Imaging tests use various forms of energy to interact with your body's tissues, creating images that help visualize internal structures [19]. Techniques such as mammography, ultrasound and Magnetic Resonance Imaging (MRI), are all categorized as imaging tests.

Mammography employs low-dose X-rays to examine breast tissue and find early indications of breast cancer, typically in the form of tumours, years before symptoms appear [7, 41, 71, 73]. Three-dimensional (3-D) mammography, also known as breast tomosynthesis, is an advanced version of the normally two-dimensional (2-D) mammography that creates a 3-D breast image using multiple angles. Its effectiveness in early detection compared to 2-D mammography remains uncertain [41, 69].

Breast ultrasound is a non-invasive imaging procedure that uses high-energy sound waves to create detailed sonograms of the breast's internal structures, using a gel-applied transducer to capture echoes [12, 32]. This technique is effective in detecting structural changes in the breast tissue and structure, such as fluid-filled cysts, and can even distinguish between fluid-filled masses and solid masses that may require further investigation [12, 32].

Breast MRI relies on the combination of powerful magnets and radio waves to produce detailed images [11, 70]. The specialized exam uses MRI equipment, a breast coil, and contrast dye injection to improve image clarity before the exam [11].

Omics

Omics technology uses high-throughput techniques to study biological molecules within an organism, providing a holistic understanding of biological systems. Key areas of omics research include genomics, transcriptomics, proteomics, and metabolomics [65].

Genetic testing is a subset of omics methods, focusing specifically on the analysis of the genome. This method is used to detect gene mutations in breast cancer patients. It involves analysing blood, saliva, or cheek swabs to detect specific or broad mutations based on the individual's background or family history [13, 68]. Results can be positive, negative, inconclusive, or variant of unknown significance (VUS), with positive results suggesting risk reduction steps, negative results indicating other mutations, and inconclusive results providing no definitive answer [13, 68].

Integration of artificial intelligence and machine learning

The integration of Artificial Intelligence (AI) and Machine Learning into image testing, particularly into mammography, in order to enhance breast cancer detection has been a subject of significant research [25]. AI was used in conjunction with human radiologists, highlighting its potential to revolutionize breast cancer detection by increasing sensitivity and specificity, ultimately reducing false positives and missed cancers [34]. Another study conducted a randomized, controlled trial evaluating the effectiveness and safety of AI-assisted mammography screening in comparison to conventional double reading by radiologists. AI-assisted screening detected slightly more cancers (6.1 per 1000 participants) compared to the control group (5.1 per 1000 participants) reducing the radiologists' screen-reading workload by 44.3%, while maintaining similar recall rates and false-positive rates. This study found AI integration in mammography screening to be safe, potentially improving screening efficiency without compromising accuracy [56]. These studies show the potential of AI and machine learning effectively measuring variables such as tumour diameter, perimeter, area, and shape metrics, which are critical for assessing tumour size, potential invasiveness and stage [35, 90].

Machine learning in the breast cancer field

Building upon the advancements in AI and machine learning, recent years have witnessed the emergence of powerful tools for medical diagnostics. Machine learning, in particular, has exhibited significant potential in improving the precision and effectiveness of disease detection across various medical fields [1, 87]. Machine learning algorithms can analyse enormous quantity of medical data, detecting patterns that may not be discernible to human clinicians [30]. This capability holds particular importance in the field of breast cancer, where early detection and accurate diagnosis significantly impacts patient outcomes [96].

Ahsan et al. (2022) [4] further elaborates on the groundbreaking potential of machine learning in breast cancer diagnosis by highlighting various techniques used to develop models capable of classifying breast cancer with notable accuracy such as Support Vector Machines (SVM), Fuzzy Logic, K-Means Clustering, and Neural Networks. Fuzzy-Logic-based systems can replicate the nuanced decision-making processes of radiologists by achieving accuracy rates around 83.34% while hybrid models combining k-means clustering and SVM have reached accuracies up to 97.38%, showcasing the potential of these techniques in handling complex datasets [4].

Naji et al. (2021) [64] also explored the potential of detecting breast cancer by applying five machine learning algorithms. Utilizing the numerical Wisconsin (Diagnostic) Breast Cancer dataset, their findings highlighted the SVM as particularly effective, attaining the greatest accuracy in breast cancer prediction and surpassing the other algorithms in terms of precision and low error rates. Additionally, the study emphasized the significance of preprocessing and feature selection procedures in enhancing the model's performance, highlighting that the careful selection of relevant features can significantly improve the predictive capabilities of machine learning models in medical diagnostics [64].

On platforms like Kaggle, users have extensively explored the Breast Cancer Wisconsin (Diagnostic) dataset, employing diverse strategies to optimize model performance. These approaches often involve the use of a variety of machine learning algorithms, coupled with advanced optimization techniques such as hyperparameter tuning, to achieve high accuracies and balanced precision-recall scores. The high-performance metrics demonstrated by Kaggle users in their analyses of numerical breast cancer datasets make their models a valuable point of comparison for this study. By analysing these Kaggle models, insights can be gained regarding the effectiveness and generalizability of various machine learning techniques for diagnosing breast cancer.

Deep convolutional neural networks (DCNs) also stood out in addressing image-based datasets, as demonstrated by Geras et al. (2017) [40]. DCNs were utilized for mammography screening, addressing challenges like low specificity and high false positive rates. They modified DCNs for medical image analysis and introduced multi-view deep convolutional networks. This approach analysed mammographic images from multiple views, such as craniocaudal and mediolateral oblique, to provide a comprehensive assessment of breast tissue. This multi-view technique improves detection rates and optimizes the screening process, maintaining critical details without loss of information from downsampling, thereby enhancing the accuracy and reliability of mammography screening [40].

Ahsan et al. (2022) [4] also discussed the role of deep learning (DL) models, such as long short-term memory (LSTM) and gated recurrent unit (GRU) networks, which have demonstrated remarkable accuracy rates approaching 99% on well-known datasets like the Wisconsin (Diagnostic) Breast Cancer dataset. While these advancements underscore the significant capabilities of machine learning algorithms in augmenting diagnostic accuracy as well as efficiency, the review also notes challenges such as data imbalance and the need for model interpretability. Addressing these issues is necessary to guarantee that machine learning algorithm models in healthcare settings are reliable and generalizable. Overall, the findings suggest that continued research and refinement of machine learning algorithms hold promise for revolutionizing breast cancer diagnosis and treatment [4].

In contrast to the emphasis placed on deep learning architectures such as LSTM and GRU, this study employs a stacking ensemble approach, integrating various classifiers to enhance accuracy and robustness. This method allows for the leveraging of diverse machine learning models, which not only improves predictive performance but also provides a different avenue for optimizing breast cancer diagnosis. By using stacking techniques, this study aims to offer a more interpretable and adaptable solution in comparison to deep learning approaches, making it more applicable to clinical settings.

Machine learning algorithms and stacking ensemble method

In the realm of machine learning each model often exhibit its own strengths and weaknesses which depend on how the model was designed and the characteristics of the data it is applied to [51]. To take advantage of different models' strengths, ensemble methods combine several base models to create a more accurate and dependable prediction model. One common technique is stacking, where multiple base models are trained, and their predictions are then combined by a meta-model to make a more accurate final decision.

Stacking, unlike other ensemble methods that typically build variations of a single base model, harnesses the complementary strengths of diverse models to enhance predictive performance [62]. As shown in Fig. 1, stacking involves two levels of modelling. At the first level, multiple base models, known as level-0 models, are trained on the original training set. To ensure diversity and maximize the potential benefits of stacking, a variety of machine learning models are used as base models. The base models' predictions, referred to as level-0 predictions, are then compiled and used as input for the meta model at the second level. The meta model, also called the level-1 model, can be any machine learning algorithm and is trained on the base models' predictions. It learns how to combine these predictions optimally to output the final prediction [62]. Figure 1 depicts a schematic showing Level 0 base classifiers generating prediction outputs that are combined and used by the Level 1 meta-classifier to produce the final prediction.

Mahajan et al. (2023) [58] performed a systematic review showing that the application of hybrid and stacking ensemble algorithms in medical diagnosis improves predictive

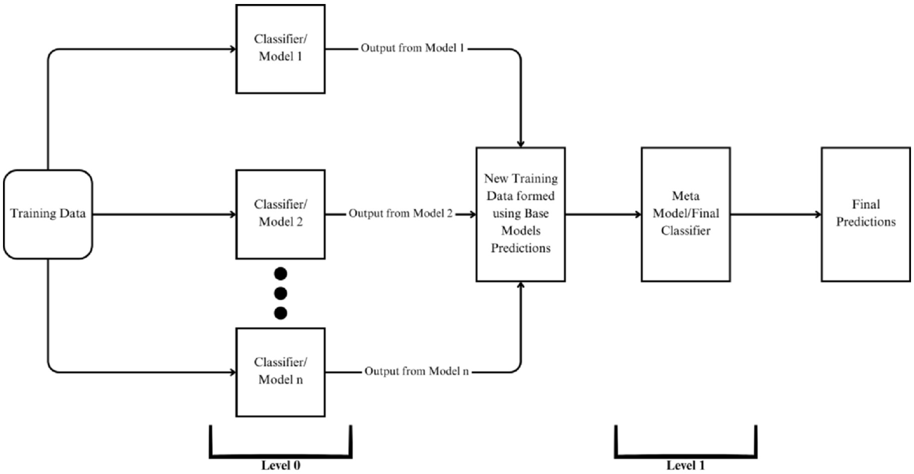


Fig. 1 Architecture of a stacking ensemble learning model

accuracy across various diseases. For instance, studies on heart disease and skin cancer prediction have reported high accuracy rates, with stacking achieving up to 100% accuracy in certain cases. Additionally, the method has proven effective in diagnosing chronic conditions such as kidney disease, diabetes, and liver disease, with one study reporting a 97.1% accuracy rate for chronic kidney disease diagnosis using a stacking approach [58].

Comparative analyses of ensemble methods indicate that stacking, though less frequently used than bagging and boosting, consistently delivers superior accuracy. This method has been found to be the most accurate in 82.6% of reviewed cases, underscoring its robustness in medical diagnostics. These results underscore the potential of stacking ensemble methods to enhance early diagnosis and treatment, thereby improving patient outcomes across a range of medical conditions [58].

This approach often yielded superior performance compared to the individual base models. Its versatility allowed it to tackle a wide range of problems, and its flexibility accommodated various base model types. However, overfitting was a problem with stacking as there are many predictors predicting the same target, and those predictions are amalgamated into an input for the meta-model [62]. To overcome this, cross-validation method is used by partitioning the data into k-folds, training and evaluating the model on each fold, and then iteratively evaluating the model on the remaining fold. This method helps to combat overfitting by offering a more reliable estimate of model performance through averaging performance metrics across iterations, reducing variance in model evaluation caused by random data splits [26].

The stacking ensemble used in this current study incorporates Support Vector Machines (SVM), Naive Bayes, and K-Nearest Neighbours (KNN), with two used as base models and one as the meta-model.

Support Vector Machine (SVM) is a supervised learning model that classifies data by finding an optimal hyperplane and optimizing the separation between classes. It performs well in high dimensions and handles complex data but can be computationally intensive and sensitive to tuning [46]. Naive Bayes operates as a probabilistic classifier grounded in Bayes' theorem, assuming independence among features. It efficiently calculates class probabilities but may struggle with correlated features and imbalanced datasets [45, 79]. K-Nearest Neighbours (KNN) assigns class labels to data by considering the majority class among its *k* closest neighbours. This algorithm is straightforward and generally effective, yet its performance can be hindered by large datasets, noise, and the selection of *k*. Additionally, it tends to struggle with high-dimensional data [74].

These models were selected based on the results of a comparative study of various machine learning algorithms for classification tasks conducted by Sheth et al. (2022) [86], which identifies them as the best performing models. Sheth et al. (2022) [86] analysed the performance of four classification models: Decision Tree, SVM, Naive Bayes, and KNN, across five different datasets and found that SVM, Naive Bayes, and KNN were the most effective algorithms, demonstrating the ability to handle diverse data types. This result makes them promising candidates for integration into the stacking model used in this paper, potentially improving diagnostic accuracy in breast cancer detection.

Additionally, while Sheth et al. (2022) [86] identified SVM, Naive Bayes, and KNN as top performers across multiple datasets, it's essential to validate their effectiveness specifically for the Breast Cancer Wisconsin (Diagnostic) dataset. Prior studies have demonstrated the efficacy of these models individually on this dataset. For instance, Asri et al. (2016) [21] utilized SVM, Naive Bayes, and KNN, achieving high accuracy in breast cancer classification. Similarly, Das et al. (2023) [33] employed these three models, reinforcing their applicability in this domain. Islam et al. (2017) [48] focused on SVM and KNN, while Sharma et al. (2018) [85] used Naive Bayes and KNN, both reporting strong performance metrics. However, these studies did not explore the potential benefits of combining these models through stacking. By integrating SVM, Naive Bayes, and KNN into a stacking ensemble, we aim to leverage their individual strengths and enhance diagnostic accuracy beyond what each model can achieve alone.

Materials and methods

The dataset utilized in this study is open-source and was retrieved from the University of California, Irvine (UCI) Machine Learning Repository [99, 100]. It contains 32 features of cell nuclei characteristics in a digitized image of a fine needle aspirate of a breast mass for 569 patients from the University of Wisconsin, Madison, Department of Surgery and Human Oncology. The target variable which contains 357 benign and 212 malignant cases (diagnosis), and the 10 potential predictors (features) used in this study are presented in Table 1. All other variables that do not contribute to the study's objective were excluded from the analysis.

It is important to note that the dataset does not include demographic or clinical variables such as patient age, ethnicity, or treatment history. As acknowledge in the limitation section, this constrains the generalizability of the study's findings to broader or more diverse populations. However, the dataset remains widely used in the literature for benchmarking machine learning models in cancer classification due to its quality, consistency, and diagnostic relevance.

Since the dataset originates from a single medical institution, there is potential for sampling bias and limited population diversity. This could affect the external validity of the model if applied to different clinical or demographic contexts. To minimize these effects, the analysis focused exclusively on cellular features rather than population-specific variables, ensuring that the model's performance depended solely on morphological data. Additionally, cross-validation (10-fold) was employed to ensure that the models were evaluated on multiple data partitions, reducing the likelihood of overfitting to specific patterns within the dataset.

Exploratory Data Analysis (EDA) was conducted such as descriptive analysis and data visualization. Duplicated records and missing values were checked for; the dataset contained no missing values. Using the Inter-Quartile Range (IQR) approach, outliers were identified and addressed through the Winsorization method to enhance model accuracy, reduce noise and prevent overfitting.

All assumptions were checked and the relevant transformation was applied to meet the assumption of normality. Independent samples t-tests were conducted on each feature to evaluate significant differences in mean values between malignant and benign cases. Features not demonstrating significant differences were excluded in the modelling process since they do not contribute significantly to the diagnosis.

A heatmap was generated to examine potential correlations between the dependent features, assessing multicollinearity issues that could impact model performance and contribute to overfitting. Simple Binary Logistic Regression (BLR) models were used

Table 1 Description of breast cancer features should appear here

	Description
Id	Identification number for each image
Target Variable	
Diagnosis	This denotes whether the digitized image of the cell nuclei show signs of being from a benign tumour or a malignant tumour.
Features	**Description**
Predictors	
Radius	The average distance from the centre (centroid) of a cell nucleus to its outer boundary points.
Perimeter	The mean of the total length of the boundary enclosing a cell nucleus.
Area	The mean of the total surface area enclosed by the nuclear boundary, calculated as the number of pixels within the nucleus plus half the number of pixels on the boundary.
Compactness	The average quantitative measure of how closely a cell nucleus resembles a perfect circle. It is calculated as the square of the perimeter divided by the area. Higher values indicate a less circular shape.
Smoothness	The average measure of the uniformity of a nuclear contour, calculated by analysing the variation in the length of radial lines drawn from the centroid to the boundary.
Concavity	The average degree to which a cell nucleus exhibits inward curves or indentations. It is quantified by measuring the depth of these concavities relative to the overall nuclear shape.
Concave Points	The average number of distinct inward curves or indentations present on a nuclear boundary.
Symmetry	The average measure of how evenly a cell nucleus can be divided into two mirror-image halves along its longest axis. It is calculated by comparing the distances between perpendicular lines drawn from this axis to the nuclear boundary.
Fractal Dimension	An average mathematical value that indicates the complexity and roughness of a nuclear boundary. It is estimated by measuring the perimeter of the nucleus using different-sized 'rulers' and analysing the relationship between ruler size and perimeter length. Higher values correlate with more irregular shapes.
Texture	The average description of the greyscale intensity variations within a cell nucleus. It is quantified by calculating the standard deviation of pixel values.

This table summarizes the features used in the breast cancer dataset. The Target Variable indicates whether a cell nucleus is from a benign or malignant tumour. The Predictor Variables describe various quantitative characteristics of the nuclei on these features that showed high correlation to identify which feature has the greatest influence on the target variable by observing the Pseudo R^2. This feature was then maintained for the model construction. This approach assessed the independent predictive power of each multicollinear feature, guiding its inclusion in the model-making process. Additionally, features showing no multicollinearity were maintained for the model-making process.

Traditional feature selection methods, such as Recursive Feature Elimination (RFE), were avoided due to their computational expense, risk of overfitting, and potential to select highly correlated features [52]. Instead, a combination of univariate statistical tests and multivariate correlation analysis was used, combining the strengths of filter-based techniques while addressing their limitations. This approach ensured the inclusion of relevant, independent features and enhanced model robustness. Statistical tests are also less computationally intensive than iterative methods like RFE, making them a resource-efficient choice while maintaining rigorous feature selection criteria.

This strategy aligns with recent studies that emphasize efficient handling of high-dimensional data streams and the use of multi-level filtering to enhance predictive model performance [22, 23]. Such frameworks highlight the value of computationally efficient feature management, particularly in medical diagnostics where large and complex datasets require scalable data processing pipelines.

SVM, Naïve Bayes and KNN models were developed. To ensure objectivity in model comparison, three stacking configurations were constructed, where in each case, two of the algorithms were used as base learners and the remaining one as the meta-classifier. This rotating meta-classifier design allowed for empirical validation of which model best combined the base models' predictions. Stacking ensemble models were constructed using the StackingClassifier from the sklearn.ensemble library [28, 77].

The selection of SVM, Naïve Bayes, and KNN as the algorithms used in this stacking ensemble was based on both theoretical complementarity and empirical performance. These three models represent distinct learning paradigms: SVM as a geometric classifier, Naïve Bayes as a probabilistic model, and KNN as a non-parametric, instance-based learner, each capturing unique data characteristics. This diversity is particularly valuable in stacking, where combining heterogeneous algorithms helps reduce correlated errors and improve generalization.

SVM performs well in high-dimensional spaces and defines clear class boundaries through margin maximization, while Naïve Bayes provides probabilistic robustness under limited data assumptions [46]. KNN, in contrast, excels at capturing local neighbourhood relationships without requiring distributional assumptions [74]. Together, their complementary strengths help mitigate one another's weaknesses such as SVM's sensitivity to noise, Naïve Bayes' independence assumption, and KNN's dependence on distance metrics, thereby enhancing ensemble robustness.

To validate the models' performance, k-fold cross-validations were used. The number of folds was set to ten for each model, striking a balance between overfitting and underfitting while ensuring reliable estimates of model performance [49].

To evaluate and compare the stacking models, the following performance metrics were employed: "accuracy", which quantifies the proportion of correctly predicted malignant and benign cases among all evaluated cases; "precision", which measures how accurately the model identifies a case as malignant among those predicted to be malignant; "recall", which assesses the model's capability to detect actual malignant cases from all malignant cases present and "F1-Score", to the show how accurate the model is in its predictions and how comprehensive it is in capturing malignant cases [95]. The mean accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristics Curve (ROC-AUC) for the 10-fold cross validation per model were calculated. ROC curves and their corresponding mean ROC curves per model were produced to check model performance.

All analyses were conducted on a local workstation equipped with an Intel® Core™ i5 processor (1.7 GHz), 8 GB RAM, and Windows 11 (64-bit) operating system. No cloud-based services were used. The models and data analyses were implemented using Python (version 3.12.4) and R (version 4.3.1), executed via Visual Studio Code (version 1.104) and RStudio (version 2024.09.0), respectively. Key Python libraries included scikit-learn (1.4.2), pandas (2.2.2), numpy (1.26.4), scipy (1.13.1), matplotlib (3.8.4), seaborn (0.13.2), kneed (0.8.5), statsmodel (0.14 2), and PIL (10.4.0). This configuration ensured consistent execution and reproducibility of all analytical processes [37, 78, 93].

To improve processing efficiency, the dataset was handled directly in memory during analysis, reducing read/write delays and improving computation speed. This approach aligns with recent research emphasizing efficient in-memory data processing for large and complex healthcare datasets [84]. Figure 2 illustrates a stepwise workflow of the main phases of the analysis, including data import, preprocessing, exploratory analysis, statistical evaluation, individual model development, stacking ensemble construction, and final model validation.

Results

Exploratory data analysis

The dataset consists of data from 569 patients from the University of Wisconsin, Madison, Department of Surgery and Human Oncology, where 357 (62.7%) cases were benign, and 212 (37.3%) were malignant. No duplicate records or missing values were detected in the dataset. Significant outliers were observed and the Winsorization method was applied to address both malignant and benign outliers.

Table 2 presents a comparison of means and standard deviations for benign and malignant cases based on the 10 cell features listed in Table 1. All features showed higher means for malignant tumours compared to benign ones, with slight differences observed in smoothness (Benign: $\mu = 0.09$, $\sigma = 0.01$; Malignant: $\mu = 0.10$, $\sigma = 0.01$) and symmetry (Benign: $\mu = 0.17$, $\sigma = 0.02$; Malignant: $\mu = 0.19$, $\sigma = 0.03$) and no differences were found in the Fractal Dimension ($\mu = 0.06$, $\sigma = 0.01$).

The distribution of the features was observed to be positively skewed thus deviating from normality. Therefore, the square root transformation was applied to each feature, resulting in a distribution that appeared closer to normality as shown in Fig. 3. This was

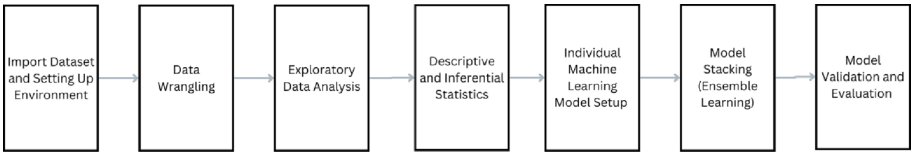


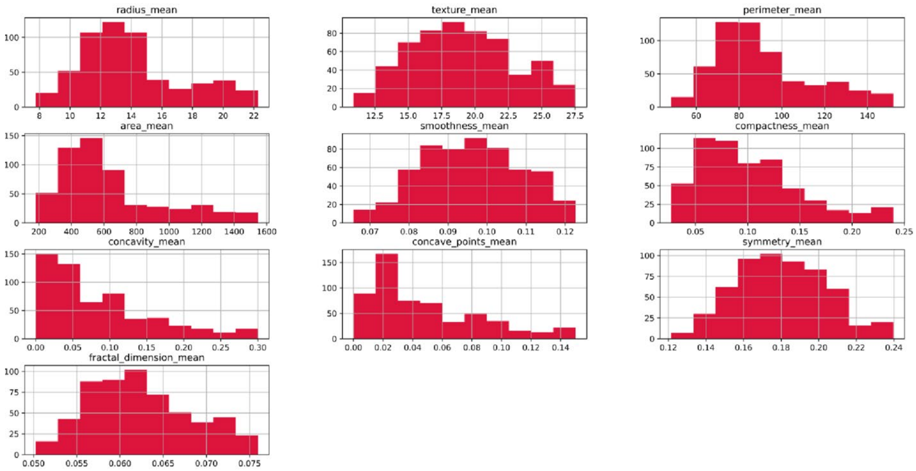
Fig. 2 Overview of the study methodology

Table 2 Comparison of averages and standard deviations of cell features between malignant and benign cases

Features	Benign		Malignant	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
Radius	12.15	1.78	17.46	3.20
Perimeter	78.08	11.81	115.37	21.85
Area	462.79	134.29	978.38	367.94
Compactness	0.08	0.03	0.15	0.05
Smoothness	0.09	0.01	0.10	0.01
Concavity	0.05	0.04	0.16	0.08
Concave Points	0.03	0.02	0.09	0.03
Symmetry	0.17	0.02	0.19	0.03
Fractal Dimension	0.06	0.01	0.06	0.01
Texture	17.91	4.00	21.60	3.78

mean and standard deviations of key cell nucleus features for benign and malignant cases. malignant nuclei generally show larger size, greater irregularity, and higher concavity, highlighting patterns useful for classification

(a) Histograms of Features



(b) Histograms of Features After Square Root Transformation

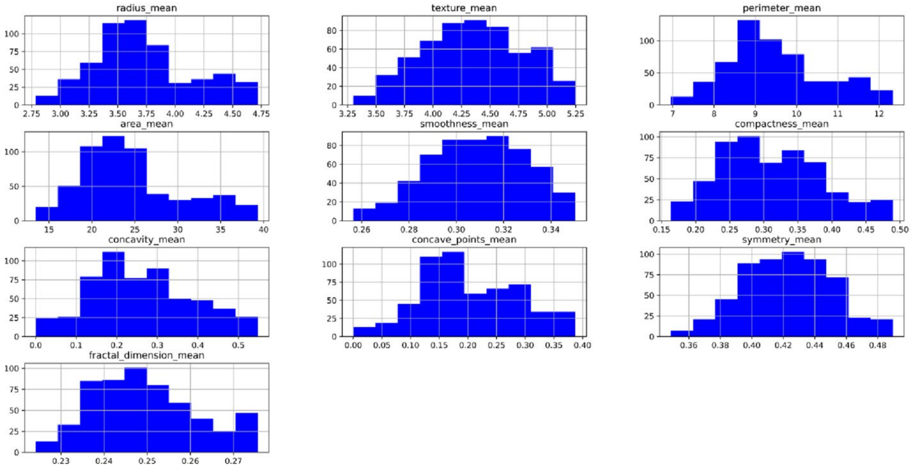


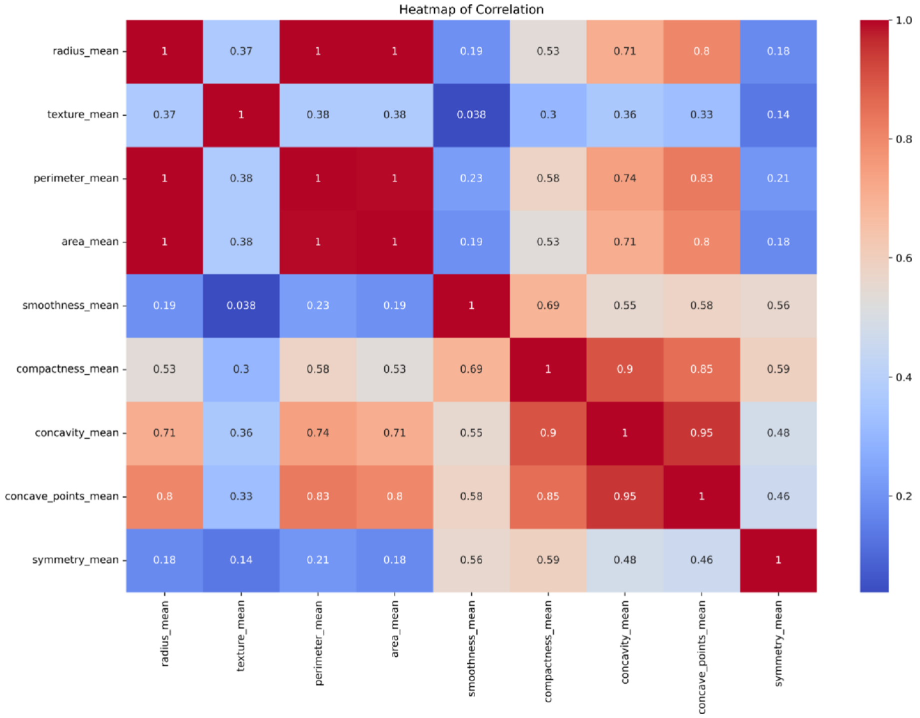
Fig. 3 Distribution of features before and after transformation. **a** Histograms showing the original distribution of key features in the Wisconsin Breast Cancer dataset. **b** Histograms of the same features after applying a square-root transformation, illustrating improved symmetry and reduced skewness

Table 3 Independent sample t-test of cell features between malignant and benign cases

Features	***P*-Value**
Radius	$<$ 0.001*
Perimeter	$<$ 0.001*
Area	$<$ 0.001*
Compactness	$<$ 0.001*
Smoothness	$<$ 0.001*
Concavity	$<$ 0.001*
Concave Points	$<$ 0.001*
Symmetry	<0.001*
Fractal Dimension	0.98
Texture	$<$ 0.001*

*Significant at 5% level of significance

P-values from independent t-tests comparing each cell feature between malignant and benign cases. Features with $p < 0.05$ (marked with *) show statistically significant differences, indicating that most nuclear characteristics differ between the two groups, except for fractal dimension



	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave_points_mean	symmetry_mean
radius_mean	1	0.37	1	1	0.19	0.53	0.71	0.8	0.18
texture_mean	0.37	1	0.38	0.38	0.038	0.3	0.36	0.33	0.14
perimeter_mean	1	0.38	1	1	0.23	0.58	0.74	0.83	0.21
area_mean	1	0.38	1	1	0.19	0.53	0.71	0.8	0.18
smoothness_mean	0.19	0.038	0.23	0.19	1	0.69	0.55	0.58	0.56
compactness_mean	0.53	0.3	0.58	0.53	0.69	1	0.9	0.85	0.59
concavity_mean	0.71	0.36	0.74	0.71	0.55	0.9	1	0.95	0.48
concave_points_mean	0.8	0.33	0.83	0.8	0.58	0.85	0.95	1	0.46
symmetry_mean	0.18	0.14	0.21	0.18	0.56	0.59	0.48	0.46	1

Fig. 4 Heatmap of Breast Cancer Features showing Multicollinearity. This figure displays the Pearson correlation coefficients among all features in the dataset. Higher correlation values indicate stronger linear relationships between variables. The heatmap highlights potential multicollinearity, supporting decisions related to feature reduction and model optimization

further validated using the distribution statistics of skewness and kurtosis, where both values for each feature were close to zero.

Table 3 shows that there exist significant differences between malignant and benign cases for all features except the average fractal dimension ($p = 0.98$) at the 5% level of significance. Consequently, the average fractal dimension was excluded in the model-making process since there is not sufficient evidence to support the claim of significant difference in means between malignant and benign cases.

Features showing weak to moderate positive correlations were retained for the stacking model. Figure 4 showed that radius, perimeter, area, compactness,

Table 4 Wald's test and Pseudo R^2 of multicollinear features

Features	*P*-Value	R^2
Simple BLR of features showing perfect correlation with each other		
Radius	< 0.001*	0.8
Perimeter	< 0.001*	0.8
Area	< 0.001*	0.81
Compactness	< 0.001*	0.49
Concavity	< 0.001*	0.67
Concave Points	< 0.001*	0.78

*Significant at 5% level of significance

Results of Wald's tests and corresponding Pseudo R^2 values for features exhibiting multicollinearity. All listed features show significant associations ($p < 0.05$) and high Pseudo R^2 indicating strong linear relationships and potential redundancy among these predictors in the dataset

Table 5 Performance comparison of stacked models using SVM, Naive Bayes, and KNN as base and Meta-Classifiers

Model	Base models	Meta-model	Accuracy	Precision	Recall	F1-score
1.	• Naïve Bayes • SVM	KNN	0.94	0.97	0.88	0.92
2.	• Naïve Bayes • KNN	SVM	0.95	0.96	0.90	0.93
3.	• KNN • SVM	Naïve Bayes	0.92	0.92	0.87	0.89

This table compares the predictive performance of stacking models using different combinations of base and meta-classifiers. Metrics include accuracy, precision, recall, and F1-score. Model 2, using Naive Bayes and KNN as base models with SVM as the meta-classifier, achieves the highest overall performance, highlighting the effectiveness of this combination for classifying breast cancer cases

concavity and concave points had perfect and strong positive correlation with each other ($0.7 \leq r \leq 1.0$). These features were therefore placed in simple BLR models where "area" showed the highest percentage variation in a patient's diagnosis, as shown in Table 4. Hence, amongst the six features, area was retained for the model-making process.

Modelling

For the KNN base model, k values from 1 to 25 were tested, with 25 selected as a conservative estimate. This aligns with the rule of thumb to use the square root of the sample size ($\sqrt{569} \approx 24$) as an initial estimate for optimal k [83]. The optimal k was found to be 11.

A comparative analysis of the three stacked models, each employing distinct combinations of SVM, Naïve Bayes and KNN as base and meta-models is shown in Table 5. Model 1, with Naïve Bayes and SVM as base models and KNN as the meta-model, achieved an accuracy of 94%. Model 2, using Naïve Bayes and KNN as base models with SVM as the meta-model, achieved the highest accuracy of 95%. Finally, Model 3, with KNN and SVM as base models and Naïve Bayes as the meta-model, reached 92% accuracy. Model 2 also showed higher recall, and F1-score, outperforming the other models.

The ROC curves for the 10-folds cross-validation for Stacking Models 1, 2, and 3 showed AUC scores of 0.98 ± 0.02, 0.97 ± 0.03, and 0.96 ± 0.03 respectively as presented in Fig. 5. The highest AUC scores for each stacking model were 1.00, 1.00, and 0.99, respectively, while the lowest scores were 0.95, 0.91, and 0.91.

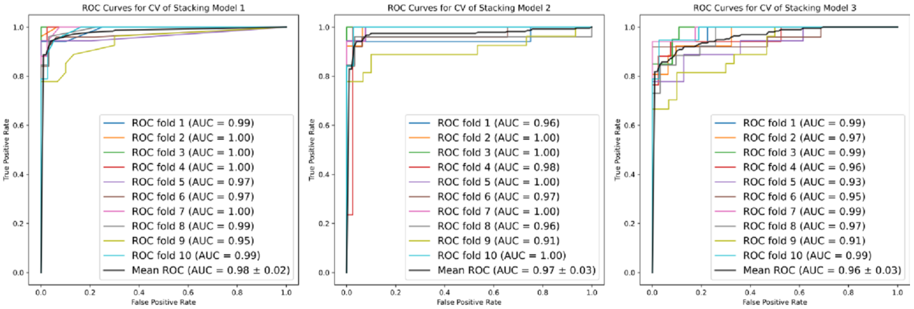


Fig. 5 ROC Curves for Cross-Validation of Stacking Models. ROC curves showing the performance of stacking models across cross-validation folds. AUC values indicate each model's overall classification accuracy and robustness

Discussion

Breast cancer remains a global health concern, representing the most common cancer diagnosis among women and a major contributor to cancer-related deaths worldwide [89]. While it is less prevalent in men, the challenges in early detection are more pronounced, often resulting in delayed diagnoses and poorer prognoses [97]. These complexities necessitate advancements in diagnostic approaches, highlighting the critical role of predictive models that leverage diverse data sources, such as imaging and omics. By integrating robust machine learning frameworks like stacking ensembles, this study contributes to the expanding corpus of research that emphasizes improved accuracy and early detection [30].

Despite advancements in breast cancer diagnosis, traditional methods like biopsies and imaging still face limitations. Biopsies, while essential for histopathological confirmation, are invasive procedures that can result in discomfort or complications, such as bruising, for patients [60]. Imaging methods such as mammography, ultrasound, and MRI, though crucial for detecting breast cancer, have their own drawbacks. For example, mammography can lead to false positives and false negatives, which might cause unnecessary anxiety, follow-up procedures, or even delayed treatment in some cases [76]. When paired with modern technologies, these methods gain new potential. For instance, integrating AI into imaging has shown improved sensitivity and specificity, reduced false-positive rates and enhanced early detection accuracy [34, 56].

Building upon the advancements of AI in enhancing sensitivity and specificity, this study contributed to existing knowledge by integrating SVM, Naïve Bayes, and KNN into a stacking ensemble framework. Sheth et al. (2022) [86] demonstrated the complementary strengths of these algorithms: SVM excels in handling high-dimensional data and is effective for classification tasks that involve complex boundaries, Naïve Bayes provides a probabilistic framework that efficiently processes both discrete and continuous data, and KNN is adept at capturing local data patterns due to its reliance on feature similarity [45, 46]. Together, these algorithms addressed key challenges in breast cancer diagnosis, such as varying feature distributions and the need for precise classification. By combining these algorithms in a stacking model, this study harnesses their unique advantages, achieving improved diagnostic accuracy, as demonstrated by Model 2's leading performance metrics. This ensemble approach not only supports the findings of Sheth et al. (2022) [86] but also builds on similar studies that have successfully utilized the three models mentioned, demonstrating their effectiveness in clinical applications [21, 48, 64].

By advancing their combined application, it illustrates how such model integrations can provide robust, interpretable results, enhancing diagnostic processes in healthcare.

However, it is important to note that although previous studies using the Breast Cancer Wisconsin (Diagnostic) dataset have applied algorithms such as SVM, Naïve Bayes, and KNN either individually or in limited combinations, relatively few have explored the specific stacking configuration implemented in this study in which two models function as base learners and the third as a meta-learner. This configuration, coupled with the rigorous feature selection and cross-validation procedures applied, represents a methodological refinement that distinguishes our approach from prior ensemble-based studies.

Unlike bagging and boosting techniques, which primarily aim to reduce variance or bias by iteratively resampling data or reweighting misclassified instances, stacking focuses on meta-learning, that is, combining diverse base models that capture different structural patterns within the data. Bagging methods such RF, aggregate homogeneous models to stabilize predictions, while boosting such as Gradient Boosting, sequentially improves weak learners. In contrast, stacking exploits the complementarity among heterogeneous classifiers such as the geometric decision boundaries of SVM, the probabilistic reasoning of Naïve Bayes, and the instance-based adaptability of KNN to capture nonlinear relationships that single or homogeneous ensemble methods may overlook.

By optimizing this heterogeneous integration, the present study demonstrates that stacking can yield improved diagnostic robustness and generalization across varying data partitions. Furthermore, the approach effectively addresses challenges inherent in biomedical classification tasks, such as class imbalance and the interpretability of multi-model predictions. This methodological contribution emphasizes that the intelligent design of stacking ensembles, beyond conventional bagging or boosting, can enhance both the accuracy and reliability of predictive models in breast cancer diagnosis.

Stacking ensembles, while powerful, are susceptible to overfitting. This risk is heightened when integrating diverse base learners, as biases can be amplified if not carefully managed. To combat this, 10-fold cross-validations were implemented to evaluate the model's performance on various data subsets, helping to identify and mitigate overfitting. Another risk to the stacking models was multicollinearity, which negatively impacts the performance of the model [5, 43, 47]. Features with strong correlations were analysed individually to guarantee the inclusion of only the most informative predictors in the three stacking models. From Table 4, "Area" emerged as the most informative feature, aligning with research by Abdalla et al. (2009) [2], which links size-related features to aggressive tumour behaviour and poorer survival outcomes. This result is further supported by studies emphasizing the importance of nuclear morphometry in distinguishing benign from malignant cases, reinforcing the clinical significance of this feature in breast cancer diagnostics [50, 66].

Acknowledging the importance of nuclear morphometry in breast cancer diagnosis, this study highlights the advantages of stacking ensemble models over both traditional diagnostic techniques, such as Fine Needle Aspiration Cytology (FNAC) and mammography. FNAC has been widely used to diagnose breast lesions, with reported diagnostic accuracy ranging from 75% to 94.7% as well as varying sensitivity ranging from 74.1% to 92.3% depending on lesion characteristics and clinical contexts [39, 91]. Mammography, another traditional diagnostic method, has an overall sensitivity of 87%, though this increases for women over 50 or those with fatty breasts. However, it is less effective for dense breast tissue, often necessitating supplementary techniques for accurate diagnosis [53]. In comparison, the Stacking Model 2 achieved a remarkable 95% accuracy and 90% recall (sensitivity), significantly outperforming these conventional methods. This aligns with findings where stacking ensembles reached up to 99.45% accuracy, demonstrating their ability to incorporate the advantages of base models like KNN and Naïve Bayes while addressing challenges such as feature variability and bias [55, 57].

In contrast to the approach adopted by Kumar et al. (2022) [55], this study took a more focused and methodical route in feature selection and model design. Kumar et al. (2022) [55] utilized all 30 features from the Wisconsin Breast Cancer dataset (including mean, standard error, and worst measurements) and incorporated the target feature, diagnosis, in their heatmap analysis. This heatmap used the Pearson correlation coefficient, which assumes that the variables of interest are measured on a continuous scale [36]. However, diagnosis is a discrete feature, making its inclusion in the heatmap potentially problematic and a source of bias. This study, on the other hand, concentrated solely on the 10 mean features and excluded the target variable from such correlation analyses, ensuring a streamlined model with enhanced interpretability. Additionally, while Kumar et al. (2022) [55] relied on Feature Importance scores for feature selection, this study addressed multicollinearity among features through a rigorous Binary Logistic Regression (BLR) analysis, retaining only the feature with the highest predictive power. This approach ensured robustness by mitigating issues related to collinearity, which were not explicitly addressed in Kumar et al.'s (2022) [55] work.

Additionally, Kumar et al. (2022) [55] explored a broader range of algorithms—10 in total, including advanced techniques like Gradient Boosting, XGBoost, and CatBoost—culminating in an Optimized Stacking Ensemble Learning (OSEL) model. This model was optimized using a genetic algorithm to identify the best combination of classifiers. Despite their effectiveness, genetic algorithms are notoriously expensive to compute, which may limit their applicability in settings with limited resources [88, 94]. In contrast, this study emphasized comparative simplicity and clarity, constructing three distinct stacking models using SVM, Naïve Bayes and KNN as base and meta-classifiers, and then directly comparing their performances. While Kumar et al.'s OSEL achieved a remarkable 99.45% accuracy, this study's Stacking Model 2 demonstrated competitive performance with 95% accuracy and 90% recall. By employing a less computationally intensive methodology, this study offers a practical and accessible framework for breast cancer diagnosis, particularly suitable for resource-constrained settings or scenarios prioritizing model interpretability.

While achieving high predictive accuracy is critical, this study's contribution extends beyond numerical performance. The stacking ensemble developed here advances the field by offering a clinically interpretable, computationally efficient, and resource-accessible framework for breast cancer diagnosis. In contrast to deep learning or complex hybrid methods that often require extensive data preprocessing, high-end GPU infrastructure, or opaque decision-making processes, this study's model emphasizes transparent and explainable learning, enabling clinicians and health data specialists to understand which morphological features most influence diagnostic outcomes.

Moreover, by demonstrating that a rigorously validated stacking model, built from heterogeneous yet complementary algorithms (SVM, Naïve Bayes, and KNN), can achieve competitive performance using only standard clinical features and modest computational resources, this study supports the translation of AI diagnostics into real-world healthcare environments, including low- and middle-income settings where access to advanced technologies may be limited. This practical adaptability represents a significant advancement over previous research that prioritized maximal accuracy at the expense of interpretability and scalability.

Importantly, this work reframes the role of ensemble learning in oncology, not merely as a tool for predictive optimization, but as a bridge between algorithmic sophistication and clinical utility. The methodological framework presented here can be extended to other diagnostic domains, encouraging the adoption of accessible, transparent, and high-performing AI tools in evidence-based medicine.

When comparing the stacking ensemble approach to other AI methods, its performance is notable in both diagnostic accuracy and practical application. Methods like the GoogLeNet-based deep learning framework, classified breast lesions in ultrasound images with an accuracy of approximately 90% and a sensitivity of 86% though it still required radiologists to manually identify regions of interest before classification [42]. Despite the GoogLeNet model's high specificity (95%), its reliance on a large dataset of 7,408 ultrasound images and significant computational resources, including an NVIDIA K40 GPU and extensive preprocessing, may have limited its practical deployment in diverse healthcare environments, especially resource-constrained settings where advanced hardware and radiological expertise for manual region selection are not readily available [42]. In contrast, Stacking Model 2 offers a balance between high diagnostic accuracy and moderate computational complexity, making it feasible for implementation even on standard CPUs (Central Processing Units), without requiring advanced GPUs (Graphics Processing Units), extensive infrastructure or reliance on radiologists for manual region selection. Its ability to effectively identify malignant cases, as evidenced by its higher recall, underscores its potential for public health applications where early and accurate detection is crucial. This adaptability ensures its utility across diverse healthcare environments, including those with limited computational resources.

The demonstrated efficacy of stacking ensemble models in improving diagnostic precision has significant implications for public health. The enhanced performance of Model 2 minimizes diagnostic errors, enabling timely and precise treatment decisions critical for improving survival rates, particularly in environments with constrained access to advanced imaging technologies, where early and accurate diagnoses play a pivotal role [6, 24].

The model's higher recall and precision can lead to more accurate identification of high-risk patients, ensuring that those most in need receive the most effective, targeted treatments, thereby improving patient outcomes. As a result, unnecessary treatments can be reduced, side effects can be minimized, and ultimately survival rates can be improved.

Scalability based on regional variations and data availability could significantly enhance machine learning models, particularly in addressing diagnostic disparities in low- and middle-income countries (LMICs) such as Latin America and the Caribbean. The highest breast cancer mortality rates are consistently reported in LMICs [24, 38, 59, 92]. By leveraging digital medical records, radiological images, and other data types, these models can provide reliable diagnostic support, thus narrowing healthcare disparities [63]. The best stacking ensemble model used in this study demonstrated significant improvements in breast cancer prediction, achieving higher accuracy and recall compared to conventional methods. This adaptability ensures that machine learning solutions can integrate with existing healthcare systems, facilitating timely diagnoses and treatment planning, which are critical for improving patient outcomes.

Limitations and recommendations

This research acknowledges potential limitations that may have influenced the study's outcomes and conclusions. Moreover, recommendations are offered to address these limitations and guide future studies in enhancing the validity and applicability of the findings.

Limitations

- The study relied on a single dataset, limiting the generalizability of the findings. The Wisconsin Breast Cancer dataset was recorded specifically for the research by W. H. Wolberg et al. (1994) [98]. The data may have been oversimplified due to limitations in computational power and imaging technologies at the time.
- There may be bias in results due to disparity in classification between benign and malignant cases, which could influence the models toward the majority class (benign), potentially resulting in a lower detection rate for malignant cases [27].
- The demographic and geographic specificity of the dataset restricts the generalizability of the findings to diverse populations, as the dataset lacks detailed demographic and racial composition data.

Recommendation

- To improve the validity and generalizability of future studies, it is recommended to validate findings using multiple datasets, ideally more recent ones, to enhance data accuracy and depth.
- Addressing potential class imbalance in the dataset, such as through oversampling or adjusting class weights, could improve detection rates for malignant cases.
- Future research should focus on using datasets with more comprehensive demographic information to ensure broader applicability of the findings across diverse populations.

Conclusion

This study explored the application of stacking ensemble methods to enhance the prediction and diagnosis of breast cancer, a critical area in medical research due to the disease's prevalence and impact. By integrating SVM, Naïve Bayes and KNN algorithms, three distinct stacking models were developed using the StackingClassifier from sklearn. ensemble. The models' performance was assessed using essential evaluation metrics to identify the most effective approach for breast cancer prediction.

From a public health perspective, the demonstrated effectiveness of Stacking Model 2 in achieving high diagnostic accuracy (95%) and recall (90%) addresses critical challenges in early breast cancer detection. These metrics reflect the model's ability to minimize missed malignant cases, which is pivotal in ensuring timely treatment and improving survival rates. Moreover, the model's adaptability to diverse healthcare settings, including those with limited computational resources, makes it an accessible tool for reducing diagnostic disparities in low- and middle-income countries, where breast cancer mortality rates remain disproportionately high. By facilitating early and accurate detection, the model contributes to better health outcomes, especially in resource-constrained environments.

The observations made in this study underscore the potential of stacking ensemble methods to improve breast cancer prediction, offering a promising direction for developing more accurate and reliable diagnostic tools. More in-depth research is vital to refine these models for broader clinical integration, enhancing their applicability across diverse healthcare settings and advancing early detection and treatment strategies for improved patient outcomes.

Abbreviations

AI	Artificial intelligence
BLR	Binary logistic regression
CPU	Central processing unit
DCN	Deep convolutional neural network
DL	Deep learning
EDA	Exploratory data analysis
FNAC	Fine needle aspiration cytology
GPU	Graphics processing unit
GRU	Gated recurrent unit
IQR	Inter-quartile range
KNN	K-nearest neighbours
LMICs	Low- and middle-income countries
LSTM	Long short-term memory
MRI	Magnetic resonance imaging
OSEL	Optimized stacking ensemble learning
RFE	Recursive feature elimination
ROC-AUC	Area under the receiver operating characteristics curve
SVM	Support vector machines
VUS	Variant of unknown significance
WHO	World Health Organization

Acknowledgements

Not applicable.

Author contributions

Conceptualization: AJ, AS and VT; Data Curation: AJ and AS; Formal Analysis: AJ and AS; Funding Acquisition: Not Applicable; Investigation: AJ and AS; Methodology: AJ and AS; Project Administration: AJ and VT; Resources: AJ and AS; Software: AS; Supervision: AJ and VT; Validation: AJ, AS and VT; Visualization: AJ and AS; Writing – Original Draft Preparation: AJ and AS; Writing – Review & Editing: AJ, AS and VT.

Funding

This study was not funded.

Data availability

As stated before, the dataset supporting the conclusions of this study is available at Kaggle and UCI Machine Learning Repository. The code used to acquire the results of this study is available at https://doi.org/10.5281/zenodo.15171573.

Declarations

Competing interests

The authors declares that there are no competing interests.

Received: 11 April 2025 / Accepted: 26 December 2025

References

1. Aamir A, Iqbal A, Jawed F, Ashfaque F, Hafsa H, Anas Z, Oduoye MO, Basit A, Ahmed S, Abdul Rauf S, Khan M, Mansoor T (2024) Exploring the current and prospective role of artificial intelligence in disease diagnosis. Ann Med Surg 86:943–949. https://doi.org/10.1097/MS9.0000000000001700
2. Abdalla F, Boder J, Markus R, Hashmi H, Buhmeida A, Collan Y (2009) Correlation of nuclear morphometry of breast cancer in histological sections with clinicopathological features and prognosis. Anticancer Res 29:1771–1776

3. Ahn JS, Shin S, Yang S-A, Park EK, Kim KH, Cho SI, Ock C-Y, Kim S (2023) Artificial intelligence in breast cancer diagnosis and personalized medicine. J Breast Cancer 26:405–435. https://doi.org/10.4048/jbc.2023.26.e45
4. Ahsan MM, Luna SA, Siddique Z (2022) Machine-Learning-Based disease diagnosis: A comprehensive review. Healthcare 10:541. https://doi.org/10.3390/healthcare10030541
5. Alexander (2015) Multicollinearity: Definition, Causes, Examples. In: Stat. To. https://www.statisticshowto.com/multicollinearity/. Accessed 11 Sept 2024
6. AlSamhori JF, AlSamhori ARF, Duncan LA, Qalajo A, Alshahwan HF, Al-abbadi M, Soudi MA, Zakraoui R, AlSamhori AF, Alryalat SA, Nashwan AJ (2024) Artificial intelligence for breast cancer: implications for diagnosis and management. J Med Surg Public Health 3:100120. https://doi.org/10.1016/j.glmedi.2024.100120
7. American Cancer Society (2022) Breast Cancer Mammogram | How Does a Mammogram Work? In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/mammogram-basics.html. Accessed 17 Aug 2024
8. American Cancer Society (2024) Breast Cancer Statistics | How Common Is Breast Cancer? In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html. Accessed 21 Aug 2024
9. American Cancer Society (2024) Key Statistics for Breast Cancer in Men. In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer-in-men/about/key-statistics.html. Accessed 21 Aug 2024
10. American Cancer Society (2023) ACS Breast Cancer Screening Guidelines. In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html. Accessed 21 Aug 2024
11. American Cancer Society (2022) What Is a Breast MRI? | Breast Cancer Screening. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/breast-mri-scans.html. Accessed 18 Aug 2024
12. American Cancer Society (2022) What Is a Breast Ultrasound? | Breast Cancer Screening. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/breast-ultrasound.html. Accessed 18 Aug 2024
13. American Cancer Society (2021) Breast Cancer Risk and Prevention: Genetic Testing. In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/risk-and-prevention/genetic-testing.html. Accessed 18 Aug 2024
14. American Cancer Society (2021) Types of Breast Cancer | About Breast Cancer. In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer.html. Accessed 21 Aug 2024
15. American Cancer Society (2021) What Is Breast Cancer? https://www.cancer.org/cancer/types/breast-cancer/about/what-is-breast-cancer.html. Accessed 21 Aug 2024
16. American Cancer Society (2024) Breast Biopsy | Biopsy Procedure for Breast Cancer. In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/breast-biopsy.html. Accessed 19 Aug 2024
17. American Cancer Society (2018) Can Breast Cancer in Men Be Found Early? In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer-in-men/detection-diagnosis-staging/detection.html. Accessed 21 Aug 2024
18. American Cancer Society (2021) Stages of Breast Cancer | Understand Breast Cancer Staging. In: Am. Cancer Soc. https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/stages-of-breast-cancer.html. Accessed 21 Aug 2024
19. American Cancer Society (2024) Imaging (Radiology) Tests for Cancer. In: Am. Cancer Soc. https://www.cancer.org/cancer/diagnosis-staging/tests/imaging-tests/imaging-radiology-tests-for-cancer.html. Accessed 18 Aug 2024
20. Breast Cancer Treatment in Bangalore | Expert Care at Apollo Hospitals. In: Apollo Hosp. Apollo Hospitals Enterprise Ltd, Bangalore (2024) https://www.apollohospitals.com/bangalore/departments/breast-cancer-treatment/. Accessed 28 Aug 2024
21. Asri H, Mousannif H, Moatassime HA, Noel T (2016) Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Comput Sci 83:1064–1069. https://doi.org/10.1016/j.procs.2016.04.224
22. Banirostam H, Banirostam T, Pedram MM, Rahmani A (2023) A model to detect the fraud of electronic payment card transactions based on stream processing in big data. J Signal Process Syst 95. https://doi.org/10.1007/s11265-023-01903-6
23. Banirostam H, Banirostam T, Pedram MM, Rahmani A (2023) Providing and evaluating a comprehensive model for detecting fraudulent electronic payment card transactions with a two-level filter based on flow processing in big data. Int J Inf Technol 15. https://doi.org/10.1007/s41870-023-01501-6
24. Barrios CH (2022) Global challenges in breast cancer detection and treatment. Breast 62:S3–S6. https://doi.org/10.1016/j.breast.2022.02.003
25. BCRF (2024) AI Breast Cancer Detection and Diagnosis. In: Breast Cancer Res. Found. https://www.bcrf.org/blog/ai-breast-cancer-detection-screening/. Accessed 18 Aug 2024
26. Berrar D (2019) Cross-Validation. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C (eds) Encyclopedia of bioinformatics and computational biology. Academic, Oxford, pp 542–545
27. Brownlee J (2020) A Gentle Introduction to Imbalanced Classification - MachineLearningMastery.com. https://machinelearningmastery.com/what-is-imbalanced-classification/. Accessed 12 Sept 2024
28. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R, VanderPlas J, Joly A, Holt B, Varoquaux G (2013) API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp 108–122
29. Cancer Research UK (2023) What is breast cancer? In: Cancer Res. UK. https://www.cancerresearchuk.org/about-cancer/breast-cancer/about. Accessed 28 Aug 2024
30. Chakraborty C, Bhattacharya M, Pal S, Lee S-S (2024) From machine learning to deep learning: advances of the recent data-driven paradigm shift in medicine and healthcare. Curr Res Biotechnol 7:100164. https://doi.org/10.1016/j.crbiot.2023.100164
31. Cleveland C (2023) Breast Cancer. In: Clevel. Clin. https://my.clevelandclinic.org/health/diseases/3986-breast-cancer. Accessed 28 Aug 2024
32. Cleveland Clinic (2024) What's a Breast Ultrasound & Why Do I Need One? In: Clevel. Clin. https://my.clevelandclinic.org/health/diagnostics/21496-breast-ultrasound. Accessed 28 Aug 2024
33. Das S, Koley S, Saha T (2023) Machine learning approaches for investigating breast cancer. Biosci Biotechnol Res Asia
34. Dembrower K, Crippa A, Colón E, Eklund M, Strand F (2023) Artificial intelligence for breast cancer detection in screening mammography in sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit Health 5:e703–e711. https://doi.org/10.1016/S2589-7500(23)00153-X

35. Dongola N (2024) Mammography in Breast Cancer: Background, X-ray Mammography, Ultrasound. https://emedicine.medscape.com/article/346529-overview?form=fpf. Accessed 19 Aug 2024
36. Fein EC, Gilmour J, Machin T, Hendry L (2022) Sect. 4.2: Correlation Assumptions, Interpretation, and Write Up
37. Fox J, Weisberg S (2018) An R Companion to Applied Regression, 3rd edition. SAGE Publications, Inc, Los Angeles
38. Francies FZ, Hull R, Khanyile R, Dlamini Z (2020) Breast cancer in low-middle income countries: abnormality in splicing and lack of targeted treatment options. Am J Cancer Res 10:1568–1591
39. Georgieva RD, Obdeijn IM, Jager A, Hooning MJ, Tilanus-Linthorst MMA, van Deurzen CHM (2013) Breast fine-needle aspiration cytology performance in the high-risk screening population. Cancer Cytopathol 121:561–567. https://doi.org/10.1002/cncy.21308
40. Geras KJ, Wolfson S, Kim SG, Moy L, Cho K (2017) High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks. ArXiv
41. Gilbert FJ, Pinker-Domenig K (2019) Diagnosis and staging of breast cancer: when and how to use Mammography, Tomosynthesis, Ultrasound, Contrast-Enhanced Mammography, and magnetic resonance imaging. In: Hodler J, Kubik-Huch RA, von Schulthess GK (eds) Diseases of the Chest, Breast, heart and vessels 2019–2022: diagnostic and interventional imaging. Springer, Cham (CH)
42. Han S, Kang H-K, Jeong J-Y, Park M-H, Kim W, Bang W-C, Seong Y-K (2017) A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys Med Biol 62:7714. https://doi.org/10.1088/1361-6560/aa82ec
43. Hayes A (2024) Multicollinearity: Meaning, Examples, and FAQs. In: Investopedia. https://www.investopedia.com/terms/m/multicollinearity.asp. Accessed 11 Sept 2024
44. IBM (2023) What Is Multicollinearity? | IBM. https://www.ibm.com/topics/multicollinearity. Accessed 11 Sept 2024
45. IBM (2021) What Are Naïve Bayes Classifiers? | IBM. In: IBM. https://www.ibm.com/topics/naive-bayes. Accessed 28 Aug 2024
46. IBM (2023) What Is Support Vector Machine? | IBM. In: IBM. https://www.ibm.com/topics/support-vector-machine. Accessed 28 Aug 2024
47. IBM (2021) What is the k-nearest neighbors algorithm? | IBM. In: IBM. https://www.ibm.com/topics/knn. Accessed 28 Aug 2024
48. Islam MM, Iqbal H, Haque MR, Hasan MK (2017) Prediction of breast cancer using support vector machine and K-Nearest neighbors. In: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, Dhaka, pp 226–229
49. James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning: with Applications in R, 1st ed. 2013, Corr. 7th printing 2017 edition. Springer, New York
50. Kashyap A, Jain M, Shukla S, Andley M (2017) Study of nuclear morphometry on cytology specimens of benign and malignant breast lesions: A study of 122 cases. J Cytol 34:10. https://doi.org/10.4103/0970-9371.197591
51. Kelleher JD, MacNamee B, D'Arcy A (2015) Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. The MIT Press, Cambridge, Massachusetts London, England
52. Khaire UM, Dhanalakshmi R (2022) Stability of feature selection algorithm: A review. J King Saud Univ - Comput Inf Sci 34:1060–1073. https://doi.org/10.1016/j.jksuci.2019.06.012
53. Komen SG (2022) Accuracy of Mammograms. In: Susan G Komen®. https://www.komen.org/breast-cancer/screening/mammography/accuracy/. Accessed 27 Nov 2024
54. Komen SG (2023) Clinical Breast Exam. In: Susan G Komen®. https://www.komen.org/breast-cancer/screening/clinical-breast-exam/. Accessed 28 Aug 2024
55. Kumar M, Singhal S, Shekhar S, Sharma B, Srivastava G (2022) Optimized stacking ensemble learning model for breast cancer detection and classification using machine learning. Sustainability 14:13998. https://doi.org/10.3390/su142113998
56. Lång K, Josefsson V, Larsson A-M, Larsson S, Högberg C, Sartor H, Hofvind S, Andersson I, Rosso A (2023) Artificial intelligence-supported screen reading versus standard double reading in the mammography screening with artificial intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol 24:936–944. https://doi.org/10.1016/S1470-2045(23)00298-X
57. Lu M, Hou Q, Qin S, Zhou L, Hua D, Wang X, Cheng L (2023) A stacking ensemble model of various machine learning models for daily runoff forecasting. Water 15:1265. https://doi.org/10.3390/w15071265
58. Mahajan P, Uddin S, Hajati F, Moni MA (2023) Ensemble learning for disease prediction: A review. Healthcare 11:1808. https://doi.org/10.3390/healthcare11121808
59. Manson EN, Achel DG (2023) Fighting breast cancer in low-and-middle-income countries – What must we do to get every woman screened on regular basis? Sci Afr 21:e01848. https://doi.org/10.1016/j.sciaf.2023.e01848
60. Mayo Clinic Staff (2023) Breast biopsy - Mayo Clinic — mayoclinic.org
61. Menon G, Alkabban FM, Ferguson T (2024) Breast cancer. In: StatPearls. StatPearls Publishing, Treasure Island (FL)
62. Mohammed A, Kora R (2023) A comprehensive review on ensemble deep learning: opportunities and challenges. J King Saud Univ - Comput Inf Sci 35:757–774. https://doi.org/10.1016/j.jksuci.2023.01.014
63. Mollura DJ, Culp MP, Pollack E, Battino G, Scheel JR, Mango VL, Elahi A, Schweitzer A, Dako F (2020) Artificial intelligence in Low- and Middle-Income countries: innovating global health radiology. Radiology 297:513–520. https://doi.org/10.1148/radiol.2020201434
64. Naji MA, Filali SE, Aarika K, Benlahmar EH, Abdelouhahid RA, Debauche O (2021) Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Comput Sci 191:487–492. https://doi.org/10.1016/j.procs.2021.07.062
65. Narad P, Kirthanashri SV (2018) Introduction to omics. In: Arivaradarajan P, Misra G (eds) Omics Approaches, technologies and applications. Springer Singapore, Singapore, pp 1–10
66. Narasimha A, Vasavi B, Kumar HM (2013) Significance of nuclear morphometry in benign and malignant breast aspirates. Int J Appl Basic Med Res 3:22. https://doi.org/10.4103/2229-516X.112237
67. Nasser M, Yusof UK (2023) Deep learning based methods for breast cancer diagnosis: A systematic review and future direction. Diagnostics 13:161. https://doi.org/10.3390/diagnostics13010161
68. National Cancer Institute (2024) BRCA Gene Mutations: Cancer Risk and Genetic Testing Fact Sheet - NCI. In: Natl. Cancer Inst. https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet. Accessed 18 Aug 2024
69. National Cancer Institute (2019) Advances in Breast Cancer Research - NCI. https://www.cancer.gov/types/breast/research. Accessed 17 Aug 2024

70. National Cancer Institute (2011) Definition of MRI - NCI Dictionary of Cancer Terms - NCI. In: Natl. Cancer Inst. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/mri. Accessed 18 Aug 2024
71. National Cancer Institute (2011) Definition of mammography - NCI Dictionary of Cancer Terms - NCI. In: Natl. Cancer Inst. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/mammography. Accessed 17 Aug 2024
72. National Breast Cancer Foundation (2024) Breast Biopsy: Procedure Types, What to Expect & Results Guide. In: Natl. Breast Cancer Found. https://www.nationalbreastcancer.org/breast-cancer-biopsy/. Accessed 28 Aug 2024
73. Olmsted Medical Center (2023) Breast Cancer: Early Detection. http://healthlibrary.olmmed.org/Library/DiseasesConditions/Adult/34,8528. Accessed 17 Aug 2024
74. Ougiaroglou S, Evangelidis G (2015) Dealing with noisy data in the context of k-NN Classification. In: Proceedings of the 7th Balkan Conference on Informatics Conference. ACM, Craiova Romania, pp 1–4
75. PAHO (2024) Breast cancer - PAHO/WHO | Pan American Health Organization. In: Pan Am. Health Organ. https://www.paho.org/en/topics/breast-cancer. Accessed 16 Aug 2024
76. PDQ® Screening and Prevention Editorial Board (2024) Breast Cancer Screening (PDQ®) - NCI. https://www.cancer.gov/types/breast/hp/breast-screening-pdq. Accessed 26 Nov 2024
77. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
78. R Core Team (2023) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
79. Ray S (2017) Naive Bayes Classifier Explained: Applications and Practice Problems of Naive Bayes Classifier. In: Anal. Vidhya. https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/. Accessed 28 Aug 2024
80. Reza Rabiei S, Mohammad, Ayyoubzadeh S, Sohrabei A, Atashi M Esmaeili (2022) Prediction of breast cancer using machine learning approaches. J Biomed Phys Eng 12. https://doi.org/10.31661/jbpe.v0i0.2109-1403
81. Rivera-Franco MM, Leon-Rodriguez E (2018) Delays in breast cancer detection and treatment in developing countries. Breast Cancer Basic Clin Res 12:1178223417752677. https://doi.org/10.1177/1178223417752677
82. Rossi C, Cicalini I, Cufaro MC, Consalvo A, Upadhyaya P, Sala G, Antonucci I, Del Boccio P, Stuppia L, De Laurenzi V (2022) Breast cancer in the era of integrating omics approaches. Oncogenesis 11:1–13. https://doi.org/10.1038/s41389-022-00393-8
83. Saha D, Manickavasagan A (2021) Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. Curr Res Food Sci 4:28–44. https://doi.org/10.1016/j.crfs.2021.01.002
84. Shamsinejad E, Banirostam H, Banirostam T, Pedram MM, Rahmani A (2025) Providing and evaluating a model for big data anonymization streams by using in-memory processing. Knowl Inf Syst 67:7791–7824. https://doi.org/10.1007/s10115-025-02417-2
85. Sharma S, Aggarwal A, Choudhury T (2018) Breast cancer detection using machine learning algorithms. 2018 Int Conf Comput Tech Electron Mech Syst CTEMS 114–118
86. Sheth V, Tripathi U, Sharma A (2022) A comparative analysis of machine learning algorithms for classification purpose. Procedia Comput Sci 215:422–431. https://doi.org/10.1016/j.procs.2022.12.044
87. Shivahare BD, Singh J, Ravi V, Chandan RR, Alahmadi TJ, Singh P, Diwakar M (2024) Delving into machine learning's influence on disease diagnosis and prediction. https://doi.org/10.2174/0118749445297804240401061128
88. Sulianta F (2024) Critical evaluation of genetic algorithms. From Computational Costs to Theoretical Gaps
89. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/caac.21660
90. Thomas TH (2019) Breast cancer tumor size chart: Factors and more. https://www.medicalnewstoday.com/articles/325669. Accessed 19 Aug 2024
91. Tripathi K, Yadav R, Maurya SK (2022) A comparative study between Fine-Needle aspiration cytology and core needle biopsy in diagnosing clinically palpable breast lumps. Cureus 14:e27709. https://doi.org/10.7759/cureus.27709
92. Unger-Saldaña K (2014) Challenges to the early diagnosis and treatment of breast cancer in developing countries. World J Clin Oncol 5:465–477. https://doi.org/10.5306/wjco.v5.i3.465
93. Van Rossum G, Drake FL Jr (2024) The Python Language Reference. In: Python Doc. https://docs.python.org/3/reference/index.html. Accessed 22 Oct 2024
94. Vié A, Kleinnijenhuis AM, Farmer DJ (2020) Qualities, Challenges and Future of Genetic Algorithms
95. Vujovic Z (2021) Classification model evaluation metrics. Int J Adv Comput Sci Appl Volume 12:599–606. https://doi.org/10.14569/IJACSA.2021.0120670
96. WHO (2024) Breast cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Accessed 19 Aug 2024
97. Wilson J, Sule AA (2022) Disparity in early detection of breast cancer. StatPearls [Internet]. StatPearls Publishing
98. Wolberg WH, Street WN, Mangasarian OL (1994) Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. Cancer Lett 77:163–171. https://doi.org/10.1016/0304-3835(94)90099-X
99. Wolberg W, Mangasarian O, Street N, Street W (2016) Breast Cancer Wisconsin (Diagnostic) Data Set. In: Kaggle. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data. Accessed 28 Aug 2024
100. Wolberg William W (1995) Breast Cancer Wisconsin (Diagnostic)

Publisher's Note