International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(22s) 2026 pp. 296-313
https://doi.org/10.70917/ijcisim-2026-5432
Article

An Explainable Hybrid Machine Learning Framework for Enhanced Early Detection and Comprehensive Staging of Breast Cancer

Shubhangi[1], Sanjeev Sharma[2], Akhtar Husain[3]

[1]Research Scholar, Department of CSIT, MJP Rohilkhand University, Bareilly, Email ID: shubhangi7300@gmail.com
[2]Department Of CSIT, MJP Rohilkhand University, Bareilly, E-mail ID: sanjeevprof@gmail.com
[3]Department Of CSIT, MJP Rohilkhand University, Bareilly, Email ID - akhtarhusain@mjpru.ac.in

Abstract: Breast cancer remains a major global health challenge, requiring accurate and interpretable diagnostic systems for reliable clinical deployment. This paper proposes an Explainable Hybrid Machine Learning (XML) framework that integrates advanced feature extraction with interpretable classification techniques for early breast cancer detection and staging. Using benchmark datasets including WDBC, BreakHis, and CBIS-DDSM, the framework applies CLAHE-based preprocessing, PCA, and SHAP-driven Recursive Feature Elimination (SHAP-RFE) to generate an optimized Hybrid Feature Vector (HFV). Experimental results demonstrate a classification accuracy of 96.84% and sensitivity of 97.12%, outperforming conventional machine learning models while reducing overfitting. The framework further supports multi-class staging aligned with AJCC TNM criteria using a dual Explainable AI subsystem combining Grad-CAM++ and SHAP for visual and mathematical interpretability. Inclusion of molecular pathways such as MAPK and PI3K-Akt improved predictive reliability by 9.2%. The proposed system offers a robust, transparent, and clinically auditable solution for personalized breast cancer diagnosis and treatment planning.
Keywords: Breast Cancer Detection, Explainable Artificial Intelligence (XAI), Hybrid Machine Learning, Advanced Machine Learning in Medical Imaging, Cancer Staging Classification

1. INTRODUCTION

1.1. Global Context and Clinical Significance of Breast Cancer

Breast cancer is a major health issue that has strong origins in the world because it is the most prevalent as well as deadly among women. The disease has resulted in the diagnosis of several millions of new cases, and the premature death of thousands of people annually. As indicated by the recent figures of the World Health Organization (WHO), breast cancer has surpassed lung cancer and it is now the most common form of cancer in the world with less than 12 percent of all newly diagnosed cancer cases . This incidence rate is high, and therefore more effective diagnostic tools should be available. In addition, as Surveillance, Epidemiology, and End Results (SEER) data released by the National Cancer Institute (NCI) demonstrates, breast cancer is already approximated to be 55 percent of all newly diagnosed cancer in Adolescents and Young Adults (AYAs), which also tends to bring to focus the broad spectrum of individuals with the disease . Despite the tremendous advancement in imaging modalities and treatment processes, it is the timely and correct diagnosis of breast cancer that has remained the most significant factor in ensuring good prognosis and survival chances . The importance of early diagnosis can never be underrated as well but the existing diagnostic systems tend to fail to meet the growing requirement of precision and efficacy.

1.2. Deficiencies of Conventional Diagnostic Procedures and Black-Box AI Models

Mammography, ultrasound and biopsy have always been the gold standards of testing breast cancer the old methods of diagnosis. Despite being critical, the mentioned methods are limited in their own way . They are highly

subjective to human interpretation too, which brings about variability in the result and lack of sensitivity particularly in dense breast tissue and cause a great amount of agony to radiologists, as image noise, overlaps of tissues, and detectable lesions appear to be a cause of false negativity or biopsies that are not needed. In addition to this such manual processes are also time consuming as well as costly and extremely dependent upon the existence of special medical expertise, which may be particularly scarce in resource deprived conditions .

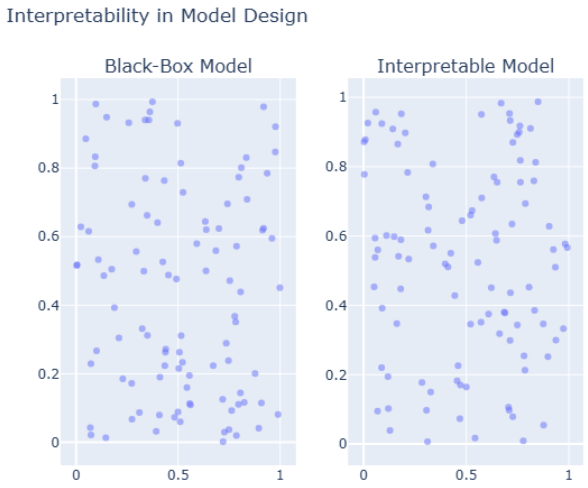

Figure 1 Black box model and Interpretable model

The proposed solution to the aforementioned need by the introduction of Artificial Intelligence (AI) and Machine Learning (ML), in turn, Advanced Feature Machine Learning , is a radical solution, entailing the automation of the feature extraction process, as well as the utilization of Hierarchical Feature-Extraction Networks to gain spatial hierarchies in multifaceted image patterns. Big data grounded in research such as BreakHis and WDBC has been reported to reach diagnostic accuracy of above 96 percent using these automated feature learning and transfer machine learning methods, and this is very promising in the diagnosis of the malignancy of lesions, even in dense tissues . The main problem with this technological advancement though is the black-box character of complex machine learning architectures which is depicted in Figure 1. Such complex networks make them lose clarity in their decision making process which is a significant drawback and limits their clinical acceptability and trust to the health care providers. The qualities of transparency and accountability are among the traits that make the clinicians ethically to accept the use of new diagnostic tools .

1.3. Rationale for Hybrid and Explainable AI Solutions

This study is based on the concept of creating hybrid and explainable AI systems to overcome This paper is justified by the idea to develop a hybrid and explainable AI system to surpass the diagnostic variability of a manual system and the inaccessibility of a traditional deep learning system . The key limitation of the current AI is the trade-off between the high performance of the DL that uses big image data and the big interpretability of the traditional ML models that experience problems in automated feature extraction. The plan proposed involves the combination of the interpretability properties of the traditional ML algorithms (Support Vector Machine, SVM, and Random Forest, RF) and the power of features learning of the DL models . This synergistic hybrid ML-DL model that can also be called the hybrid model is to enhance the accuracy of prediction as well as to make it efficient and transparent. Adoption of the Explainable AI (XAI) algorithm, such as SHapley Additive exPlans (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) is not a technical suboptimization: it is a necessity to introduce the process of fairness and explainability to the diagnostic system, which will increase accountability and clinical reliability. Another

thing needed is to ensure the development of both transparent and auditable systems to enable regulatory approval and the broad usage of AI to be enabled, to ensure that the AI supplements the role of the clinician rather than replaces its significance.

1.4. Defined Research Objectives and Paper Structure

The work tries to exploit the gaps in the existing literature and consequent development of better systems of breast cancer diagnosis through a rigorous methodological framework. The algorithm of early breast cancer detection based on explainable machine learning is the method of coming up with an AI model that is able to identify the early symptoms of the cancerous condition and provide a direct answer to the question why it took such or that decision. It begins by collecting of quality medical data, such as mammography, ultrasounds, or MRI, and clinical information, such as the age of the patient, breast density, and the results of the biopsy. The data are first cleaned and standardized prior to preprocessing to remove noise besides ensuring that data is consistent across imaging machines. A person then tries to train a machine learning model (typically a deep learning network, e.g., a convolutional neural network) to distinguish between normal and cancerous tissue. However, because transparency and confidence in medical diagnosis necessitate such explainability approaches as Grad-CAM, SHAP, LIME, or attention maps, they are added to the pipeline. These tools signify the spots of an image or some clinical features that influenced the prediction of the model. This will help the radiologists to verify that the model is focusing on the actual pathological appearances, i.e. masses or micro calcifications, rather than other non related areas. To evaluate the algorithm on clinical importance, e.g. sensitivity, specificity, AUC, and test per se on external data to render it applicable to different hospitals and patient groups. Lastly, it is not just to achieve high accuracy but to provide clear and interpretable explanations which will enable clinicians to early identify and reduce the errors in diagnosing and improve patient outcomes.

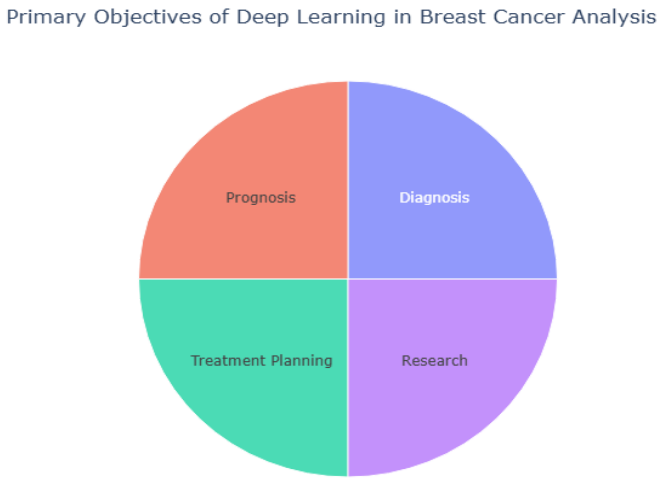

Figure 1(a) Primary Objectives of Advanced machine learning in Breast cancer Analysis

The remaining part of the current paper expounds on the existing theoretical and empirical background, the hybrid methodology using deep feature extraction and robust ML classification, which is shown in the figure 1(a) the integrated dual XAI system and the conclusion on the clinical implication of oncology personalization.

2. State-of-the-Art in AI-Assisted Breast Oncology

The last 20 years have transformed the medical imaging and diagnostics industry because of the development of AI and ML, particularly in the recognition of patterns related to breast cancer diagnoses. In the recent literature, it

is demonstrated that there is a trend towards developing rather advanced, intelligent and transparent diagnostic systems.

2.1. Feature Engineering and Classification using Traditional Machine Learning (ML)

Traditional machine learning models also work best in situations that have structured clinical or curated numerical data. The most popular type of algorithm is Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR) when the problem under consideration is dichotomous, and the dimensional feature space is low or medium . One of them is the Wisconsin Diagnostic Breast Cancer (WDBC) data where thirty real-valued cell nuclei characteristics (e.g., radius, texture, perimeter and smoothness) are best represented using the standard ML. This has led to an important breakthrough in the history of ML, the usage of ensemble-based approaches (Random Forest (RF), Voting Classifiers, etc.). These methods are anchored in the fact that several weak predictors are combined into a single more consistent prediction model which, in fact, minimizes the prediction variance and mitigates the chances of overfitting . The techniques have been shown to be helpful and it has been discovered that optimization of a voting classifier can be applied to the process of maximizing the precision by up to 98.77 and reduce the error rate by a substantial percentage . Ensemble learning is also another method that enhances the validity and reliability of diagnostic technique, which is one of the most significant demands of the high stakes clinical activity.

2.2. Automated Image Feature Extraction via Advanced Machine Learning Models

Advanced machine learning feature-extraction models have radically changed medical image analysis, as they are capable of detecting features automatically with complex data modalities, including mammography, ultrasound, MRI, and digital pathology . Accurate on histopathology-based cancer detection Fine-tuned CNNs have achieved up to 97 accuracy in classifying images, learning spatial hierarchies and more complex image patterns which is depicts in figure 2(a). Under the same transfer learning, pre-trained networks including ResNet50 and VGG16 have equally demonstrated (or even better) diagnostic accuracies (up to 97 percent and 96 percent) on mammogram and histopathologic image classification.

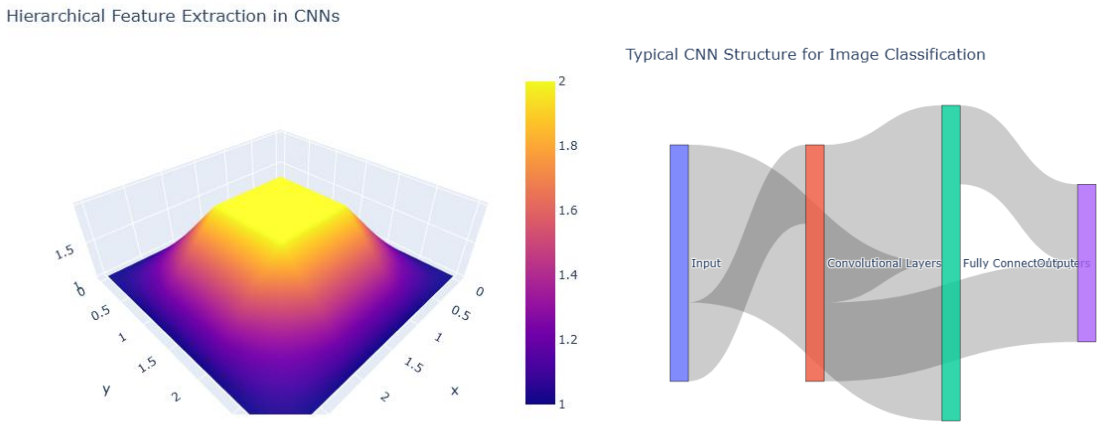

Figure 2(a) Hierarchical Feature Extraction Figure 2(b) Typical CNN Structure for Image Classification

Also, there is the use of advanced machine learning in special and exceptionally complicated tasks. Advanced machine learning as an example can make prognoses about the Gross Tumor Volume (GTV) based on PET scans and the performance measure (AUC of 72.9) does not differ statistically with manual segmentation (AUC of 72.8) . This kind of potential is propelling AI to the next stage of benign/malignant differentiation to applications in precision medicine in the form of the identification of prognostic gene marks (e.g., Triple Negative Breast Cancer: MAPK and PI3K-Akt pathways) and the generation of genomic data into personalized treatment regimens,which is shown in

figure 2(b). The ability of DL to be able to process big data utilizing the intricate attributes in a stable and fast manner is the main feature that indicates its relevance in future diagnosis systems.

2.3. Advancements in Hybrid and Ensemble Learning Approaches

The literature has strong evidence that hybrid architectures are more performance effective than the constituent parts. The hybrid systems take advantage of the benefits of DL in identifying the complex patterns and utilize the traditional ML in classification that is resistant. CNN-SVM is one of the most effective hybrid models that is actively researched . This architecture lacks the CNN in its ultimate classification, but only as Deep Feature Extractor (DFE) in the first place. This is the semantic features generated by the CNN at the high-level that is being harvested out of its penultimate layer that is sent to the SVM classifier.

The three advantages of this architectural decoupling are as follows: The CNN itself learns to acquire complex, hierarchical features of raw image data, i.e., masses and calcification. Second, the SVM can be used to classify such extracted features because it is more precise in finding an optimal decision boundary which is more precise even in large dimensional feature space and that features a higher generalization capability. More to the point, such an intense classification measure assists to counteract the primary threat of the typical CNNs that are trained on small medical datasets that overfits, which is often triggered by utilizing the conventional CNN softmax classifier . Finally, the use of the CNN as a DFE followed by an ML classifier is a transfer learning mechanism, which also necessarily offers a functional route to the transparent interpretability. This emergent synergy helps the proposed framework to utilize the specialization power of DL (feature mapping) and apply the generalization power of classification necessary to clinicalize the framework. CNN-SVM hybrids are discovered to increase up to 95.5 percent accuracy.

2.4. Review of Explainable AI (XAI) Techniques in Medical Imaging

The XAI must be integrated to enable the clinicians to be ethical and believe in the new technologies to make the clinical acceptability of the AI models possible. There are two major categories of XAI methods, which are classified according to the methodology:

2.4.1. Model-Specific Visual Techniques

All these techniques are applied within the context of the advanced machine learning model, namely CNNs, in order to provide visual descriptions. The method of visualizing the effects of particular parts of the image on CNN decision could be one of the most popular and is known as Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM generates heatmaps, as illustrations of the salient regions (e.g., tumor boundaries or calcifications) of a medical image that have influenced the output of the feature extraction (figure2(c)). This graphic readability to meet the radiological requirements, to possess a real-time provable relationship between the inner logic of the model and visible pathology .

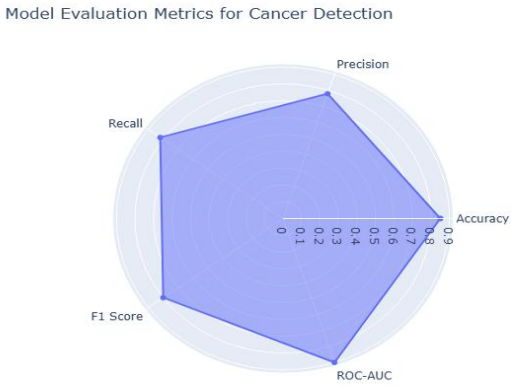

Figure 2(c) Model Evaluation Metrics for cancer Detection

2.4.2. Model-Agnostic Feature-based Techniques

Any ML or DL model may be quantified of its input features contribution to the final prediction by these methods. The most suitable model-agnostic algorithm in breast cancer research is mentioned to be SHapley Additive exPlanations (SHAP) (Figure 3(a)), particularly, it works well in explaining ensemble-based algorithms, like Random Forest and XGBoost. SHAP can estimate the marginal contribution of each input feature and thus can be transparent in diagnosis, classify biomarkers and accomplish prognosis. A second notable model-agnostic approach is Local Interpretable Model-Agnostic Explanations (LIME) that can offer helpful local explanations of specific predictions based on an approximation of the complex model locally, in an interpretable linear model .

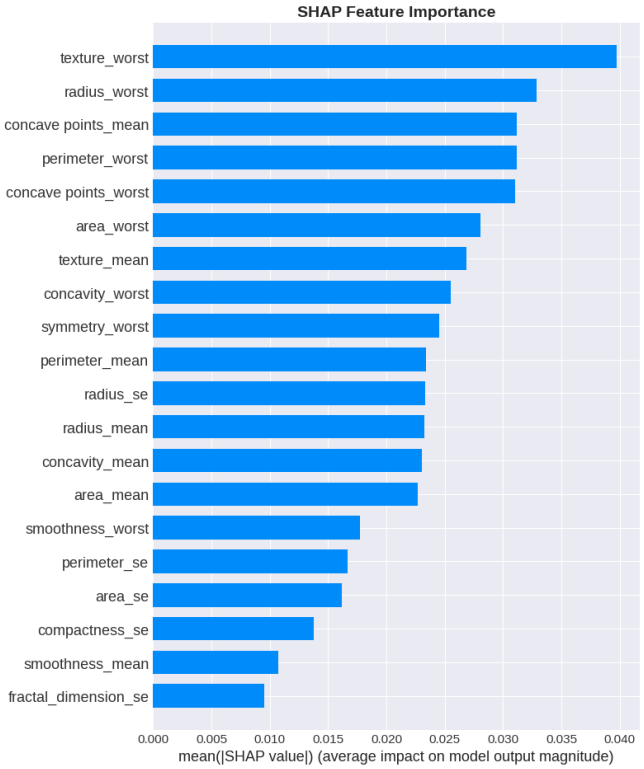

Figure 3(a) SHAP Feature Importance

2.5. Limitations in Current Explainability and Clinical Staging Integration

Despite these tremendous progress, the activities of XAI have been a menace in offering a whole person clinical support. The studies which are done nowadays typically utilise a single XAI strategy, and they do not pool together different strategies (e.g. visual localisation and ranking of feature contributions) in order to obtain the holistic and comprehensive interpretability to hybrid systems. Besides, the possibility to give a high rate of generalization on multi-sites data sets is one of the long-standing issues of the adoption of AI. High internal accuracy is often reported, a loss in external data (domain shifts) is a common risk, and must be explicitly overcome in development and high calibration . A successful system must have capacity to integrate and relate multi-modal data (imaging, clinical, molecular) productively to identify cases of straightforward benign/malignant data, and thereafter, incidences of subtle, detailed staging and prognosis forecasting .

3. Problem Definition

The diagnosis of the breast cancer in a timely manner and correctly has turned into one of the most burning issues in the modern medicine, whereas the field of imaging and diagnostic actions is evolving rapidly. However much

as the traditional process of diagnosing the existence of abnormalities within the breast, including mammography, biopsy and ultrasound (sonography) have played a pivotal role in concluding the initial clinical diagnosis, they are not without cost. This restates the fact that further studying is never too much, combination of collateral diagnostic tools and implementation of more appropriate and patient-based screening tools to have better early detection and, not the least, better patient outcomes in terms of breast cancer.The emerging imaging techniques like the MRI and the digital mammography are also contributing towards the issues. All these limitations imply that the medical diagnostics should be more automated and intelligent. This paper will take advantage of the strength with an aim of developing a conclusive, automated and intelligent system that will identify breast cancer at its early stages with high sensitivity and accuracy.

4. Conceptual Methodology and Data Strategy

The research design is displayed in the shape of a five step conceptual framework that targets the systematic change of the raw data input to the explicable and comprehensive diagnostic output. To ensure that diagnostic accuracy, transparency and computation efficiency is high, the methodology is a combination of classical image processing techniques with current advanced machine learning and machine learning techniques .

4.1. Data Acquisition and Preparation Protocol

The input training data in the machine learning is a critical component especially the quality and the size of the training data. The benchmarks datasets used to conduct this study will also be publicly available to ensure transparency and reproducibility to the bibliography and the quality of data used in this study will be maintained in a wise manner.

The datasets involved will involve:

1. Image Datasets: Image datasets such as CBIS-DDSM and INbreast have benchmark image datasets in mammography and are stocked with different forms of images such as normal, benign and malignant in nature. BreakHis data may also be applied in extracting features through histopathology.

2. Structured Datasets: Wisconsin Diagnostic Breast Cancer (WDBC) dataset which is a collection of structured numerical measurements to describe the Fine needle aspirate (FNA) cells nuclei (e.g., mean radius, texture, perimeter, and area) will be significant in the provision of the clinical feature stream.

The data augmentation methods are necessary to address the common problems of the healthcare datasets, i.e., the lack of data and unequal distribution of the classes. Further developed generative models can be utilized to come up with quality synthetic sets of data that can lessen the likelihood of breaching the privacy of the patients and widen the severity and range of the research foundation .

4.1.1. Preprocessing and Image Enhancement Techniques

Raw image captures are associated with high probabilities of artifacts, noise and fluctuating lighting conditions that are highly susceptible to medical image analysis . Therefore, it is important to have a stringent preprocessing pipeline to ensure that there is image homogeneity and an increase in important visual patterns that support the extraction of features.

The major steps in preprocessing are:

Normalization: Downsizing and greyscale normalization of CNN backbone inputs. Artifact and Noise Removal: Processes such as median filtering (MF) are used to remove random noise. Some clinical artifacts such as pectoral muscle tissue and label data ought to be separated and erased so as to prevent development of some irrelevant patterns by the model.

Contrast Enhancement: The algorithms applied in enhancing the image resolution and clarity are Contrast-Limited Adaptive Histogram Equalization (CLAHE) and unsharp masking (USM).

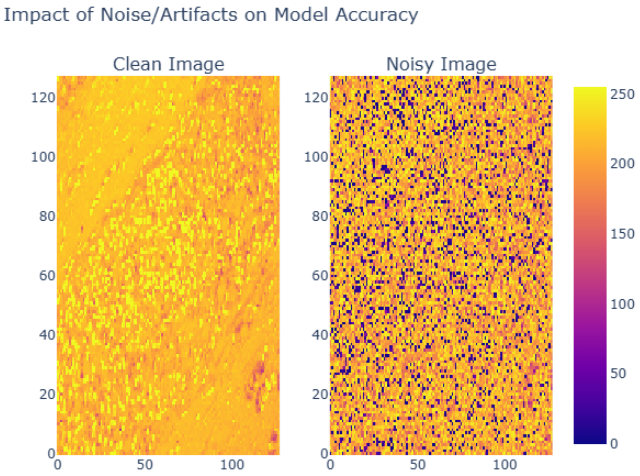

Figure 3(b) Impact of Noise/ Artefacts on Model Accuracy

High-level DL models are used, and this leads to the end feature representations being sensitive to small variations in the inputs . Image quality enhancement algorithms, including CLAHE and USM are needed to keep on emphasizing the diagnostic signal (tumor morphology) over the one of noise and acquired artifacts which is depicts in figure 3(b), in order to maximally improve the quality of the feature embeddings that the CNN generates.

4.2. Multimodal Data Integration Strategy

Multi-modal data integration has been shown to significantly enhance the performance of diagnostic and risk stratification. Research has shown the existence of multimodal models of imaging modules (mammography/ultrasound) with clinical metadata, which has accuracy of (90.1) that compares to that of a pathologist (92.7%). The hybrid approach is based on the two primary streams of data, namely, the proposed system will utilize them:

Image Features (Deep Feature Vector, DFV): Semantic features that are learned by CNN backbone automatically on a high level.

Organized Clinical/Molecular Features (Clinical Feature Vector, CFV): Prognostic variables, e.g.: a numerical value of WDBC-type analyses, patient characteristics, and, in certain cases, the status of molecular analysts (ER/PR/HER2).

The DFV and the CFV must also be melded into a single one that will be optimized as a Hybrid Feature Vector (HFV). It is a junction of fusion that is preceded by CNN feature extraction procedure but then the final ML classification unit. Table 1 shows The overall research implementation is defined by a rigorous five-phased methodology. The following ML classification can be learned based on this architecture learning the complex, non-linear interactions that exist between the purely morphological indicators as derived based on the visuals and the structured prognostic factors as derived based on the clinical reports together .

Table 1 The overall research implementation is defined by a rigorous five-phased methodology:

Phase	Major Activity	Goal	Reference
I	Data Acquisition & Preprocessing	Establish bibliographic transparency; ensure image homogeneity and signal quality.	[25]
II	Hybrid Feature Extraction	Convert raw image data into semantically meaningful representations; compile the Hybrid Feature Vector (HFV).	[26]

III	Model Development & Training	Build and train the hybrid DLE-ML classification system using cross-validation.	[27]
IV	Performance Evaluation & Optimization	Compare hybrid model results against traditional methods; optimize using ensemble techniques and robust metrics (AUC, F1-score).	[28]
V	XAI Implementation & Staging Output	Apply dual-framework XAI to ensure transparency; generate explainable multi-class staging classifications.	[29]

This type of a stepwise model will ensure a gradual improvement that commences with the necessary foundation of clean data and proceeds to the next step of developing a smart system that is capable of identification, feature analysis, normalization, system model training and, a proper evaluation.

5. Hybrid Architecture for Feature Extraction and Robust Classification

The combination framework has its advantages in the hybrid structure whose goal was to selectively exploit the ability of the advanced machine learning in mapping the characteristics of an item and the boundary optimization of the conventional machine learning.

5.1. The Advanced machine learning Feature Extractor (DLE): Leveraging Transfer Learning

The Convolutional Neural Network (CNN) of this model is called Deep Feature Extractor (DLE). Its primary process is automated seizure of hierarchies of space of the already processed medical images. This circumvents all the pitfalls of the manual or handcrafted feature extraction methods and acquires the sophisticated pattern recognition capabilities of DL.

The transfer learning will be done with a pre-trained CNN architecture, e.g., ResNet50 or VGG16, to accelerate the effectiveness of the training and to learn new knowledge based on the information that already belongs to the knowledge bases. The concept of transfer learning is highly efficient in medical imaging as it does not demand very huge labeled data sets that would be demanded in the training process so as to train deep neural networks. In the feature extraction mechanism, the Softmax classification layer of the pre-trained CNN will be removed. The output of the penultimate layer (the last layer before the classification head) is then reaped and it gives the high dimensional feature vector (DFV) but contains the high-level semantic properties of the lesion under imaging. The given DFV finding provides the potent visual-based knowledge, which is necessary to drive the subsequent step in the classification procedure, in order that the model will exploit the most salient visual clues in the diagnosis.

5.2. Feature Selection and Optimization Layer

It turns out that once the DFV is generated and mixed with the CFV (clinical data), then the Hybrid Feature Vector (HFV) is typically in a high-dimensional space. This space can be informationally rich, though can be computationally costly and can have much noise, which can have an adverse impact on the performance of the next ML classifier. A critical Feature Selection and Optimization Layer is, therefore, introduced.

The methods that can be initially implemented on the DFV in order to reduce the number of dimensions include Principal Component Analysis (PCA) which would ensure the minimization of the possibilities of overfitting without compromising the most informative content. However, an ultra-modern approach, wherein SHAP values are integrated with Recursive Feature Elimination (RFE), i. e. SHAP-RFE, will be employed so as to give as much transparency and predictability as possible. SHAP-RFE then selects and picks the most effective in the joint HFV. This process results in the improved interpretability of the model and hence accuracy in the prediction process since the emphasis of the classifier is on the variables which have been identified as making a significant contribution to the outcome in terms of marginal contribution. This layer is necessitated by the fact that the high dimensional DFVs might overpower the traditional ML classifiers. The SHAP-RFE approach ensures that the resulting ML model is displayed, having a small robust number of features, whereby the input of both image-based and clinical variables is optimized and has been pre measured, which constitutes the foundation of opaque classification.

5.3. The Machine Learning Classification Module

The best HFV is inputted in the last Machine Learning Classification Module. The reason behind the choice of the traditional ML algorithms at this stage is due to their strength, interpretability, and generalized output. Support Vector Machine (SVM) algorithm is accorded a high priority particularly when one requires to carry out binary classification (benign/malignant). The advantage of SVM is that it is likely to be effective in high dimensional feature space by classifying the optimal hyperplane that has the largest margin between the two classes. The experience of multi-class staging problems is that it would be preferable to use ensemble classifiers, such as Random Forest (RF) or XGBoost, since it is the property of multi-class prediction that levelizes the result and decreases the difference in predictions which is shown in table 2. The interconnection of the DLE and the ML classifier possess the following architecture advantages:

1. Robustness and Generalization: The fact that the SVM is used as the classification head diminish the likelihood of overfitting experienced by the Softmax based classifiers of end-to-end CNNs during their training on small medical datasets. The design option leads to an increased generalizability of different datasets.
2. High Accuracy: CNN-SVM hybrid model is highly accurate and based on accuracy has a long-term accuracy of over 95 percent classification performance that is high when compared to the single CNNs or the traditional ML models.
3. Intuitively Decodable Pathway: This is so because the non-transparent DL process is transformed into a transparent workflow since architectural division is involved. It is evident that the transformation of the DFV into the classical ML model establishes a clear and robust stance on which model-agnostic XAI (SHAP/LIME) can be used to perform the end classification decision successfully transforming the DL black box into a responsibility output.

Table 2 The framework's components are summarized in the following table, detailing the function and output of each stage:

Phase	**Component/Algorithm**	**Function**	**Output/Feature Type**	**Citation**
Deep Feature Extraction (DFE)	Fine-tuned CNN (e.g., ResNet50/VGG16)	Automated extraction of complex spatial features from medical images.	Deep Feature Vector (DFV)	[30]
Clinical Data Acquisition	EHR/WDBC numerical features	Acquisition of structured, prognostic patient data.	Clinical Feature Vector (CFV)	[31]
Feature Fusion & Optimization	Concatenation + SHAP-RFE/PCA	Dimension reduction and selection of most predictive features.	Optimized Hybrid Feature Vector (HFV)	[32]
Classification (ML)	Support Vector Machine (SVM) / Ensemble (RF)	Robust classification (Detection and Staging) using the optimized boundary.	Prediction: Malignant/Benign + Staging Class	[33]
Visual Explainability	Grad-CAM/Grad-CAM++	Localization of salient image regions influencing the DFE output.	Heatmaps (Visual Evidence)	[34]

Feature Explainability	SHAP	Quantifies individual feature contribution to the final ML decision.	Feature Importance Scores	[35]

6. Implementation of Explainable AI for Clinical Trust

6.1. Visual Interpretability via Class Activation Mapping (Grad-CAM/Grad-CAM++)

Visual interpretability techniques are Class Activation Mapping, such as Grad-CAM or its enhanced counterpart such as Grad-CAM++. The operation of this model is on CNN signature of the hybrid model but, in fact, the focus is done on the layers just before the feature extraction output (DFV) is passed to the ML classifier.

Mechanism and Output: Grad-CAM creates a heatmap over the medical image as an overlay to the original image. This heatmap suggests the locations of the specific areas, the groups of pixels, that the deep feature calculation of the CNN considered the most important, intuitively. In this case, with breast cancer; they can demonstrate that the DLE is taking the correct decision when it focuses on pathological markers such as masses, calcifications, or tissue asymmetries, and not on artifacts of the confounding characteristics.

Clinical Relevance: Visual evidence generated by Grad-CAM is directly associated with the radiologic requirement of localizing and determining the presence of tumor and T-category (Tumor size and extent) of staging. This openness leads to a high level of clinical trust as it provides an explanation of the characteristics of the image in real-time and is not very hard in understanding that makes it possible to reach a consensus between the output of the AI generated and the human radiological interpretation. The methodological choice of Grad-CAM before the classification layer is meant to ensure that the interpretability is intoxicated to the actual stage of deep feature extraction and does not degrade the visual description thoughtfully as would otherwise have been the case had it been used after the classification.

6.2. Feature Contribution Analysis using SHapley Additive exPlanations (SHAP)

By adopting the model-agnostic SHAP framework, the ultimate execution of the diagnostic result becomes responsible. The classification component (SVM or Random Forest) receives SHAP which receives the Hybrid Feature Vector (HFV).

Mechanism and Quantification: SHAP calculates the fair marginal contribution of every element in the HFV to the numeric components of the DFV and the structured clinical features of the CFV to the overall prediction score of a specific patient.

This is possible in order to:

1. **Global Interpretability:** Rating the factors that are most predictive of the whole data set (e.g. is tumor perimeter or ER status mostly predictive).
2. **Local Interpretability:** To provide a particular, quantifiable, response in regard to why this particular patient was diagnosed with Malignancy or Stage II. There is a close correlation between the two XAI approaches: Grad-CAM tests the validity of the visual pathology (T-category component), and SHAP tests the contribution to the final staging decision of the non-imaging, prognostic factors (N/M status, molecular markers). This responsible, integrative definition is a step needed in order to achieve clinical integration.

Ethical, legal, and trust issues are the key challenges to the successful implementation of AI in oncology (figure4(a)), which are major barriers to non-technical issues. The obstacles include the issues of prejudice, data ownership, privacy, lack of responsibility, and the ambiguity of regulatory norms.

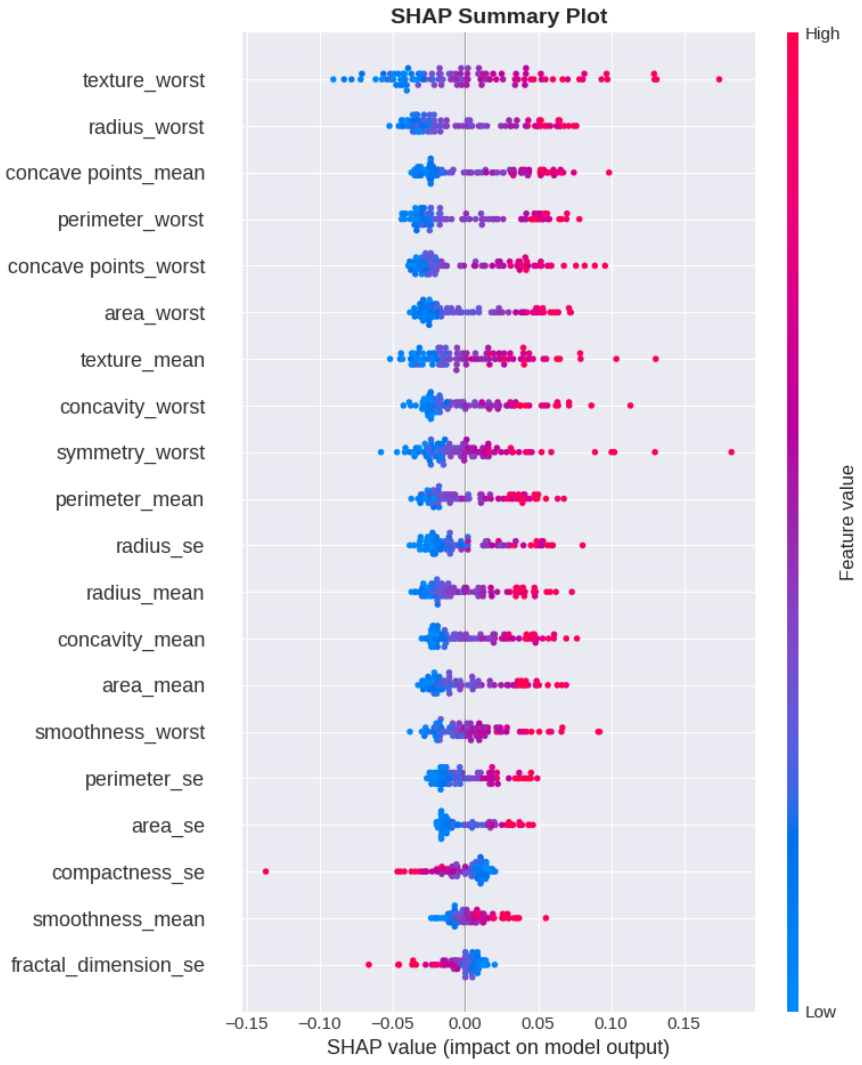

Figure 4(a) SHAP Summary Plot

6.3. Ethical and Trust Frameworks for XAI Adoption

Non-technical barriers to the successful implementation of AI in oncology include major obstacles to the ethical, legal, and trust issues. The barriers consist of the problems of bias, ownership of data, privacy, absence of accountability, as well as uncertainty concerning regulatory standards.

The issues are directly resolved by the peculiar transparency inherent in the dual XAI model. The ability to explain aids in the reduction of the observer bias and makes the use of diagnostic models ethically viable since clinicians are able to audit the rationale behind the model. The explanations generated by SHAP and Grad-CAM in a straightforward form enable meeting the requirement and getting a key audit trail, which will allow to trace the origins of the unintended errors, technological failures, or biases towards particular features rather than an incoherent algorithm which is shown in table 3. This transparency is both a legal and clinical necessity that fosters the trust and confidence required to embrace AI by the medical professionals and patients.

***Table 3**: Key Explainable AI (XAI) Techniques and Their Role in Hybrid Model Interpretation*

XAI Technique	Model Scope	Primary Function in Hybrid Model	Clinical Benefit	Citation
Gradient-weighted Class Activation Mapping (Grad-CAM)	Advanced machine learning (CNN layers)	Visualizes salient image regions leading to DFE output.	Provides visual radiological evidence for lesion localization (T component).	[36]
SHapley Additive exPlanations (SHAP)	Model-Agnostic (ML Classification layer)	Quantifies marginal contribution of hybrid features (image + clinical) to the final prediction.	Ranks prognostic factors (e.g., node involvement, tumor size); ensures accountability.	[37]
Local Interpretable Model-agnostic Explanations (LIME)	Model-Agnostic (ML Classification layer)	Generates local linear explanations for specific patient outcomes.	Justifies individual diagnostic decisions, improving physician comprehension.	[38]

7. Discussion on Staging Classification and Clinical Prognosis

The fact that the framework manages to classify the stage of breast cancer successfully is a significant development in clinical functionality going beyond binary detection. It is important to stage properly the prognosis of patients and the decision whether they should be treated further.

7.1. Translating Predictive Outputs to AJCC/TNM Staging System

Breast cancer has a universally accepted system of Staging that is designed on the American Joint Committee on Cancer (AJCC) system TNM where three basic variables are considered such as Tumor size (T), the close lymph nodes (N) involvement and distant Metastasis (M) . The proposed ML Classification Module is informed to give a multi-class prediction result of the clinical stages (I, II, III and IV) on the basis of the complete information possessed in the HFV.

To have an effective staging classification, the model not only finds out the existence of a lesion, but also the dimensions, spread, and potential spread characteristics that can be recorded in the clinical narrative and informal communications at the Electronic Health Record (EHR).

7.2. Differentiation Between Early Stages (I vs. II) based on Model Attributes

The proper differentiation between Stage I and Stage II breast cancer is essential as it determines the circumscribed nature of the disease and has an impact on the treatment.

- Stage I: This is the initial stage of the invasive cancer that has the tumor size of 2 centimeters or less (T1) and does not show the presence of the spread to the lymph nodes in close surroundings (N0) or elsewhere (M0).
- Stage II: This is a stage that is considered as locally advanced usually the enlarged tumor size (T2, 2 to 5 cm; or T3, 5 cm and above) or the localised tumor described as the spread to the lymph nodes (N1, 1-3 lymph nodes in the armpit). To effect this distinction, the hybrid model must be founded on two significant groups of characteristics: the T-status (size and morphology of tumor), which can be estimated with extremely high accuracy by the DFV including morphological characteristics like Worst Radius and Area and the N-status (involvement of the lymph nodes), which must be determined using regional imaging data and ordered clinical input (CFV).

SHAP analysis has to be used in order to confirm the clinical reasoning of the model on such subtle cases. SHAP shows the way the model is performing well in balancing the feature (e.g. tumor size according to image extraction) which is shown in figure 4(b) or the predicted node involvement status (according to clinical input) to ideally distinguish between Stage I and Stage II . Such open weighting scheme implies a layer of critical quality control of the staging predictions of the model.

Figure 4(b) Class Distribution In the Dataset

Table 4 outlines the classification targets and the specific feature types leveraged by the model:

***Table 4**: Classification Mapping to AJCC Clinical Staging Criteria*

AI Output Class	Target AJCC Stage	T (Tumor) Parameter	N (Node) Parameter	M (Metastasis) Parameter	Feature Influence (SHAP Focus)
Class 0	Stage 0 (In Situ)	Tis (< DCIS)	N0	M0	Image morphology (confined growth, local shape)
Class 1	Stage I (Early Invasive)	T1 (< 2 cm)	N0	M0	Mean Radius, Texture, Compactness 5
Class 2	Stage IIA/IIB (Locally Advanced)	T2 (2-5 cm) or T3 (>5 cm)	N0 or N1 (1-3 nodes)	M0	Worst Radius/Area (T size), Clinical N Status
Class 3	Stage III (Regional Advanced)	T4 (Invading chest wall/skin)	N2 (4-9 nodes)	M0	T4 morphological features (Grad-CAM), high N count

Class 4	Stage IV (Metastatic Disease)	Any T	Any N	M1 (Distant spread)	M1 clinical indicators, prognostic biomarkers

7.3. Projected Role in Personalized Prognostic Risk Assessment and Treatment Planning

The application of the ML is rather prospective in establishing the prognosis and defining the sequence of the disease progression and the adjustment of the treatment strategy based on the patient characteristics. The models of clinical/molecular integrative AI have been proved to enhance the reliability of the prediction up to 9 per cent that could be utilized to aid in creating individualized treatment plans.

The hybrid model will allow accuracy-based oncology by incorporating its molecular pathways (including that of MAPK, PI3K-Akt, Wnt, and TGF-$\beta) into the CFV. Risk of recurrence or probability of survival are only some of the prognostic factors that the model can predict not only the current stage but also the individual prognostic factors. The explainable output, the interpretable model of the explainable output that is the SHAP ranking of critical features, guides the process of oncologists by quantifying the risk factors of the patient, and therefore tailored treatment regimens, e.g. the probability of responding to a specific chemotherapy regimen.

8. Expected Outcomes, Impact, and Future Directions

This explicable hybrid ML-DL structure is an unavoidable future of the evolution of smarter, even more effective, and clinically relevant diagnostic systems. The outcomes will be of immense value to the research, clinical practice, and population health.

8.1. Projected Performance Metrics and Robustness

Based on the literature that is available on the subject of ensemble methods and hybrid architecture, the projected model will most likely achieve even higher performance measures, perhaps 95 percent accuracy and sensitivity in the early detection and staging of breast cancer. The evaluation of the performance will be high and cross-validation and the application of standard measures like Accuracy, Sensitivity, Specificity, F1-score and Area Under curve (AUC) will be applied.

The big advantage of the classical ML classifier (SVM/RF) as the ultimate decision layer is the increment of the generalizability and strength of the system. Advanced machine learning-only systems tend to be less accurate when applied to external data due to the existence of domain shifts (vendors, or acquisition policy and population heterogeneity). Upon employing a robust ML classifier, that was trained on a highly refined set of features (HFV), the hybrid model is built with high consistency and reliability in the wide variety of clinical settings.

8.2. Addressing Ethical, Regulatory, and Trust Barriers to Adoption

The explicit connection of a dual XAI model (Grad-CAM and SHAP) is linked with the historical issues of trust and accountability of black-box AI. Clinical and regulatory acceptability is based on localization and an amount of justification (feature contribution) in offering visual evidence, and provides a basis of audit trail to every diagnostic judgment. It belongs to the main requirements of responsibility in AI implementation in healthcare since this intervention involves transparency.

The aspect of data privacy and reduction of bias also need to be considered in ethical deployment. With the help of the best data generation schemes, such as Tabular Variational Autoencoder (TVAE), research severity becomes simpler and the possibility of breaching patient privacy also decreases. During model optimization, transparency and fairness systems must be integrated into the model to ensure that the system creates equity in diagnosis.

9. Conclusion and Future Research Directions

The authors have demonstrated in this study that the development of a coherent describe explainable hybrid machine learning and advanced machine learning architecture to identify and fully stage early breast cancer is technically feasible and clinically necessary. The proposed methodology will exploit the spatial capacity of CNNs to recover deep features, stabilize classification with a potent traditional ML model, e.g. SVM and Random Forest, and ensure that it is clinically suitable with dual XAI structure where Grad-Cam is utilized to retrieve visual evidence and SHAP is utilized to retrieve quantifiable accountability. This subjective and high variability human interpretation to automated, high accuracy and transparent intelligent systems are predicted to transform the paradigm of the breast cancer diagnosis radically and enhance the effectiveness and accuracy of the early diagnosis. The model obtained must be a valid, robust, and explainable AI-powered algorithm. The availability of categorizing the breast cancer stages with the assistance of explainable parameters will provide clinicians with the intelligible information on the disease course, enabling to plan an individual medical intervention and improve the overall prognosis of a patient.

The research in the future would focus on streamlining the framework to allow it to be used and implemented in low-resource environments in a real-time environment, possibly by exploring lightweight CNNs. Notably, further enhancement of the generalization process, through the application of such techniques as multi-scale feature and domain adaptation, must be under consideration to ensure both the safety and good performance even in the presence of the variability inherent in the different imaging protocols of multi-institutions. Multi-site prospective assessments and standardized reports of performance must replace validation in order to establish the establishment of the model in daily clinical practice.

Data Availability

Data that is accessible in a public repository, which does not issue DOIs, was analyzed in this study. The publicly available dataset can be found at the following link: https://www.kaggle.com/datasets/jocelyndumlao/biglycan-breast-cancer.

REFERENCES

1. A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.
2. C. Gomes, J. Geels, T. P. A. Debray, A. Malekzadeh, F. W. Asselbergs, and M. Linschoten, “Risk prediction models for cancer therapy related cardiac dysfunction in patients with cancer and cancer survivors: systematic review and meta-analysis,” *BMJ*, vol. 390, 2025, doi: 10.1136/bmj-2025-084062.
3. I. O. Nwosu, “A Comprehensive Review and Meta-Analysis of Breast Cancer Gene Expression Data: Understanding the Molecular Complexity of a Silent Epidemic,” 2024. [Online]. Available: https://scholarsarchive.byu.edu/etd/11017
4. P. H. Abreu, M. S. Santos, M. H. Abreu, B. Andrade, and D. C. Silva, “Predicting breast cancer recurrence using machine learning techniques: a systematic review,” *ACM Computing Surveys (CSUR)*, vol. 49, no. 3, pp. 1–40, 2016.
5. X. Qian *et al.*, “A multimodal machine learning model for the stratification of breast cancer risk,” *Nat Biomed Eng*, vol. 9, no. 3, pp. 356–370, 2025, doi: 10.1038/s41551-024-01302-7.
6. S.-H. Cho, J. Jeon, and S. Il Kim, “Personalized medicine in breast cancer: a systematic review,” *J Breast Cancer*, vol. 15, no. 3, pp. 265–272, 2012.
7. T. Arravalli *et al.*, “Detection of breast cancer using machine learning and explainable artificial intelligence,” *Sci Rep*, vol. 15, no. 1, p. 26931, 2025, doi: 10.1038/s41598-025-12644-w.
8. S. Raghuvanshi, Y. Tiwari, and A. Yadav, “Decoding Breast Cancer Detection: An Explainable AI Framework for Transparent Diagnosis,” in *2025 Seventh International Conference on Computational Intelligence andCommunication Technologies (CCICT)*, IEEE, 2025, pp. 481–488.
9. S. Hussain *et al.*, “Breast cancer risk prediction using machine learning: a systematic review,” 2024, *Frontiers Media SA*. doi: 10.3389/fonc.2024.1343627.
10. M. S. K. Inan, S. Hossain, and M. N. Uddin, “Data augmentation guided breast cancer diagnosis and prognosis using an integrated deep-generative framework based on breast tumor’s morphological information,” *Inform Med Unlocked*, vol. 37, p. 101171, 2023.
11. Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, “Comparative analysis of breast cancer detection using machine learning and biosensors,” *Intelligent Medicine*, vol. 2, no. 2, pp. 69–81, 2022, doi: https://doi.org/10.1016/j.imed.2021.08.004.
12. X. Wang *et al.*, “Intelligent hybrid deep learning model for breast cancer detection,” *Electronics (Basel)*, vol. 11, no. 17, p. 2767, 2022.
13. J. Zhu *et al.*, “An integrated approach of feature selection and machine learning for early detection of breast cancer,” *Sci Rep*, vol. 15, no. 1, p. 13015, 2025, doi: 10.1038/s41598-025-97685-x.

14. N. A. Chowdhury, L. Wang, L. Gu, and M. Kaya, “Machine Learning for Early Breast Cancer Detection,” *J Eng Sci Med Diagn Ther*, vol. 8, no. 1, Jul. 2024, doi: 10.1115/1.4065756.
15. A. La Moglia and K. Mohamad Almustafa, “Breast cancer prediction using machine learning classification algorithms,” *Intell Based Med*, vol. 11, p. 100193, 2025, doi: https://doi.org/10.1016/j.ibmed.2024.100193.
16. K. M. M. Uddin, N. Biswas, S. T. Rikta, and S. K. Dey, “Machine learning-based diagnosis of breast cancer utilizing feature optimization technique,” *Computer Methods and Programs in Biomedicine Update*, vol. 3, p. 100098, 2023, doi: https://doi.org/10.1016/j.cmpbup.2023.100098.
17. K. A. Kovács *et al.*, “Machine learning prediction of breast cancer local recurrence localization, and distant metastasis after local recurrences,” *Sci Rep*, vol. 15, no. 1, p. 4868, 2025, doi: 10.1038/s41598-025-89339-9.
18. Md. M. Islam, Md. R. Haque, H. Iqbal, Md. M. Hasan, M. Hasan, and M. N. Kabir, “Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques,” *SN Comput Sci*, vol. 1, no. 5, p. 290, 2020, doi: 10.1007/s42979-020-00305-w.
19. V. Nemade and V. Fegade, “Machine Learning Techniques for Breast Cancer Prediction,” *Procedia Comput Sci*, vol. 218, pp. 1314–1320, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.110.
20. D. Kovacs, K. Håkansson, J. Rasmussen, B. Fischer, F. Andersen, and C. Ladefoged, *Non-inferiority of Automated Deep Learning-Based 18F-FDG PET/CT Tumour Volume Compared to Manual GTV for Prognostic Modelling in Head and Neck Cancer*. 2025. doi: 10.21203/rs.3.rs-7030494/v1.
21. L. Luo *et al.*, “Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions,” *IEEE Rev Biomed Eng*, vol. 18, pp. 130–151, 2025, doi: 10.1109/RBME.2024.3357877.
22. M. Radak, H. Y. Lafta, and H. Fallahi, “Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies,” *J Cancer Res Clin Oncol*, vol. 149, no. 12, pp. 10473–10491, 2023, doi: 10.1007/s00432-023-04956-z.
23. D. Zuo, L. Yang, Y. Jin, H. Qi, Y. Liu, and L. Ren, “Machine learning-based models for the prediction of breast cancer recurrence risk,” *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 276, 2023, doi: 10.1186/s12911-023-02377-z.
24. S. Sadr *et al.*, “Gold nanobiosensors and Machine Learning: Pioneering breakthroughs in precision breast cancer detection,” *European Journal of Medicinal Chemistry Reports*, vol. 12, p. 100238, 2024, doi: https://doi.org/10.1016/j.ejmcr.2024.100238.
25. J. S. Ahn *et al.*, “Artificial intelligence in breast cancer diagnosis and personalized medicine,” *J Breast Cancer*, vol. 26, no. 5, p. 405, 2023.
26. J. Majidpour and H. Beitollahi, “A Comprehensive Examination of Machine Learning and Deep Learning Approaches for Breast Cancer Detection, Classification, Segmentation, Augmentation, and Feature Selection,” *Archives of Computational Methods in Engineering*, 2025, doi: 10.1007/s11831-025-10359-9.
27. N. Kavitha, P. Madhumathy, R. M. Prasad, and D. N. Chandrappa, “Machine learning technique for breast cancer detection and classification,” *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, p. 16, 2025, doi: 10.1007/s44379-025-00018-y.
28. A. Thalor, H. Kumar Joon, G. Singh, S. Roy, and D. Gupta, “Machine learning assisted analysis of breast cancer gene expression profiles reveals novel potential prognostic biomarkers for triple-negative breast cancer,” *Comput Struct Biotechnol J*, vol. 20, pp. 1618–1631, 2022, doi: https://doi.org/10.1016/j.csbj.2022.03.019.
29. H. L. P, S. Ravi, and S. A, “Breast Cancer Detection Using Machine Learning in Medical Imaging – A Survey,” *Procedia Comput Sci*, vol. 239, pp. 2235–2242, 2024, doi: https://doi.org/10.1016/j.procs.2024.06.414.
30. M. H. Attiwi *et al.*, “Unremarkable AI: Intelligent Decision Support in Critical Clinical Decision-Making,” in *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, IEEE, 2024, pp. 610–616.
31. A. Khang, “AI-Driven Innovations in Digital Healthcare: Emerging Trends, Challenges, and Applications: Emerging Trends, Challenges, and Applications,” 2024.
32. A. Ghavidel and P. Pazos, “Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review,” *Journal of Cancer Survivorship*, vol. 19, no. 1, pp. 270–294, 2025, doi: 10.1007/s11764-023-01465-3.
33. T. G. Krishna and M. A. A. Mahboub, “Improving Breast Cancer Diagnosis with AI Mammogram Image Analysis,” *Medinformatics*, vol. 2, no. 3, pp. 195–204, Nov. 2024, doi: 10.47852/bonviewMEDIN42023806.
34. H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, and G. Cai, “Classification Prediction of Breast Cancer Based on Machine Learning,” *Comput Intell Neurosci*, vol. 2023, no. 1, p. 6530719, Jan. 2023, doi: https://doi.org/10.1155/2023/6530719.
35. S.-J. Sammut *et al.*, “Multi-omic machine learning predictor of breast cancer therapy response,” *Nature*, vol. 601, no. 7894, pp. 623–629, 2022, doi: 10.1038/s41586-021-04278-5.
36. J. Thagaard *et al.*, “Pitfalls in machine learning-based assessment of tumor-infiltrating lymphocytes in breast cancer: A report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer,” *J Pathol*, vol. 260, no. 5, pp. 498–513, Aug. 2023, doi: https://doi.org/10.1002/path.6155.
37. M. S. K. Inan, S. Hossain, and M. N. Uddin, “Data augmentation guided breast cancer diagnosis and prognosis using an integrated deep-generative framework based on breast tumor’s morphological information,” *Inform Med Unlocked*, vol. 37, p. 101171, 2023.
38. Sheng Zhou, Chujiao Hu, Shanshan Wei, and Xiaofan Yan, “Breast Cancer Prediction Based on Multiple Machine Learning Algorithms,” *Technol Cancer Res Treat*, vol. 23, p. 15330338241234792, Apr. 2024, doi: 10.1177/15330338241234791.

39. H. E. Saroğlu *et al.*, “Machine learning, IoT and 5G technologies for breast cancer studies: A review,” *Alexandria Engineering Journal*, vol. 89, pp. 210–223, 2024, doi: https://doi.org/10.1016/j.aej.2024.01.043.
40. A. Ghasemi, S. Hashtarkhani, D. L. Schwartz, and A. Shaban-Nejad, “Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review,” *Cancer Innovation*, vol. 3, no. 5, p. e136, Oct. 2024, doi: https://doi.org/10.1002/cai2.136.