Kerbala Journal for Engineering Sciences/ Vol. 06, No. 01 (2026) ISSN: 2709-6718
https://doi.org/10.63463/kjes1209

Kerbala Journal for Engineering Science
https://kjes.uokerbala.edu.iq

AI-Driven Insights in Breast Cancer: A Deep Learning Review Integrating Medical Imaging and Genomic

Rusul Hatem Hanoon [1,] *, Hameed Rasool Farhan [1], Arwa H. Mohammed Taqi [1]

[1] Department of Electrical and Electronic Engineering, College of Engineering, University of Kerbala, Kerbala, Iraq.

* Corresponding author, Email: rusul.hatem@s.uokerbala.edu.iq

Received: 20 September 2025; Revised: 15 March 2026; Accepted: 26 March 2026; Published:31March 2026

Abstract

One of the main causes of cancer-related mortality among women globally is breast cancer. It presents a significant barrier to community public health, and better patient retrieval and intervention depend on early diagnosis and careful planning. Since breast cancer is one of the most significant or exclusive predictors of survival, machine learning and deep learning techniques have a great potential to increase the accuracy of both classification and identification. With a focus on visual data (mammograms) and molecular data (genomic biomarkers), this study offers a thorough summary of current developments in early breast cancer detection strategies. Strong techniques for learning from diverse data sources are provided by conventional deep learning models. However, its scalability, interpretability, and efficiency are dubious when applied to large-dimensional biological datasets. It has been suggested that topical research on explainable AI could improve the precision, openness, and dependability of early breast cancer diagnosis. By bypassing the black-box aspect of traditional AI systems, these technologies are intended to improve the slide and clinical credibility of figures on deep learning models. In order to speed up data analysis and improve prediction skills in healthcare imaging, researchers have also investigated well-established machine learning techniques. The development of AI-driven diagnostic tools that are more user-friendly and appropriate for practical medical applications is made possible by this advancement.

Keywords: CNN, deep learning, genetics, breast cancer, mammography, prognosis prediction.

1. Introduction

According to the World Health Organization (WHO), breast cancer is the most common cancer amongst women worldwide and the foremost cause of cancer-related death in women [1]. Due to its high incidence and the critical role early diagnosis plays in improving patient survival outcomes, breast cancer continues to be one of the most researched diseases. Early and accurate identification is a major difficulty in breast cancer society since fatality rates are still strongly correlated with the stage at which the disease is discovered, despite significant advancements in treatment techniques. Practically speaking, this difficulty has spurred much research into developing sophisticated screening and diagnostic techniques that can identify cancers at their earliest and most curable stages [2][3]. Mammography has been time-honored as the main imaging modality for breast cancer screening and early detection [4]. It was brought to widespread clinical attention for its capacity to detect fundamental abnormalities and microcalcifications subsequent to the emergence of clinical symptoms. Despite the statement that mammography amplification has been technically static, technically difficult, and has drawbacks, equally reduced sensitivity in dense breast tissue, image noise, tissue overlay, and inter-observer variability [5]. In line with these limitations, developments in computational and data-driven techniques, for instance, deep learning and machine learning, have enriched the diagnostic capabilities, including dependability, classification accuracy, and feature extraction from mammography pictures [2]. Individually, advances in genomic machinery have assisted in the identification of molecular and genetic markers associated with breast cancer risk, tumor antagonism, and treatment response. Genomic biomarkers provide complementary information that is not captured by imaging alone, such as mutations in powerlessness genes and differences in gene-expression profiles. Happily, women with breast cancer have a better chance of survival when it is detected early [6]. Shows that effective defensive treatment can reduce the risk of breast cancer enduring clinical experience [7]. As a result, early identification of breast cancer enhances the possibility of therapy and improves survival. Over the years, continuing research and scientific activities have been decisive in improving the treatment of breast cancer because scientists have stayed conscious of the risks concomitant with concentrating on cancer from the start. Genetic data display high-dimensional, heterogeneous feature spaces that involve sophisticated modeling and integration techniques to be therapeutically significant from an engineering and data analysis standpoint. Although genetic markers have provided insights into cancer biology, their independent use for detection and diagnosis was limited by cost, data complexity, and clinical interpretability. By providing a thorough and critical assessment of mammography and genetic biomarker-based breast cancer screening procedures [7].

The presented study examines recent developments in computational modeling, multimodal fusion procedures, and feature extraction and calculation, to highlight the profits and drawbacks of current methods and highlight unresolved issues. Additionally, by focusing on mechanical designs and methodological changes rather than straightforward medical knowledge, it assesses efforts to assist

professionals in developing more accurate, dependable, and clinically useful breast cancer diagnostic instruments.

2. Methods of Breast Cancer Detection

Breast cancer detection relies on two main systems: mammography and genomic biomarkers. An exact type of X-ray imaging technique, titled mammography, permits doctors to view the internal structures of the breast to detect and diagnose breast cancer early. Figure 1 shows several approaches to detecting breast cancer, each emphasizing different ways the disease is detected. Mammography has remained one of the most commonly used imaging modalities for breast cancer screening and, from a computational viewpoint, poses a challenging image analysis problem owing to low contrast, significant interpatient variability, and high noise levels [8][9].

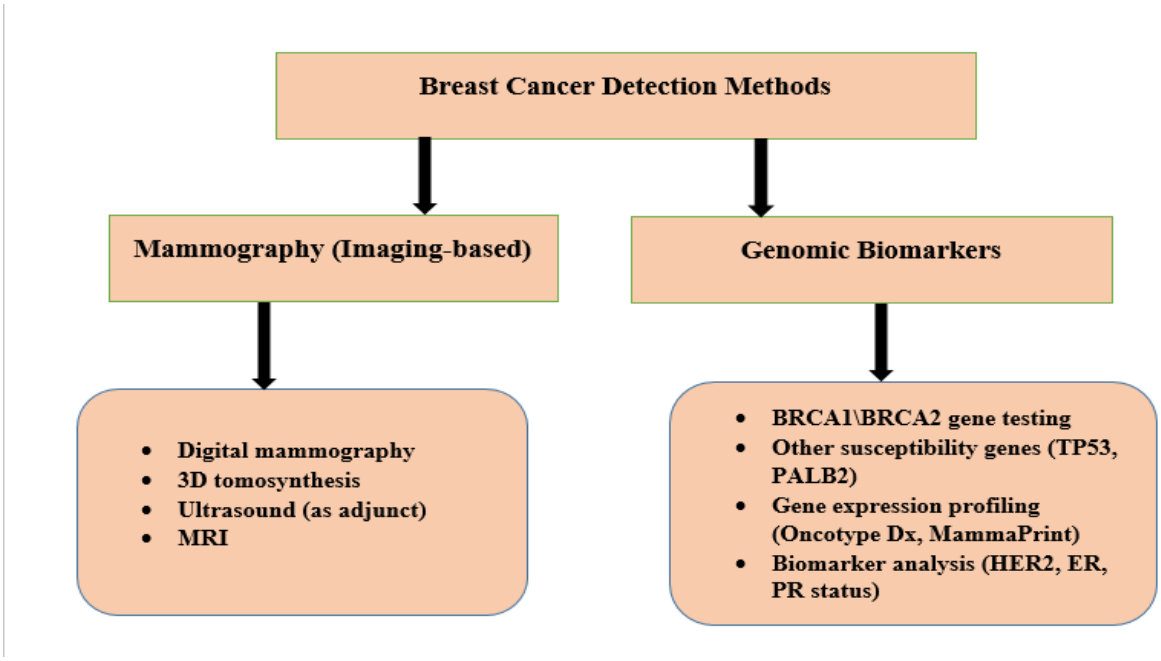

Figure 1. Breast cancer detection methods.

These physiognomies have attracted significant research into mechanical and AI-based diagnostic systems, particularly those that rely on deep learning methods for robust feature extraction and classification [7]. Genomic data can transport complementary molecular-level information that can improve diagnostic accuracy when integrated with imaging features [10]. Existing studies have recognized that deep learning can effectively model both mammographic images and high-dimensional genomic biomarkers, leading to more accurate and refined breast cancer detection [10]. Therefore, rather than focusing on their clinical or procedural features, this study highlights computational techniques that use genomes and mammography as data sources.

2.1. Breast Cancer Diagnosis Using Mammography

With the advent of fusion imaging, which integrates many imaging modalities with 3D mammography, imaging technologies have advanced significantly. The precision and effectiveness of breast cancer staging, diagnosis, and treatment planning have significantly improved thanks to these advancements.

2.1.1. Mammography Methods: Their Function in Early Diagnosis of Breast Cancer

The most popular imaging method for detecting breast cancer has been mammography. However, its clinical application as the primary source of data for computerized diagnosis has been greatly surpassed by its significance in subsequent studies [11][12]. From an engineering perspective, mammography images provide special difficulties because of their high resolution, low signal-to-noise ratio, low lesion prevalence, and straightforward class imbalance between normal and malignant cases [13]. Because of these features, comprehensive mammography has become a typical problem for the development and assessment of deep learning-based diagnostic systems and computer-aided diagnosis (CAD) [12] [14]. Recent developments in deep learning have superficially altered mammography-based detection procedures by enabling automatic feature learning from raw or little-processed pictures. Convolutional neural networks (CNNs) have proven effective in identifying masses and microcalcifications, particularly when paired with attention mechanisms and multi-scale feature extraction [15]. The creation of volumetric and slice-based deep learning models was prompted by pseudo-3D images of breast tissue, which further enhanced data complexity. DBT, or digital breast tomosynthesis [7][16].

Despite these advances, automated mammography analysis was challenging because of the lack of labeled datasets, the development of a number of information access systems, and the challenge of identifying tiny lesions like microcalcifications [17]. Pathetic learning and full-image classification without region-of-interest annotations were significant teething issues, according to recent research on transfer learning algorithms exhausting large-scale natural or medical image datasets [18]. Mammography was the primary research focus for walkable AI-driven breast cancer screening. These methods aim to improve oversimplification performance in the future, even though depressing the need for pricey expert annotations [19].

2.1.2. Image Processing and Feature Extraction for Mammography

Effective image preprocessing and feature extraction procedures were essential for applying deep learning successfully to mammographic studies. Subsequently, mammograms are typically stored in high-resolution DICOM files, which often surpass the memory and processing capacities of standard deep learning frameworks. To ensure proper data management and consistent model optimization, various preprocessing steps have been used, as explained in Figure 2 [20].

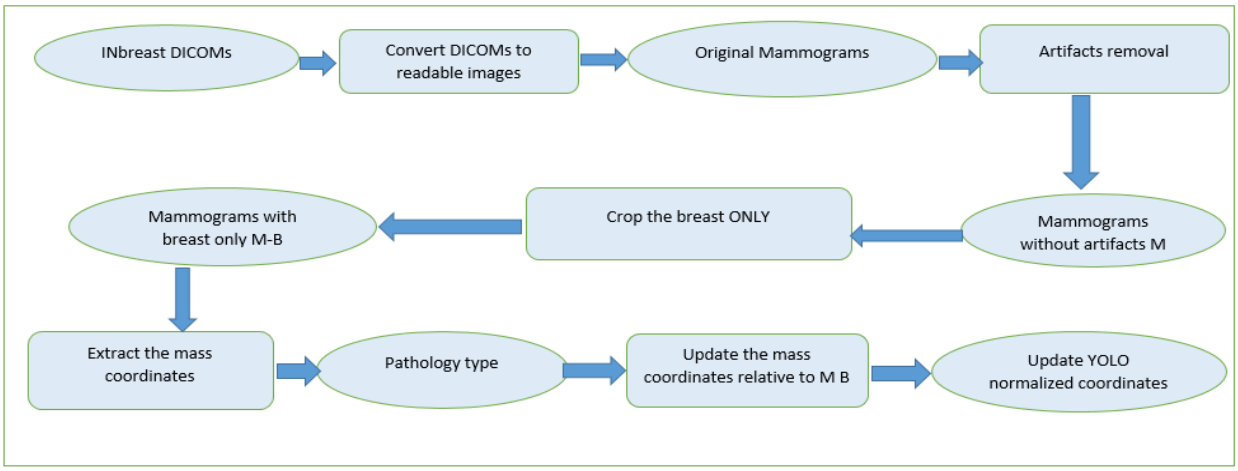

Figure 2. Mammograms preprocessing phase [20]**.**

Since the INbreast database is utilized in this figure, alternative datasets could be used instead; nonetheless, key processes such as eliminating artifacts, subduing context, segmenting the breast area, and normalizing intensity are essential. The small size and small differences in malignant findings relative to the total breast area make feature extraction in mammography more difficult [21].

Modern deep learning procedures use convolutional frameworks that effectively harvest hierarchical symbols directly from images, whereas previous CAD systems relied on region-focused evaluations and human feature engineering, which caused reliability problems [22]. Conversely, the lack of granular pixel or area labels in publicly obtainable data led to the development of alternative techniques, in place of patch-based training, multiple-instance learning, and end-to-end full-image classification. According to recent research, full-image learning combined with patch-based supervision enhances detection performance without compromising efficacy. Additionally, methods like transfer learning and self-supervised pre-training have improved model adaptability across various imaging modalities and decreased the need for manual annotation [2]. The computational main of mammography-based AI companies consists of preprocessing and representation methods that directly disturb clinical usability, scalability, and diagnostic reliability [15].

2.1.3. Challenges in Deep Learning for Mammography

Deep learning has been long-established remarkable outcomes in the diagnosis and detection of breast cancer by means of mammography. However, there are numerous significant problems with its extensive clinical application. There are a few huge, reliable, and high-quality mammography databases, but data input is difficult due to privacy constraints [11]. Also, creating generic copies is challenging because of variations in imaging gear, methods, and resolutions. A more difficult problem with annotation is that labeling needs accomplished radiologists, and even

in such circumstances, inter-observer variability makes ground truth questionable [17]. Technically, language and deep learning models frequently encounter overfitting, high processing costs, and domain-specific issues when applied to various datasets. The limited interpretability of CNNs and comparable architectures is another concern for clinicians seeking transparent, explainable AI solutions. Because models' strength is not able to generalize across a range of groups, which raises questions about bias and justice, integrating them into clinical procedures is still difficult [23]. The deployment is constrained by moral factors like acceptance, resistance, trust, and obligation. Creating multi-institutional standardized datasets, reducing data scarcity through joint and transfer learning, and increasing transparency through explainable AI (XAI) methodologies are some potential answers to these issues [23]. Moreover, more precise diagnostic systems could be produced by merging mammography with other modalities, including MRI, ultrasound, and genetic biomarkers. To turn energy deep learning from research into reliable clinical tools for breast cancer screening, these issues need to be resolved [23] [24].

2.2 Genomic Biomarkers in Breast Cancer Detection

Fundamentally, breast cancer is a hereditary disease characterized by complex genetic alterations. Large-scale characterization of these genetic alterations is now possible thanks to developments in next-generation sequencing (NGS) technologies, which produce high-dimensional datasets that record tumor heterogeneity at previously unheard-of resolution [25][26]. Genomic biomarkers were a rich but analytically difficult data channel for prognosis modeling, hazard stratification, and the computerized diagnosis of breast cancer [27][28]. Because gene expression profiling can identify clinically significant differences between tumor subtypes, it has become a popular molecular data type [28][29]. Nevertheless, transcriptome datasets often include thousands of features, small sample sizes, and tens, which can result in problems including feature redundancy, the curse of dimensionality, and model overfitting [26].

To extract discriminative genetic markers for breast cancer diagnosis and classification, these features have played a key role in advancing algorithms for feature selection, dimensionality reduction, and machine learning [30]. Compared with traditional clinical or histological criteria, molecular subtyping based on genomic biomarkers has demonstrated greater discriminative power. Luminal A, Luminal B, HER2-enriched, and Basal-like breast cancer subtypes are frequently identified by computational models using immunohistochemistry-derived markers (ER, PR, HER2) or gene-expression signatures (e.g., PAM50). Additionally, genomic indicators are essential for hazard assessment and predictive modeling of therapy response [31][32]. As multigene expression tests like Oncotype DX and MammaPrint, HER2 amplification, and BRCA1/2 mutation status have all been thoroughly examined as predictive and prognostic markers [31][33]. Robust learning algorithms that can handle a range of data types and varying noise levels are necessary for incorporating such biomarkers into computational frameworks. In order to increase prediction accuracy and personalized treatment regimens, recent research has increasingly resorted to machine learning and deep learning approaches that imitate these intricate genetic patterns. Subtype-aware modeling is crucial to precision oncology since these subtypes have

distinct genetic profiles, prognostic outcomes, and treatment responses [33]. From a data-analysis standpoint, this stratification problem has often been presented as a multi-class classification issue in high-dimensional, unjust circumstances [34]. The identification of genetic biomarkers has been vulnerable from an engineering standpoint due to a number of significant issues, including excessive data dimensionality, a lack of labeled samples, intrinsic biological variability, and interpretability limitations. Given their acknowledged therapeutic significance, genetic biomarkers thus constitute a crucial area for computational advancements in breast cancer detection systems. Table 1 shows research linking genetic markers to breast cancer treatment strategies.

Table 1. Key Clinical Trials Linking Genomic Biomarkers to Treatment Planning in Breast Cancer (2020–2025).

Reference	Targeted Biomarker	Disease Stage / Design	Primary Outcome	Key Finding
Robson et al., NEJM 2017 [31]	BRCA1/2 mutations	Advanced/Metastatic; RCT (talazoparib / olaparib vs chemo)	PFS, OS	PARP inhibitors improved PFS and QoL; OS impact was influenced by later therapy.
Modi et al., NEJM 2022 [32]	HER2-low	Metastatic; RCT (T-DXd vs chemo)	PFS, OS	When T-DXd was used instead of the doctor's preferred chemotherapy, PFS and OS were considerably improved.
Tutt et al., NEJM 2021 [33]	BRCA1/2 germline mutations	High-risk early; Adjuvant; RCT (olaparib vs placebo)	IDFs, DDFS, OS	Adjuvant olaparib improved IDFs and OS in BRCA1/2 carriers.
Hortobagyi et al., NEJM 2022 [35]	HR+ advanced (ER/PR+)	Metastatic; RCT (ribociclib + letrozole vs letrozole)	OS, PFS	Ribociclib + letrozole significantly improved OS and PFS in the first-line setting.
Martoň et al., NEJM 2024 [36]	HR+ (ER/PR+), HER2-negative	Stage II–III; Adjuvant; RCT (ribociclib + NSAI vs NSAI)	IDFs	Ribociclib + NSAI significantly improved iDFS compared to ET alone.

2.2.1 Methods for Analyzing Genomic Data

A- Machine Learning

Examining genetic data related to breast cancer is a high-dimensional learning problem with large feature sets, inherent biological heterogeneity, and small sample sizes. Millions of DNA or RNA sequences are produced in each sample by modern genomic technologies, such as next-generation sequencing (NGS) and microarrays; traditional statistical analysis is inadequate for detecting and forecasting healthy trends. In order to extract discriminative genetic characteristics and create prediction models for detection, prognosis, and therapy response, machine learning techniques become essential [37][38]. The documentation of mutations, copy-number variations, and gene-expression changes has been enabled by the basic genomic and transcriptomic characteristics provided by NGS platforms [30]. For instance, in the context of breast cancer, mutations in genes such as BRCA1/2 and TP53, as well as HER2 amplification, were vital inputs for machine learning models that perform hazard assessment and classification

[37]. Before models can be trained, feature selection and dimensionality reduction techniques are needed since the large number of sizes in NGS data currently causes issues such as overfitting, susceptibility to noise, and extra features [37]. Because of its affordability and the availability of standardized analytical pipelines, microarray-based gene expression profiling continues to be widely used. Strong predictive value has been shown by microarray-derived expression-based markers like the clinically validated Oncotype DX and MammaPrint. From a computational perspective, the comparison of expression profiles from malignant and normal tissues was completed using machine learning algorithms that were searched for subtype-specific patterns and predictive biomarkers [39]. Bioinformatics tools such as RNA-seq analysis events and genotyping tools similar to the Genome Analysis Toolkit (GATK) play a crucial preprocessing role, transforming raw sequencing data into intended feature matrices [28] [40]. However, downstream machine learning models that can manage large dimensionality, heterogeneous feature distributions, and missing data were still necessary for clinical implications. Machine learning frameworks provide a scalable foundation for precision oncology and tailored breast cancer care by integrating genomic, transcriptomic, and clinical features.

B- Deep Learning

By enabling autonomous symbol learning from high-dimensional genetic data, deep learning techniques surpass traditional machine learning. Deep neural networks were able to capture the nonlinear correlations seen in gene expression contours and mutation patterns, in contrast to manually developed feature-based methods. In breast cancer genomics, the extremely context-dependent and highly nonlinear connections between genes and tracks are advantageous [22]. One of the most often used deep learning architectures for genomic analysis is the Multilayer Perceptron (MLP), chiefly for gene expression-based classification and prognosis prediction [41]. MLPs are frequently paired with dimensionality reduction methods such as Principal Component Analysis (PCA) or Autoencoders to overcome the curse of dimensionality [3]. Extreme Learning Machines (ELMs) and MLPs both exhibit competitive performance in breast cancer genomic classification tasks, according to comparative research [3]. Because autoencoder-based models can learn dense, physically meaningful representations from high-dimensional genomic data, they have attracted significant interest. Specifically, by concurrently modeling gene expression patterns and therapeutic characteristics, Variational Autoencoders (VAEs) have been successfully used to expand biomarkers and predict therapy response [42]. By extracting latent representations that improve classification robustness and interpretability, these models successfully tackle the "large p, small n" issue. Deep learning techniques outperform conventional machine learning by enabling autonomous symbol learning from high-dimensional genetic data. Different manually constructed feature-based procedures, deep neural networks were able to capture the nonlinear relationships found in gene expression contours and mutation patterns. The highly nonlinear and context-dependent relationships between genes and tracks in breast cancer genomics greatly benefit from this capability [43]. All things considered, deep learning frameworks offer a solid computational

foundation for developing accurate algorithms for modeling complicated genetic designs and identifying breast cancer.

3. Related work

This section reviews and explains important prior research on detection procedures in order to emphasize the vital significance that early identification of breast cancer plays in enhancing patient survival. It emphasized the significance of key approaches that allow for early tumor detection, namely imaging technologies and genetic biomarkers. Additionally, many previous investigations have used efficient detection techniques like machine learning and deep learning because of their increasing relevance and dependable performance.

3.1 Using Mammography

WU et al. (2020) [44] molded high-resolution mammograms using BI-RADS data. The approach used machine learning, multi-view (CC & MLO), heatmap channels from a patch-level network, and pixel- and breast-level labels. The two-stage CNN is based on a customized ResNet architecture and consists of a patch-level network and a breast-level network. In a reader study, 14 radiologists each read 720 exams using this method. An independent screening population was used to test a hybrid model that combined AI and a radiologist. Thus, (89.5%) AUC for malignant detection and (75.6%) AUC for benign detection were obtained for accuracy. A small set of tests. The entire diagnostic process is not covered by limited clinical validation. Although the model design is straightforward, it could be improved. only detects cancer that is apparent at the time of screening.

P. E. Jebarani et al. (2021) [45] used MIAS mammogram data with the K-Means GMM method. The image size was 1024 pixels after the database was padded, clipped, and trimmed to a 200-micron pixel edge. The accuracy was 95.50%, but the method has limitations, including artifacts introduced during image preparation, which, as a result of additional lesion spots being marked, affect the image.

U. Naseem et al. 2022 [46] used the Wisconsin Breast Cancer (Diagnosis) collection, which contains 569 cases and 32 attributes (including an ID and a target variable), to conduct their study. The writers utilized BI-RADS data to enhance mammography screening and to assess the efficacy of different machine learning models for breast cancer diagnosis. Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Logistic Regression (LR) were among the classifiers that were compared. According to the experiment's findings, SVM had the highest accuracy (98.83%), followed by LR (98%), NB, and DT (93.56% and 91.22%, respectively). These responses show that SVM and LR function well on complex medical datasets and have good classification accuracy. The study also demonstrates how machine learning may assist radiologists by offering precise decision-support tools for early detection of breast cancer.

Aziz, S. et al. (2023) [47] authorized a study that used images of breast cancer histology collected from Jimma University of Medical Sciences. BC_IDC_Grade_1, BC_IDC_Grade_2, and BC_IDC_Grade_3, three indicators of offensive ductal carcinoma, were represented by 906 image

samples at four magnification levels: 4×, 10×, 20×, and 40×. The efficacy of categorization was tested using several state-of-the-art CNN architectures, including VGG16, ResNet50, InceptionNetV3, MobileNetV3, EfficientNetV3, and IVNet; some of these models have undergone ImageNet pre-training to aid transfer learning. Additionally, sophisticated picture preparation techniques were employed to improve feature extraction and increase the models' capacity for generalization. The study discovered that differences in magnification levels offer crucial information for determining the severity of the disease and that CNN-based models showed significant potential in discriminating among IDC grades. The scientists came to the conclusion that, when paired with advanced preprocessing, deep learning techniques can be effective diagnostic tools for breast cancer pathology.

MO et al. (2023) [48] proved the use of a pre-trained HoVer-Trans model that combines a CNN and a transformer to diagnose breast cancer ultrasound images. The purpose of the HoVer-Trans block is to create anatomical prior knowledge in BUS images. When trained on a smaller dataset, such as UDIAT, the suggested model has shown poor classification performance due to its complexity, achieving accuracies of 0.85%, 0.84%, and 0.81%.

W. Arshad et al.(2023) [49] studied 277,524 50 × 50 picture patches for breast histopathology photos in the Kaggle dataset. The dataset involved 78,786 Invasive Ductal Carcinomas (IDC+) patches and 198,738 non-cancerous patches from 162 breast cancer slide pictures. The authors used cutting-edge convolutional neural network (CNN) architectures, such as VGG-16, DenseNet-121, and MobileNetV2, to create and evaluate classification models. Model construction has been made easier and more efficient by utilizing the TensorFlow and Keras frameworks. With accuracy rates of 98%, 99%, and 97% in differentiating between benign and malignant tissue, all three topologies demonstrated outstanding performance. This study established that deep learning models trained on histopathology data can be extremely helpful to pathologists, enabling early and precise diagnosis of breast cancer.

VO et al. (2025) [50] recycled the Imaging Modality Model (IMM) to assess breast cancer using mammography data sets. The methodology has attained 97.8% accuracy and 96.3% sensitivity by preprocessing and refining raw mammogram images using a state-of-the-art frozen vision-language model and a 5-fold cross-validation strategy. Nevertheless, this method has drawbacks, such as its exclusive use to mammograms without genomic integration, and has been required on a single dataset without external validation.

S. A. Qureshi et al. (2025) [51] created an automated method that uses mammography images from the DDSM and MIAS datasets to identify breast cancer. The authors' use of preprocessing and data augmentation techniques to improve image quality and boost dataset diversity boosted the training process. The evaluation was conducted using a 70,30 train-test split and a recently developed deep CNN architecture. With 96.4% classification accuracy, the suggested model performed competitively with traditional methods. The study has several shortcomings, including relatively small datasets and the lack of integration of genetic or clinical

data. Also, no multi-modal method had been tested, which limits the generalizability of the consequences to broader clinical applications.

Table 2 presents the most important preceding studies on mammography-based breast cancer detection approaches, the algorithms used, their results, and the problems they encountered.

Table 2. Reviewed Literature Studies on Breast Cancer Detection Using Mammography.

Reference	Dataset	Method / Model	Accuracy	Limitations
Wu et al. (2020) [44]	BI-RADS	Two-stage CNN + ResNet	AUC 89.5% (Malignant)	Small test set, limited clinical validation
Jebarani et al. (2021) [45]	MIAS	GMM + K-means	95.5%	Artifacts in preprocessing
U. Naseem et al. (2022) [46]	BI-RADS	SVM, LR, NB, DT	98.8%	Small traditional dataset
S. Aziz et al. (2023) [47]	Jimma Univ. Medical Sciences	IVNet, CNNs (VGG16, ResNet50)	—	Focus on preprocessing; no external validation
Mo et al. (2023) [48]	UDIAT, BUSI, GDPH&SYSUCC3	HoVer-Trans, CNN	0.85%,0.84%,0.81%	Model complexity
Arshad et al. (2023) [49]	Kaggle (Histopathology)	MobileNetV2, VGG-16, DenseNet-121	97%,98%, and 99%	limited dataset size, only three architectures
Vo et al. (2025) [50]	EMBED, CBIS-DDSM	Pretrained Vision-Language Model	97.8%	No genomic data, no external validation
Qureshi et al. (2025) [51]	MIAS, DDSM	Deep CNN	96.4%	No genomic data

An overview of the mammography datasets commonly used in deep learning-based breast cancer diagnosis is publicized in Table 3, which also details the quantity of samples, clinical focus, and intrinsic constraints of publicly accessible databases [52] [53]. These datasets vary greatly in terms of image quality, annotation detail, and class balance, all of which have a direct impact on model performance and generalizability. For the majority of image-based research, they provide the fundamental framework. This comparison attracts attention to both the benefits of commonly used standards and the determined issues with clinical representativeness and data variety.

Table 3. Public Mammography Imaging Datasets.

Dataset Name	Number of Samples	Clinical Focus	Limitations
MIAS (Mammographic Image Analysis Society)	322 images	Early detection of benign and malignant breast lesions; widely used for algorithm benchmarking	Small dataset size; low spatial resolution; limited clinical diversity

DDSM (Digital Database for Screening Mammography)	~2,620 cases (~10,000 images)	Detection of masses and calcifications in screening mammography	Film-based images; requires digitization; annotation quality varies
CBIS-DDSM	~3,000 mammography studies	Standardized subset of DDSM for mass and calcification detection	Limited demographic information; still based on digitized film images
INbreast	410 images (115 patients)	High-resolution full-field digital mammography (FFDM) with precise annotations	Small dataset size; limited population diversity
BCDR (Breast Cancer Digital Repository)	~1,000 images (various versions)	Risk assessment and lesion characterization using digital mammography	Restricted access for some subsets; inconsistent annotation formats
OPTIMAM Mammography Image Database	>2.5 million images	Large-scale population screening and AI-based diagnostic development	Access restrictions; requires ethical approval; limited public annotations
VinDr-Mammo	5,000 images	BI-RADS classification and lesion detection in digital mammography	Limited longitudinal data; single-country population
CMMD (Chinese Mammography Database)	~5,000 images	Dense breast analysis and cancer detection in Asian populations	Limited external validation; demographic bias
BCS-DBT (Breast Cancer Screening – Digital Breast Tomosynthesis)	~16,000 DBT volumes	3D tomosynthesis-based breast cancer detection	High computational cost; limited availability of annotations

3.2 Using Genomic Biomarker

R. Lupat et al. (2023) [54] used TCGA and METABRIC data to clarify somatic mutation, copy number variation, and gene expression. They also employed a deep learning approach and approximately 47,000 characteristics from more than 15,000 genes (multi-omics). With a 70% training, 30% validation split, stratified K-Fold, and an independent test set, the semi-supervised Autoencoder + Multi-task Neural Network (Moanna) yields accuracy values. ER: 96.5%, Basal: 98.4%, PAM50: 85.6%, and this system had several challenges. cannot forecast a single sample, requires normalized reference samples, and integrates diverse data types early.

E. K. Jadoon et al. (2023) [55] designed an outline based on heterogeneous stacking to augment breast cancer prognosis. A carefully selected subset of 130 highly significant pathological features, drawn from an initial pool of 1990 and 2020 features from gene expression CNNs and DNNs, was used to train the design. The goal of this dimensionality reduction was to eliminate superfluous data and enhance the computing efficiency of the prediction model. The stacking collaborative technique used many classifiers and their balancing abilities to produce reliable results. In an experimental test, the proposed technique outperformed multiple baseline models with an overall accuracy of 97% and a sensitivity of 92%. These findings demonstrate how

stacking-based designs and feature selection may enhance breast cancer prognostic and diagnostic systems.

Rahman et al. (2024) [56] used the Wisconsin Breast Cancer (Diagnostic) database and an extreme gradient boosting method to identify breast cancer. They chose just 13 features that improve the gradient classifier, and used the Bonferroni correction. With 99.12% accuracy, 0.9767 predictive accuracy, 1.0 sensitivity, 0.9861 specificity, and 0.9882 F1 score, our model outdoes state-of-the-art approaches. Furthermore, our model outperformed the other models in terms of computational speed, according to the experimental results. The consequences of this study improve the diagnosis and prognosis of breast cancer. To advance performance, future research will focus on optimizing the model's limits.

R. Zeng et al. (2024) [57] used the general BreaKHis dataset to analyze breast cancer histopathology images. The dataset includes 7,909 photos from 82 patients, spanning both benign and malignant tumor collections. To guarantee consistency throughout the model training process, the corresponding image was preprocessed by resizing to a standard resolution of 224×224 pixels. The authors create FastLeakyResNet-CIR, a novel deep learning architecture that combines enhanced activation functions with residual learning to improve feature representation. With a classification accuracy of 98.94%, the experimental results outperformed several current benchmark techniques. This study demonstrates that sophisticated deep learning models can be highly effective for histopathology-based breast cancer detection and focuses on the role of specialized CNN architectures in improving diagnostic performance.

Tsiknakis et al. (2024) [30] combined TCGA-BRCA data for whole slide images (H&E-stained) and TP53 mutation status as a biomarker, using 20× and 10× patches and domain-specific and ImageNet features, aggregated via concatenation or linear transformation. They used nine modified state-of-the-art Multiple Instance Learning (MIL) models with a multi-resolution aggregation framework, trained on TCGA and Cohort 1, and tested on all cohorts (including TransNeo) for grade, TP53 mutation, and survival prediction. Given an accuracy rate of 89.5%, the method has limitations, including the need for additional validation and a preference for multi-scale models, which are easier to interpret and more reliable.

T. Khater et al. (2025) [58] used WBC and WDBC data. The permutation method states that the bare nuclei are the most significant feature. The more accurately the model predicts the malignant type, the higher the value of bare nuclei, which denotes the absence of a cell membrane. There can be an indirect relationship between the two characteristics, though. For instance, the WDBC dataset's "Bare Nuclei" feature, which identifies aberrant cell division and proliferation, may affect the "Worst Area" feature. Larger, irregularly shaped cell nuclei are another possible outcome. achieved a model 97.7% accuracy and 98.2% precision using KNN on the WBC dataset. With 98.6% accuracy and 94.4% precision, the ANN achieves the best performance on the WDBC dataset. More research and analysis are required to investigate any potential connections between

these characteristics and the biological processes they are thought to represent. Table 4 shows the literature on breast cancer detection using genomic biomarkers.

Table 4. Reviewed Literature Studies on Breast Cancer Detection Using Genomic Biomarkers.

Reference	Dataset	Method / Model	Accuracy	Limitations
Tsiknakis et al. (2024) [30]	TCGA-BRCA	Multi-resolution MIL models	89.5%	Lower performance on biopsies, more validation needed
Lupat et al. (2023) [54]	METABRIC, TCGA	Moanna (Semi-supervised Autoencoder + Multi-task NN)	ER: 96.5%, Basal: 98.4%, PAM50: 85.6%	Requires normalized samples; cannot predict a single case
Jadoon et al. (2023) [55]	METABRIC	CNN	97%	Limited selected features
Rahman et al. (2024) [56]	Wisconsin Breast Cancer (genes)	CNN + eXtreme Gradient Boosting	99.12%	Computationally expensive, needs balanced data
Zeng et al. (2024) [57]	BreaKHis (genes)	FastLeakyResNet-CIR	98.9%	Limited dataset
T. Khater et al. (2025) [58]	WBC and the WDBC (genes)	SVM, RF, and XG-boost	97.7%, 98.2%, 98.6%	Handcrafted features, no genomic integration

Table 5 delivers a comparative valuation of the main genomic and multi-omics datasets used in breast cancer diagnosis and forecast studies to supplement imaging-based approaches [59][60]. The table displays important information, including dataset size, molecular emphasis, and practical limitations similar data heterogeneity, normalization needs, and limited availability. These genetic insights facilitate molecular studies of tumor biology as well as progressive machine learning and deep learning models for hazard prediction and subtype classification. Though the difficulty of combining several omics layers and the heterogeneity of data-collection procedures remain major obstacles to the development of reliable, therapeutically effective models.

Table 5. Public Breast Cancer Genomic Datasets.

Dataset Name	Number of Samples	Clinical Focus	Limitations
TCGA-BRCA (The Cancer Genome Atlas – Breast Invasive Carcinoma)	~1,100 patients	Gene expression, mutation analysis, subtype classification, and prognosis prediction	High dimensionality, missing values, and limited normal tissue samples
METABRIC (Molecular Taxonomy of Breast Cancer	~2,000 patients	Molecular subtyping, survival analysis, and genomic-clinical association	Limited RNA-seq data; batch effects between cohorts

International Consortium)			
GEO Breast Cancer Datasets (e.g., GSE96058, GSE42568)	Varies (100–3,000 samples)	Gene expression profiling and biomarker discovery	Heterogeneous platforms; inconsistent preprocessing
SCRNA-seq Breast Cancer Datasets	~10,000–100,000 cells (study-dependent)	Tumor heterogeneity, cell-type identification, and microenvironment analysis	High noise; limited sample size per patient; high computational cost
ICGC Breast Cancer Projects	~1,000 patients	Somatic mutation analysis and cancer genome characterization	Limited transcriptomic depth; access restrictions
CPTAC Breast Cancer Dataset	~1,000 samples	Proteogenomic analysis integrating genomics and proteomics	Complex data integration; limited public tools
ArrayExpress Breast Cancer Studies	Varies	Gene expression and epigenetic analysis	Dataset fragmentation; variable annotation quality

The comprehensive review introduced in this study demonstrates significant improvements in sensitivity, specificity, and diagnostic accuracy achieved utilizing AI-driven methodologies despite ongoing technological challenges and unresolved issues. Results from image-based detection studies demonstrate that deep learning models consistently outperform traditional machine learning methods and earlier computer-aided detection organisms when evaluating mammography and histopathology images. For example, WU et al. (2020) [44] achieved an AUC of 89.5% for malignancy diagnosis using a two-stage CNN on multi-view mammograms, demonstrating the potential of incorporating AI into radiologist clarification. The same research by VO et al. (2025) [50] and R. Zeng et al. (2024) [57] uses vision-language models and exact CNN architectures to achieve impressive accuracy (up to 98.94%) on histopathological images. These outcomes imply that advanced deep learning models, by way of CNNs, transformers, and ensemble models, can identify subtle imaging features vital for early diagnosis, such as tissue heterogeneity and microcalcifications. Despite these developments, there are still certain restrictions: Due to dataset limitations, many researchers use very small or single-origin datasets (such as DDSM, MIAS, and WBC), which restrict their application to larger groups. Most image-based models do not incorporate imaging with genetic or clinical data, which could increase diagnosis accuracy, due to a lack of multimodal integration. Additionally, complex projects that require a lot of processing power, such as multi-resolution MIL or HoVer-Trans, may increase computational complexity and compromise clinical acceptability. The results essentially demonstrate that deep learning models increase diagnosis accuracy; nonetheless, their use in clinical settings needs validation across larger, multi-institutional datasets and organizations with longer-term data. Deep learning and machine learning techniques are becoming more widely used in genomic biomarker research to evaluate breast cancer risk, classify subtypes, and forecast patient outcomes. Notable studies include R. Lupat et al.'s (2023) [54] application of Moanna Autoencoders to multi-omics data (from METABRIC and TCGA), which produced up to 98.4%

accuracy for basal-like subtypes. Similarly, E. K. Jadoon et al. (2023) [55] attained a 97% accuracy in prognostic prediction by combining feature selection with stacking ensembles. These methods, including MLPs, Variational Autoencoders, and ensemble deep learning, are instrumental in processing vast, diverse genomic datasets to identify clinically relevant biomarkers. The thousands of features and various scales present in multi-omics datasets due to data dimensionality and heterogeneity complicate model training and elucidation, notwithstanding some challenges with genomic-based techniques. Since good prediction frequently necessitates normalized orientation samples, which limit their application to specific patient instances, normalization and calibration are required. Inadequate integration with imaging: Most genomic models do not incorporate combined imaging-genomic data, which eliminates the potential for multimodal early detection.

4. Conclusion

Breast cancer is a serious disease that needs to be detected early to ensure an effective treatment plan. Deep and machine learning techniques have made recent advances in artificial intelligence and hold great promise for improving breast cancer diagnosis and prognosis. One of these methods is the use of medical imaging techniques, such as mammography, and genomic biomarkers; this review sought to examine the many approaches in this field and classify them into two categories. This review provides an overview of the latest detection and classification techniques. Among the reviewed studies, the mammogram-based detection method by Jebarani et al. (2021) [45] achieved the highest overall accuracy. Using a GMM + K-means clustering technique and the MIAS mammography dataset, it identified breast cancer with 95.5% accuracy. However, methods based on genomic biomarkers demonstrated exceptional molecular-level diagnostic efficacy. Lupat et al. (2023) [54] employed the Moanna semi-supervised autoencoder. reported 96.5% to 98.4% accuracy, while Jadoon et al. (2023) [55] achieved 97% accuracy using deep neural networks trained on the METABRIC dataset. These models require well-normalized genomic datasets and are constrained by data complexity, despite their effectiveness in detecting breast cancer subtypes. These models have a lot of potential in terms of accuracy and computational efficiency, but their complexity and high processing cost have severely limited their practical applicability. These research findings demonstrate how deep learning and machine learning methods can be applied to the diagnosis and prognosis of breast cancer. When compared to traditional methods, these methods have demonstrated encouraging results in clinical decision-making, early detection, and diagnostic accuracy. Despite recent advancements in the use of genetic and imaging approaches to diagnose breast cancer, some concerns still need to be resolved. Another issue that needs to be addressed is the presence of natural data and incorrect diagnoses that cause patients to disregard their symptoms. To improve biological interpretability and diagnostic accuracy, future research should focus on using multimodal data sources, such as genetic biomarkers and mammography images. Furthermore, to ensure that current deep learning methods are generalizable, clinical validation on sizable, varied datasets is necessary. Last but not least, developing understandable AI frameworks may aid in bridging the trust gap between medical professionals and high-performance models, paving the way for more precise and open breast cancer diagnosis.

Declaration of generative AI and AI-assisted technologies in the writing process

-None

Disclosures

The authors have no conflicts of interest to declare in relation to this manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Term
3D	Three-Dimensional
AI	Artificial Intelligence
AUC	Area Under the Curve
BC	Breast Cancer
BCDR	Breast Cancer Digital Repository
BRCA1/2	Breast Cancer Gene 1/2
BUS	Breast Ultrasound
CAD	Computer-Aided Diagnosis
CC	Cranio-Caudal View
CMMD	Chinese Mammography Database
CNN	Convolutional Neural Network
CPTAC	Clinical Proteomic Tumor Analysis Consortium
DBT	Digital Breast Tomosynthesis
DDFS	Distant Disease-Free Survival
DDSM	Digital Database for Screening Mammography
DICOM	Digital Imaging and Communications in Medicine
DNA	Deoxyribonucleic Acid
DT	Decision Tree
DX	Diagnosis
ELMs	Extreme Learning Machines
EMBED	Electronic Medical Breast Data
ER	Estrogen Receptor
FFDM	Full-Field Digital Mammography
GATK	Genome Analysis Toolkit
GEO	Gene Expression Omnibus
GMM+K-	Gaussian Mixture Model + K-means
GNN	Graph Neural Network
HER2	Human Epidermal Growth Factor Receptor 2
HR+	Hormone Receptor Positive
ICGC	International Cancer Genome Consortium
ID	Identification
IDC+	Invasive Ductal Carcinoma Positive
IDFs	Invasive Disease-Free Survival
IVNet	Inception Variant Network
KNN	K-Nearest Neighbors
LR	Logistic Regression
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MIAS	Mammographic Image Analysis Society
MIL	Multiple Instance Learning
MLO	Medio-Lateral Oblique View
MLP	Multi-Layer Perceptron

MRI	Magnetic Resonance Imaging
NB	Naive Bayes
NGS	Next-Generation Sequencing
NSAI	Non-Steroidal Aromatase Inhibitor
OS	Overall Survival
PAM50	Prediction Analysis of Microarray 50
PCA	Principal Component Analysis
PFS	Progression-Free Survival
PR	Progesterone Receptor
RCT	Randomized Controlled Trial
RNA	Ribonucleic Acid
scRNA	Single-cell RNA sequencing
TCGA	The Cancer Genome Atlas
TCGA-BRCA	TCGA Breast Cancer Dataset
T-DXd	Trastuzumab Deruxtecan
UDIAT	Breast Ultrasound Dataset (UDIAT)
VAEs	Variational Autoencoders
WBC	Wisconsin Breast Cancer
WDBC	Wisconsin Diagnostic Breast Cancer
WHO	World Health Organization
XAI	Explainable Artificial Intelligence

References:

[1] Y. Zhang *et al.*, "Global burden of female breast cancer: new estimates in 2022, temporal trend and future projections up to 2050 based on the latest release from GLOBOCAN," *J. Natl. Cancer Cent.*, vol. 5, no. 3, pp. 287–296, 2025, doi: 10.1016/j.jncc.2025.02.002.

[2] A. Amin, D. A. U, P. Koteshwara, S. P C, and S. Mathew, "A systematic literature review on mammography: deep learning techniques for breast cancer detection with global and Asian perspectives," *BMC Cancer*, vol. 25, no. 1, 2025, doi: 10.1186/s12885-025-14876-5.

[3] J. R. Pérez-Núñez, C. Rodríguez, L. J. Vásquez-Serpa, and C. Navarro, "The Challenge of Deep Learning for the Prevention and Automatic Diagnosis of Breast Cancer: A Systematic Review," *Diagnostics*, vol. 14, no. 24, 2024, doi: 10.3390/diagnostics14242896.

[4] B. Abdikenov, T. Zhaksylyk, A. Imasheva, and D. Rakishev, "From Mammogram Analysis to Clinical Integration with Deep Learning in Breast Cancer Diagnosis," *Informatics*, vol. 12, no. 4, p. 106, 2025, doi: 10.3390/informatics12040106.

[5] N. C. da Rocha *et al.*, "Enhancing Breast Density Assessment in Mammograms Through Artificial Intelligence," *J. Imaging Informatics Med.*, no. C, 2025, doi: 10.1007/s10278-025-01657-6.

[6] M. Ma *et al.*, "Contrast-enhanced mammography-based interpretable machine learning model for the prediction of the molecular subtype breast cancers," *BMC Med. Imaging*, vol. 25, no. 1, 2025, doi: 10.1186/s12880-025-01765-3.

[7] S. Hussain, Y. Lafarga-Osuna, M. Ali, U. Naseem, M. Ahmed, and J. G. Tamez-Peña, "Deep learning, radiomics and radiogenomics applications in the digital breast tomosynthesis: a systematic review," *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–21, 2023, doi: 10.1186/s12859-023-05515-6.

[8] S. W. Duffy *et al.*, "Beneficial effect of consecutive screening mammography examinations on mortality from breast cancer: A prospective study," *Radiology*, vol. 299, no. 3, pp. 541–547, 2021, doi: 10.1148/radiol.2021203935.

[9] K. Feigin, "Quality assurance in Mammography: An overview," *Eur. J. Radiol.*, vol. 165, no. June, p. 110935, 2023, doi: 10.1016/j.ejrad.2023.110935.

[10] D. Simancas-racines, N. M. Román-galeano, J. P. Vásquez, D. J. Gavilanes, R. Vijayan, and C. Reytor-gonzález, "Liquid Biopsy and Multi-Omic Biomarkers in Breast Cancer : Innovations in Early Detection , Therapy Guidance , and Disease Monitoring," pp. 1–32, 2025.

[11] J. Ahmad, S. Akram, A. Jaffar, M. Rashid, and S. M. Bhatti, "Breast Cancer Detection Using Deep Learning: An Investigation Using the DDSM Dataset and a Customized AlexNet and Support Vector Machine," *IEEE Access*, vol. 11, no. October, pp. 108386–108397, 2023, doi: 10.1109/ACCESS.2023.3311892.

[12] P. Oza *et al.*, "Digital mammography dataset for breast cancer diagnosis research (DMID) with breast mass segmentation analysis," *Biomed. Eng. Lett.*, vol. 14, no. 2, pp. 317–330, 2024, doi: 10.1007/s13534-023-00339-y.

[13] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," pp. 1–23, 2025, doi: 10.7717/peerj.6201.

[14] A. Baccouche, B. G. Zapirain, and A. S. Elmaghraby, "An integrated framework for breast mass classification and diagnosis using stacked ensemble of residual neural networks," *Sci. Rep.*, no. 0123456789, pp. 1–17, 2022, doi: 10.1038/s41598-022-15632-6.

[15] C. Ben Rabah, A. Sattar, A. Ibrahim, and A. Serag, "A Multimodal Deep Learning Model for the Classification of Breast Cancer Subtypes," *Diagnostics*, vol. 15, no. 8, pp. 1–17, 2025, doi: 10.3390/diagnostics15080995.

[16] S. Pramanik, D. Bhattacharjee, and M. Nasipuri, "Wavelet based thermogram analysis for breast cancer detection," *2015 Int. Symp. Adv. Comput. Commun. ISACC 2015*, pp. 205–212, 2016, doi: 10.1109/ISACC.2015.7377343.

[17] R. Gudur, N. Patil, and S. T. Thorat, "Early Detection of Breast Cancer using Deep Learning in Mammograms," *J. Pioneer. Med. Sci.*, vol. 13, no. 2, pp. 18–27, 2024, doi: 10.61091/jpms202413204.

[18] Y. Khourdifi, A. El Alami, M. Zaydi, and Y. Maleh, "Early Breast Cancer Detection Based on Deep Learning : An Ensemble Approach Applied to Mammograms," pp. 2338–2373, 2024.

[19] G. Hamed, M. Marey, S. E. Amin, and M. F. Tolba, "Automated Breast Cancer Detection and Classification in Full Field Digital Mammograms Using Two Full and Cropped Detection Paths Approach," *IEEE Access*, vol. 9, pp. 116898–116913, 2021, doi: 10.1109/ACCESS.2021.3105924.

[20] S. Loibl *et al.*, "Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up ☆," *Ann. Oncol.*, vol. 35, no. 2, pp. 159–182, 2024, doi: 10.1016/j.annonc.2023.11.016.

[21] M. Wei *et al.*, "A Benign and Malignant Breast Tumor Classification Method via Efficiently Combining Texture and Morphological Features on Ultrasound Images," *Comput. Math. Methods Med.*, vol. 2020, 2020, doi: 10.1155/2020/5894010.

[22] E. Strelcenia and S. Prakoonwit, "Improving Cancer Detection Classification Performance Using GANs in Breast Cancer Data," *IEEE Access*, vol. 11, no. July, pp. 71594–71615, 2023, doi: 10.1109/ACCESS.2023.3291336.

[23] K. J. Geras, R. M. Mann, and L. Moy, "Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives," *Radiology*, vol. 293, no. 2, pp. 246–259, 2019, doi: 10.1148/radiol.2019182627.

[24] A. Yala, T. Schuster, R. Miles, R. Barzilay, and C. Lehman, "A deep learning model to triage screening mammograms: A simulation study," *Radiology*, vol. 293, no. 1, pp. 38–46, 2019, doi: 10.1148/radiol.2019182908.

[25] I. Bedrosian *et al.*, "Germline Testing in Patients with Breast Cancer: ASCO-Society of Surgical Oncology Guideline," *J. Clin. Oncol.*, vol. 42, no. 5, pp. 584–604, 2024, doi: 10.1200/JCO.23.02225.

[26] J. J. Jeong *et al.*, "The EMory BrEast imaging Dataset (EMBED): A Racially Diverse, Granular Dataset of 3.4 Million Screening and Diagnostic Mammographic Images," *Radiol. Artif. Intell.*, vol. 5, no. 1, pp. 1–9, 2023, doi: 10.1148/ryai.220047.

[27] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018, doi: 10.3322/caac.21492.

[28] S. Piscuoglio *et al.*, "Genomic and transcriptomic heterogeneity in metaplastic carcinomas of the breast," *npj Breast Cancer*, no. October, 2017, doi: 10.1038/s41523-017-0048-0.

[29] A. N. Babatunde *et al.*, "Breast Cancer Classification Using Densenet121 And K-Means

Segmentation With Augmented Data," vol. 49, pp. 79–102, 2025.

[30] N. Tsiknakis *et al.*, "Unveiling the Power of Model-Agnostic Multiscale Analysis for Enhancing Artificial Intelligence Models in Breast Cancer Histopathology Images," *IEEE J. Biomed. Heal. Informatics*, vol. 28, no. 9, pp. 5312–5322, 2024, doi: 10.1109/JBHI.2024.3413533.

[31] M. E. Robson *et al.*, "OlympiAD final overall survival and tolerability results: Olaparib versus chemotherapy treatment of physician's choice in patients with a germline BRCA mutation and HER2-negative metastatic breast cancer," *Ann. Oncol.*, vol. 30, no. 4, pp. 558–566, 2019, doi: 10.1093/annonc/mdz012.

[32] S. Modi, "Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer: a plain language summary of the DESTINY-Breast04 study," *Futur. Oncol.*, vol. 21, no. 4, pp. 367–380, 2025, doi: 10.1080/14796694.2024.2434388.

[33] A. N. J. Tutt *et al.*, " Adjuvant Olaparib for Patients with BRCA1 - or BRCA2 -Mutated Breast Cancer ," *N. Engl. J. Med.*, vol. 384, no. 25, pp. 2394–2405, 2021, doi: 10.1056/nejmoa2105215.

[34] V. Shukla, Kaarthika, A. Mathur, P. Narayan, and K. Kishor, "A Multi-Modal Approach for the Molecular Subtype Classification of Breast Cancer by Using Vision Transformer and Novel SVM Polyvariant Kernel," *IEEE Access*, vol. 13, no. June, pp. 97545–97558, 2025, doi: 10.1109/ACCESS.2025.3575126.

[35] G. N. Hortobagyi *et al.*, "Overall Survival with Ribociclib plus Letrozole in Advanced Breast Cancer," *N. Engl. J. Med.*, vol. 386, no. 10, pp. 942–950, 2022, doi: 10.1056/nejmoa2114663.

[36] D. Slamon *et al.*, "Ribociclib plus Endocrine Therapy in Early Breast Cancer," *N. Engl. J. Med.*, vol. 390, no. 12, pp. 1080–1091, 2024, doi: 10.1056/nejmoa2305488.

[37] A. Arimura *et al.*, "TP53 and/or BRCA1 Mutations Based on CtDNA Analysis as Prognostic Biomarkers for Primary Triple-Negative Breast Cancer," *Cancers (Basel).*, vol. 16, no. 6, pp. 1–14, 2024, doi: 10.3390/cancers16061184.

[38] M. W. Qoronfleh and N. Al-Dewik, "Cancer Biomarkers: Reflection on Recent Progress, Emerging Innovations, and the Clinical Horizon," *Cancers (Basel).*, vol. 17, no. 18, pp. 1–9, 2025, doi: 10.3390/cancers17182981.

[39] H. S. Das, A. Das, A. Neog, S. Mallik, K. Bora, and Z. Zhao, "Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach," *Front. Genet.*, vol. 13, no. January, pp. 1–14, 2023, doi: 10.3389/fgene.2022.1097207.

[40] S. Schmidt, "Measuring absorptive capacity," *Proc. Int. Conf. Intellect. Capital, Knowl.*

Manag. Organ. Learn., pp. 254–260, 2009, doi: 10.1101/gr.107524.110.20.

[41] M. A. S. Al Husaini, M. H. Habaebi, S. A. Hameed, M. R. Islam, and T. S. Gunawan, "A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks," *IEEE Access*, vol. 8, pp. 208922–208937, 2020, doi: 10.1109/ACCESS.2020.3038817.

[42] Q. Wei and S. A. Ramsey, "Predicting chemotherapy response using a variational autoencoder approach," *BMC Bioinformatics*, vol. 22, no. 1, pp. 1–22, 2021, doi: 10.1186/s12859-021-04339-6.

[43] F. Z. Nakach, A. Idri, and E. Goceri, "A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification," *Artif. Intell. Rev.*, vol. 57, no. 12, 2024, doi: 10.1007/s10462-024-10984-z.

[44] N. Wu *et al.*, "Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening," *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1184–1194, 2020, doi: 10.1109/TMI.2019.2945514.

[45] P. E. Jebarani, N. Umadevi, H. Dang, and M. Pomplun, "A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection," *IEEE Access*, vol. 9, pp. 146153–146162, 2021, doi: 10.1109/ACCESS.2021.3123425.

[46] U. Naseem *et al.*, "An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers," *IEEE Access*, vol. 10, no. April, pp. 78242–78252, 2022, doi: 10.1109/ACCESS.2022.3174599.

[47] S. Aziz, K. Munir, A. Raza, M. S. Almutairi, and S. Nawaz, "IVNet: Transfer Learning Based Diagnosis of Breast Cancer Grading Using Histopathological Images of Infected Cells," *IEEE Access*, vol. 11, no. November, pp. 127880–127894, 2023, doi: 10.1109/ACCESS.2023.3332541.

[48] Y. Mo *et al.*, "HoVer-Trans: Anatomy-Aware HoVer-Transformer for ROI-Free Breast Cancer Diagnosis in Ultrasound Images," *IEEE Trans. Med. Imaging*, vol. 42, no. 6, pp. 1696–1706, 2023, doi: 10.1109/TMI.2023.3236011.

[49] W. Arshad *et al.*, "Cancer Unveiled: A Deep Dive Into Breast Tumor Detection Using Cutting-Edge Deep Learning Models," *IEEE Access*, vol. 11, no. November, pp. 133804–133824, 2023, doi: 10.1109/ACCESS.2023.3335604.

[50] H. Q. Vo *et al.*, "Frozen Large-Scale Pretrained Vision-Language Models are the Effective Foundational Backbone for Multimodal Breast Cancer Prediction," *IEEE J. Biomed. Heal. Informatics*, vol. 29, no. 5, pp. 3234–3246, 2025, doi: 10.1109/JBHI.2024.3507638.

[51] S. A. Qureshi *et al.*, "Breast Cancer Detection Using Mammography: Image Processing to Deep Learning," *IEEE Access*, vol. 13, no. April, pp. 60776–60801, 2025, doi:

10.1109/ACCESS.2024.3523745.

[52] J. Logan, P. J. Kennedy, and D. Catchpoole, “A review of the machine learning datasets in mammography , their adherence to the FAIR principles and the outlook for the future,” pp. 1–11, 2023, doi: 10.1038/s41597-023-02430-6.

[53] J. H. Jaradat, R. Amro, R. Hamamreh, A. Musleh, and M. S. Abdelgalil, “From Data to Diagnosis : Narrative Review of Open-Access Mammography Databases for Breast Cancer Detection,” pp. 1–10, 2024, doi: 10.59707/hymrPFNZ8344.

[54] R. Lupat, R. Perera, S. Loi, and J. Li, “Moanna: Multi-Omics Autoencoder-Based Neural Network Algorithm for Predicting Breast Cancer Subtypes,” *IEEE Access*, vol. 11, no. December 2022, pp. 10912–10924, 2023, doi: 10.1109/ACCESS.2023.3240515.

[55] E. K. Jadoon, F. G. Khan, S. Shah, A. Khan, and M. Elaffendi, “Deep Learning-Based Multi-Modal Ensemble Classification Approach for Human Breast Cancer Prognosis,” *IEEE Access*, vol. 11, no. July, pp. 85760–85769, 2023, doi: 10.1109/ACCESS.2023.3304242.

[56] M. A. Rahman *et al.*, “Enhancing Early Breast Cancer Detection Through Advanced Data Analysis,” *IEEE Access*, vol. 12, no. September, pp. 161941–161953, 2024, doi: 10.1109/ACCESS.2024.3483095.

[57] R. Zeng *et al.*, “FastLeakyResNet-CIR: A Novel Deep Learning Framework for Breast Cancer Detection and Classification,” *IEEE Access*, vol. 12, pp. 70825–70832, 2024, doi: 10.1109/ACCESS.2024.3401729.

[58] T. Khater *et al.*, “An Explainable Artificial Intelligence Model for the Classification of Breast Cancer,” *IEEE Access*, vol. 13, no. August 2023, pp. 5618–5633, 2025, doi: 10.1109/ACCESS.2023.3308446.

[59] R. Altaf, H. Nadeem, M. M. Babar, U. Ilyas, and S. A. Muhammad, “Genome - scale meta - analysis of breast cancer datasets identifies promising targets for drug development,” *J. Biol. Res.*, pp. 1–15, 2021, doi: 10.1186/s40709-021-00136-7.

[60] H. Zhang, H. Lin, and S. Dong, “Novel radiation-derived gene blueprint stratifying patients with breast cancer,” pp. 1–17, 2025.

رؤى مدفوعة بالذكاء الاصطناعي في سرطان الثدي: مراجعة معمقة للتعلم الآلي تدمج التصوير الطبي والبيانات الجينية

الخلاصة: *أصبح الكشف المبكر عن سرطان الثدي أحد أهم مجالات البحث الطبي نظرًا لأثره الكبير في تحسين معدلات التشخيص والحد من المضاعفات. تعتمد الدراسات الحديثة على تقنيات التعلم العميق لتحليل البيانات الطبية بطرق متقدمة. ففي التحليل الشعاعي، تستعمل الشبكات العصبية العميقة للتعرف على التغيرات النسيجية والكتل والتكلسات في صور الماموغرام، مما يتيح تصنيف الحالات بدقة عالية. بينما في مجال العلامات الجينومية، تُستخلص المعلومات من البيانات الجزيئية مثل التعبير الجيني والطفرات لتقييم مخاطر الإصابة وتحديد خصائص الورم. تتيح كل تقنية تحسين دقة التشخيص المبكر عند استخدامها بشكل منفصل، مع ضرورة توافر بيانات دقيقة ومنظمة لدعم النتائج وجعلها قابلة للتطبيق سريريًا. وعلى الرغم من*

التقدم الملحوظ في كلا المجالين، إلا أن كل منهما يواجه تحديات خاصة، مثل محدودية حجم البيانات، وعدم تجانسها، وارتفاع التكلفة الحاسوبية للنماذج المتقدمة. كما أن النماذج المعتمدة على البيانات الجينية تعاني من تعقيد الأبعاد ووجود قيم مفقودة، مما يتطلب تقنيات معالجة مسبقة دقيقة. لذلك، تتجه الأبحاث الحديثة نحو دمج البيانات متعددة الأنماط، مثل الجمع بين صور الماموغرام والبيانات الجينية، بهدف تحسين دقة التنبؤ وزيادة موثوقية النتائج. إضافةً إلى ذلك، يُعد التحقق السريري باستخدام بيانات واسعة ومتنوعة أمرًا ضروريًا لضمان قابلية تعميم هذه النماذج في البيئات الطبية الحقيقية. كما أن تطوير نماذج ذكاء اصطناعي قابلة للتفسير يسهم في تعزيز ثقة الأطباء بهذه الأنظمة ودعم استخدامها في اتخاذ القرارات الطبية. ومن المتوقع أن يسهم التكامل بين تقنيات التعلم العميق والبيانات متعددة المصادر في تحقيق تقدم كبير نحو تشخيص أكثر دقة وفعالية لسرطان الثدي في المست .

الكلمات المفتاحية: سرطان الثدي، سي إن إن، التعلم العميق، علم الجينوم، صور الثدي الشعاعية، التنبؤ بالتشخيص.