Engineering Sciences
ISSN: 1308 7231
Article ID: 1A0501

Status : Review
Received: 05.07.2025
Accepted: 20.10.2025

Maral A. Mustafa
Gazi University, maralanwer@ntu.edu.iq, Ankara-Türkiye
O. Ayhan Erdem
Gazi University, ayerdem@gazi.edu.tr, Ankara-Türkiye
Esra Söğüt
Gazi University, esrasogut@gazi.edu.tr, Ankara-Türkiye

DOI	http://dx.doi.org/10.12739/NWSA.2025.20.4.1A0501		
ORCID ID	0000-0002-0601-3457	0000-0001-7761-1078	0000-0002-0051-2271
Corresponding Author	Esra Söğüt		

ARTIFICIAL INTELLIGENCE IN BREAST CANCER DIAGNOSIS: CURRENT APPLICATIONS, CHALLENGES, AND THE ROLE OF EXPLAINABLE AI

ABSTRACT

Breast cancer is the most commonly diagnosed cancer in women worldwide, and its prevalence is constantly increasing despite significant advancements in early diagnosis and personalised treatment. Nevertheless, current workflows for diagnostic interventions face challenges such as overdiagnosis in low-risk populations, increasing workloads for radiologists and pathologists, and inconsistent interpretation of imaging and pathological study findings. Artificial intelligence (AI) has proven to be an effective solution to these issues by enhancing image analysis, automating labour-intensive processes, and facilitating clinical decision-making. This paper presents a narrative review of recent AI implementations in breast cancer screening and diagnosis, including malignancy detection and classification, tumour segmentation, molecular subtype prediction, and recurrence or metastatic risk assessment. Data sources in both imaging and non-imaging domains are analysed, including mammography, ultrasound, magnetic resonance imaging (MRI), histopathology, clinical variables, and multimodal data integration. The reviewed articles also identify explainable artificial intelligence (XAI) methods, such as SHAP, Grad-CAM, and LIME, as being key to improving transparency, interpretability, and clinician confidence in AI-assisted systems. Overall, the current evidence suggests that AI-based tools have the potential to enhance diagnostic accuracy, reduce inter-observer variability, and facilitate personalised risk evaluation and treatment planning. However, there are still multiple obstacles to the widespread clinical implementation of AI in breast cancer care, such as the heterogeneity of datasets, a lack of external and prospective validation, interpretability issues, and constraints based on real-world application. Therefore, future studies must focus on creating more and better-quality data, establishing standard assessment guidelines and solid explainability models, and conducting future clinical trials, in order to enable the safe, productive and fair integration of AI into routine breast cancer care.

Keywords: Screening, Explainable AI, Malignancy Classification, Recurrence Prediction, Image Segmentation.

1. INTRODUCTION

Breast cancer is gradually becoming more prevalent, and is currently the most frequently diagnosed malignancy in women worldwide, surpassing lung cancer in terms of incidence [1 and 2]. While the burden is increasing, screening and the development of personalised treatment options have led to

How to Cite:
Maral, A.M., Erdem, O.A., and Söğüt, E., (2025). Artificial intelligence in breast cancer diagnosis: current applications, challenges, and the role of explainable Al. Engineering Sciences, 20(4):129-147, DOI: 10.12739/NWSA.2025.20.4.1A0501.

a dramatic decrease in breast cancer-related mortality and improved patient outcomes [1 and 3]. Nevertheless, significant issues remain in the diagnostic pathway, including overdiagnosis in low-risk groups, increased strain on radiology and pathology services, and inconsistent image and specimen interpretation [4, 5, 6 and 7]. Furthermore, a lack of access to diagnostic tests and the high cost of advanced tests still impede the provision of timely and equitable care in many areas [8, 9, 10, 11 and 12]. The breast cancer diagnostic process is multi-step and usually involves screening, imaging assessment, tissue biopsy, pathology, staging, molecular and biomarker profiling, treatment planning and, in some cases, neoadjuvant therapy, surgery and systemic adjuvant treatments [17 and 18]. However, existing screening procedures fail to adequately consider personal risk diversity, potentially exposing lower-risk groups to unnecessary recall and therapy. In the United Kingdom, an independent analysis of randomised trials estimated a 19% risk of overdiagnosis with screening [4]. Despite risk stratification tools being used to inform intensified surveillance, including annual MRI scans for women with a lifetime breast cancer risk of ≥20%, most models are based on variables that are not routinely measured, have low predictive power (often AUC <0.7) and tend to select cancers with a better prognosis, thus restricting their influence at a population level [20, 21 and 22].

At the same time, shortages in the workforce and growing imaging volumes further burden clinical services. It is estimated that the United Kingdom will be short of radiologists by 40% in 2027, and the workload of pathologists in the United States has increased by 41.73% in the last ten years. These stresses are compounded by the time taken for sophisticated imaging, including DBT and MRI, the additional workload of pathological procedures (e.g., preparing extra slides), and the lack of consistency in agreement (between 75 and 88 per cent), even in focused diagnostic settings. The proficiency of radiologists in making diagnoses is strongly influenced by their experience and training, which are known to contribute to interpretive accuracy and error susceptibility [24]. However, substantial inter- and intra-reader variability in mammography interpretation persists, reflecting differences in training, experience, and interpretive approaches among radiologists [25]. Additionally, expensive infrastructure-based tests such as contrast-enhanced mammography, MRI, and gene tests such as Oncotype Dx, are not accessible to all institutions or patients. They may require referral or specimen transfer and some tissue-destroying tests can limit follow-up biomarker or genetic measures [12]. Artificial intelligence (AI) has become one of the most promising methods to assist with the diagnosis of breast cancer, enhance image interpretation, automate time-consuming processes, and perform predictive analytics in the fields of radiology and pathology [13, 14, 15 and 16]. In line with this, this review summarises existing AI applications for all major diagnostic tasks and data types in breast cancer care, and describes the evidence base supporting their potential clinical utility.

2. RESEARCH SIGNIFICANCE

This review focuses on the potential of artificial intelligence (AI) to address the major limitations of current breast cancer diagnosis and screening workflows. AI has the potential to reduce overdiagnosis, improve the accuracy and consistency of image interpretation, lighten the workload of clinicians, and support personalised treatment planning.

Particular emphasis is placed on radiology and pathology, where AI can assist with the detection of early-stage cancers, the characterisation of lesions, and the prediction of outcomes. By integrating both image- and non-image-based data, AI systems can support risk stratification, subtype classification, and recurrence prediction, thereby helping clinicians to make more informed decisions and optimise patient management.

This study is also significant because it highlights explainable AI (XAI) approaches, which are essential for establishing trust, transparency and accountability in AI-assisted medical decision-making. In high-stakes fields such as oncology, interpretability is a key factor in clinical adoption.

Highlights:

- It provides a comprehensive and up-to-date synthesis of artificial intelligence applications in breast cancer diagnosis. It systematically outlines how AI addresses major clinical challenges, such as overdiagnosis, interpretive variability and an increasing diagnostic workload.
- It offers a focused and structured evaluation of XAI methods, including SHAP, Grad-CAM, and LIME, highlighting their role in enhancing transparency, reliability, and clinical trust — essential factors for the safe integration of AI systems in oncology.
- It critically analyses both imaging-based and non-imaging multimodal datasets, identifying current capabilities and limitations and suggesting future research directions for AI-driven risk stratification, malignancy classification, early detection, and personalised treatment planning.

3. ANALYTICAL STUDY (LITERATURE REVIEW METHOD)

This work is intended as a thorough analytical review of the literature, rather than as an experimental bench or clinical trial study. The review is in narrative format and employs a synthesised, reproducible literature search strategy. Combinations of keywords were used to search six large electronic databases: PubMed, IEEE Xplore, Scopus, Web of Science, ScienceDirect and Google Scholar. The keywords used were: breast cancer, artificial intelligence, machine learning, deep learning, screening, segmentation, classification and recurrence prediction. Common variations of terms such as 'explainable', 'transparency', 'black box', 'understandable', and 'comprehensible' were employed in this search. Peer-reviewed journal articles and conference papers utilising AI- or ML-based techniques to diagnose or prognose breast cancer based on imaging, clinical, genetic or multimodal data were included. The literature was primarily reviewed from 2012 to 2024 to provide a methodological and historical background. Articles written in languages other than English, editorials, and studies containing an insufficient methodological description, or lacking a clear focus on diagnostic or prognostic tasks in breast cancer, were excluded. After screening the titles and abstracts, a full-text assessment was performed to verify relevance and quality and identify further studies using a snowballing strategy based on the references in the included articles.

4. FINDINGS AND DISCUSSIONS

4.1. Breast Cancer Diagnosis: An Overview

Modern methods of breast cancer diagnosis increasingly incorporate both imaging and non-imaging data. These datasets can be analysed using machine learning (ML) algorithms to identify areas of concern that may indicate the presence of tumours. These methods produce clinically significant data at various stages of the diagnostic process and have been found to enhance the early detection of breast cancer [26].

One of the key diagnostic activities is malignancy classification, which defines the presence of benign or malignant abnormalities and directly affects further clinical management [27]. Machine learning (ML) models process image-derived features to support this process. These features include shape, texture, and intensity patterns. When trained on large, varied datasets, these models can determine the probability of

malignancy, thereby helping clinicians to make evidence-based decisions [28]. Furthermore, breast cancer is biologically heterogeneous and can be subclassified into molecular subtypes according to specific biomarkers, each with different prognostic and therapeutic implications. Classifying subtypes, such as triple-negative, HER2-positive and hormone receptor-positive breast cancers, is thus an essential aspect of current diagnostic and treatment planning interventions [29]. By classifying instances into different subgroups, clinicians can develop more disciplined and selective treatment plans, making them available to those in a position to incorporate them into their treatment plans. To better predict subtypes and optimise treatment plans, machine learning algorithms analyse genetic profiles, patterns of gene expression, and clinical history.

The process of dividing an image or area into segments of interest is called segmentation [26]. In the diagnosis of breast cancer, imaging data segmentation is relevant for identifying any suspicious lesions or defining tumour boundaries [30]. It is an important stage in defining the extent and configuration of the tumour, forming the basis for further work, such as determining tumour volume or the data obtained. Performance has improved significantly, which helps detect breast cancer early due to the effectiveness of machine learning, particularly deep learning, in identifying breast lesions in medical images [31].

Another major aspect of breast cancer therapy and subsequent follow-up is the ability to predict metastasis formation and recurrence of cancer [32]. This involves estimating the likelihood of cancer developing again or spreading to other parts of the body. Since clinical markers are useful for measuring the risk of relapse or metastasis, machine learning models — which, by definition, use statistical algorithms — can handle multiple forms of input data, such as imaging, genetic, clinical and electronic health records data. These predictions can therefore be successfully implemented to tailor further patient treatment, aiming to avoid adverse effects and improve outcomes.

These activities should be combined and modelled together to create a better, more dynamic breast cancer diagnosis system. Specifically, to enhance the accuracy of subtype classification, it is necessary to incorporate complementary processes such as tumour segmentation and histological grading, which provide important structural and morphological data critical for accurate cancer characterisation [33].

4.2. Explainable Artificial Intelligence (XAI)

In machine learning (ML), interpretability is the ability to understand how a model operates and makes predictions. Perhaps more importantly, this type of model does not claim to provide a specification as to why it acts in a certain way, but rather provides an organic view of how it actually operates. Therefore, it could be assumed that not every interpretable model is fully explainable, despite interpretability being one of the main factors that maximises explainability. By 'interpretability', we also imply making forecasts regarding other potential situations and evaluating the impact of modifying some inputs. Considering both obvious and subtle aspects, the purpose is to gain a deeper understanding of the ML model and provide a more detailed description of its activity.

Nonetheless, several AI models are not easily interpretable, meaning the full picture behind the result cannot be seen. Due to this lack of transparency, it is difficult to understand their decision-making processes. While asking 'why' is the most one can do to determine how a model arrived at a decision based on the inputs, explainability goes one step further by showing how a model reacts to changes in the inputs and how the output changes as a result. This distinction is particularly critical in a medical institution, where decisions have consequences for

people's lives. A doctor being unable to prescribe medication to their patients without knowing how the drugs work is an example of when the doctor should use algorithms without understanding how the recommendations are generated. Practitioners lose trust in, and the willingness to use, AI technology, especially when they are unable to ascertain the results produced by the technology. For AI approaches to be trustworthy and just, they must be clarified and comprehended, particularly in fields such as medicine, finance and law, which have far-reaching impacts.

There are two types of explainable AI: post-hoc and intrinsic. The former is referred to as transparent models because they are inherently comprehensible and explainable. Conversely, post-hoc explainability techniques fall into two categories: model-specific and model-agnostic. Post-hoc or surrogate models replicate the decision-making processes of models that would otherwise be incomprehensible. These models are unique, like a model's structure, and comprise saliency maps and Grad-CAM. However, SHAP, LIME and other model-agnostic techniques can also help shed light on the prediction mechanisms of many models.

XAI methods can be divided into two broad groups: global and local explanations. Global explainability describes the model's general behaviour, exposing its general decision-making trends and feature dependencies throughout the dataset. Conversely, local explainability focuses on explaining why specific predictions have occurred, providing case-specific information on how particular inputs affect a single model output.

The following sections discuss the XAI models that have been most actively used alongside ML and deep learning (DL) models in breast cancer studies. These models are then explained, along with a summary of the relevant studies in the field of breast cancer.

4.3. Shapley Additive Explanations (SHAP)

In ML, interpretability is the ability to understand how a model operates and makes predictions. Perhaps more importantly, this type of model did not claim to provide a specification as to why it did what it did; rather, it provided an organic view of how it actually operated. Therefore, it could be assumed that not every interpretable model is fully explainable, despite interpretability being one of the main factors that maximises explainability. By 'interpretability', we also mean making forecasts about other potential situations and evaluating the impact of changes to some inputs. Considering both obvious and subtle aspects, the purpose is to gain a deeper understanding of the ML model and to provide a more detailed description of its activity.

SHAP [34], proposed by Lundberg, is a powerful, interpretable framework for understanding the predictions of ML models and assigning importance to each feature in relation to a given output. Local and global interpretability: This allows the contribution of each attribute to the overall prediction to be quantified and used to demystify complex models.

Based on the concept of Shapley values in game theory, SHAP allocates the prediction fairly across features in a manner similar to the allocation of rewards among participants in a cooperative game.

The SHAP values can be expressed as follows:

$$\phi i(f,x) = \sum_{S \subseteq F \setminus \{i\}}^{\infty} \left(\frac{|\mathbf{S}|!(|\mathbf{F}|-|\mathbf{S}|-\mathbf{1})!}{|\mathbf{F}|!} \left[\mathbf{f}(\mathbf{S} \cup \{\mathbf{i}\}) - \mathbf{f}(\mathbf{S}) \right] \right) \quad (1)$$

where ϕi represents the SHAP value for feature i, S denotes a subset of features excluding i, F is the set of all features, and f (S) is the model's output when only the features in S are considered.

- **Local Accuracy:** This principle ensures that the sum of the SHAP values equals the difference between the model's prediction for a given input and the average model output. Formally:

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M}(\phi_i x'_i) \quad (2)$$

where ϕ0 is the average model output over the entire dataset, and ϕi are the SHAP values for each feature.

- **Feature Absence (Missingness):** This means that when a feature is not used in the model (i.e. the values of it are either zero or not present in the model), its SHAP value should be zero:

$$x_i = 0 ==> \phi_i = 0 \quad (3)$$

- **Consistency:** According to this principle, when adding a feature to the model increases the impact of a feature on the prediction The SHAP values of the feature should increase. For two models, f and f ', if:

$$f'_x(z') - f'_x(z' \setminus i) \geq f_x(z') - f_x(z' \setminus i) \quad (4)$$

for all subsets $z' \in \{0, 1\}^M$, then:

$$\phi_i(f', x) \geq \phi_i(f, x) \quad (5)$$

SHAP offers a mathematically sound and consistent framework for attributing feature importance, aiding in the interpretability and trustworthiness of machine learning models.

4.4. Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM [35] is a visual explanation tool for a wide range of convolutional neural network (CNN) models. Initially introduced by Selvaraju et al., it identifies and highlights important regions in an input image that are useful for model class predictions. This is achieved by computing the gradients of any target class with respect to the activations of the final convolutional layer to develop a heatmap of the significant regions. Grad-CAM is more versatile than CAM and can be used with other CNN architectures without any design modifications being required. It produces informative visualisations by calculating gradients and transforming the feature maps of the final convolutional layer to differentiate between classes.

- **Calculating the gradient:** Grad-CAM involves computing the gradient of the loss with respect to the activations in the final convolutional layer:

$$\frac{\partial L}{\partial A^k} \quad (6)$$

This step indicates the extent of the contribution made by each region to the loss of the activation maps which were involved in the decision making process of each region.

- **Gradient Global Average Pooling (GAP):** The significance weights αk for each channel in the activation maps of the gradients are then calculated using the GAP operation.

$$a^k = \frac{1}{z}\sum_{S \subseteq F \setminus \{i\}}^{\infty} \sum \frac{\partial L}{\partial A^k} \quad (7)$$

Here, it will be convenient to assume that the total number of components in the activation map A^k is equal to the product of H × W, where Z, and the weight of each channel's contribution to the model output is given by α^k.

- **Complete Grad-CAM Formula:** Such weights are subsequently multiplied by the activation maps, and the ReLU function is added to it to create the Grad-CAM heatmap:

$$\text{Grad} - \text{CAM}_c = ReLU \sum_k (a^k A^k) \quad (8)$$

In this formula, the roles of the weighted activation maps are demonstrated to bring out the areas of the input image that are predicted. The ReLU function assists in the promotion of interpretability in the manner by which the real behavior of the model is executed, while ensuring only characteristics that support the target class are illustrated.

4.5. Local Interpretable Model-Agnostic Explanations (LIME)

Ribeiro et al. [36] proposed LIME, an approach for generating locally faithful models and one that provides easily interpretable explanations for a decision made by a given model. The approach identifies a surface that closely models a complex decision boundary of a given machine learning model for the specific data instance that needs to be explained. To explain the behavior of the original complex model near that instance, LIME uses a simpler model with improved interpretability.

- **Comprehensible Data Display:** This is among the key attributes of LIME since it offers the framework a direction on which features are interpretable, and which are complex. LIME reduces the representation to a human-understandable level. For example, in text classification, the model itself may employ more complex features like word embeddings, whereas the explanation may utilize a binary vector that indicates the presence or absence of particular words. Similarly, even if the model employs raw pixel values or other image attributes for image classification, LIME may represent pictures based on the existence of super-pixels.

  This might be converted into an interpretable binary form $x' \in \{0, 1\}^{d'}$ in the context of an instance $x \in R^d$, where each element of x' represents a reduced feature:

$$x \in R^d \quad (9)$$

$$x' \in \{0,1\}^{d'} \quad (10)$$

The binary vector x′ is then used for generating human-understandable explanations.

- **Trade-off Between Integrity and Interpretability:** Interpretable and true to the original model explanations are the goals of LIME. The explanation model, represented by the notation g ∈ G is selected from a collection G of interpretable models, including rule-based systems, decision trees, and linear models. The interpretable feature space is represented by $\{0, 1\}^d$, which is the domain of g. The symbol Ω(g) represents the explanatory model's complexity, which may be interpreted as the number of non-zero weights in a linear model or the depth of a decision tree.

  The degree to which the explanatory model g approximates the original model f close to the instance x is measured by the fidelity function L(f, g, π_x), where $\pi_x(z)$ specifies the proximity of the instance z to x. LIME balances interpretability and fidelity by optimizing the following equation:

$$\varepsilon(x) = \arg min_{g \in G} \left(L\left(F, G, \pi_X\right) + \Omega(g) \right) \quad (11)$$

- **Local Approximation Sampling:** By sampling data points in the vicinity of x′, LIME approximates the local fidelity function L(f, g, πx). This is carried performed without making any assumptions regarding the original model's structure f. Each of these modified samples is labelled by the original model, forming a dataset Z. Using this dataset to solve equation (10) balances simplicity and fidelity while offering a locally correct and interpretable explanation ξ(x).
- **Diffuse Linear Interpretations:** LIME looks for a sparse linear model for interpretability, where G contains linear models and the fidelity function L is selected to be a locally weighted square loss. To give points near x greater weight, a kernel function πx(z) is used. For this task, the loss function is defined as follows:

$$L(f,g,\pi_X) = \sum_{z,z' \epsilon Z}^{x} \pi_x(z)\big(f(z) - g(z')\big)^2 \quad (12)$$

LIME uses regularisation techniques like Lasso to reduce the amount of characteristics in the explanation while maintaining interpretability. Next, restrictions are used to regulate the model's complexity, usually with the use of an indicator function:

$$\Omega(g) = \infty \cdot 1\left[\left|\left|w_g\right|\right|_0 > K\right] \quad (13)$$

In this case, K stands for the maximum number of non-zero weights that may be included in the linear model while maintaining a concise and understandable explanation.

By using simplified approximations, LIME efficiently generates straightforward, locally accurate explanations that assist users in comprehending and interpreting the workings of complicated models, including deep learning models.

4.6. Datasets for Breast Cancer Diagnosis

As shown in Figure 1, the dataset used to diagnose breast cancer consists of both clinical image data and non-image data [37]. Radiological and pathological pictures make up clinical image data. MRI, CT, thermal imaging, mammography, and ultrasound are examples of radiology pictures, whereas histopathology and pCLE are examples of pathology images. Clinical and non-clinical data are further classifications for non-image data. Laboratory findings, radiography and pathology reports, and narrative summaries of the patient's condition are all considered clinical data. Age, patient history, demographics, and genetic data are examples of non-clinical data [38].

Furthermore, there are two types of non-image data: organized and unstructured. Pathology reports and patient profiles are classified as structured data, whereas radiology reports and narrative patient descriptions are classified as unstructured data [39]. This work focuses mostly on histopathology-based datasets, especially in a multi-modal setting, even though there are many image and non-image datasets available for breast cancer detection. The table shows that most datasets have small sample sizes, and there are far fewer multi-modal datasets than unimodal datasets.

Many datasets, each offering distinct insights, improve the study of breast cancer histology. BRACS [40] and BreCaHAD [41] are examples of unimodal datasets that concentrate on a single kind of data. For instance, three qualified pathologists have annotated the 162 histopathology photos in BreCaHAD, which focus on malignant cases, and also include annotations for tubules, non-tubules, tumor nuclei, apoptosis, and mitosis.

Multi-modal datasets, on the other hand, integrate many data sources to provide a more thorough picture of the pathophysiology of breast cancer

[42] combines pathology pictures, CNVs, and gene expression information from 1.098 patients with breast cancer. A deeper understanding of the molecular and histological features of breast cancer is possible because of this multifaceted dataset. The IMPRESS dataset, which includes 126 patients' whole-slide images (WSIs) stained with Hematoxylin and Eosin (HE), along with biomarker annotations and extra clinical data, is another example. 96 WSIs and clinical information, such as the status of the human epidermal growth factor receptor 2 (HER2), progesterone receptor (PR), and estrogen receptor (ER), are provided by the Post-NAT BRCA38 dataset [43].

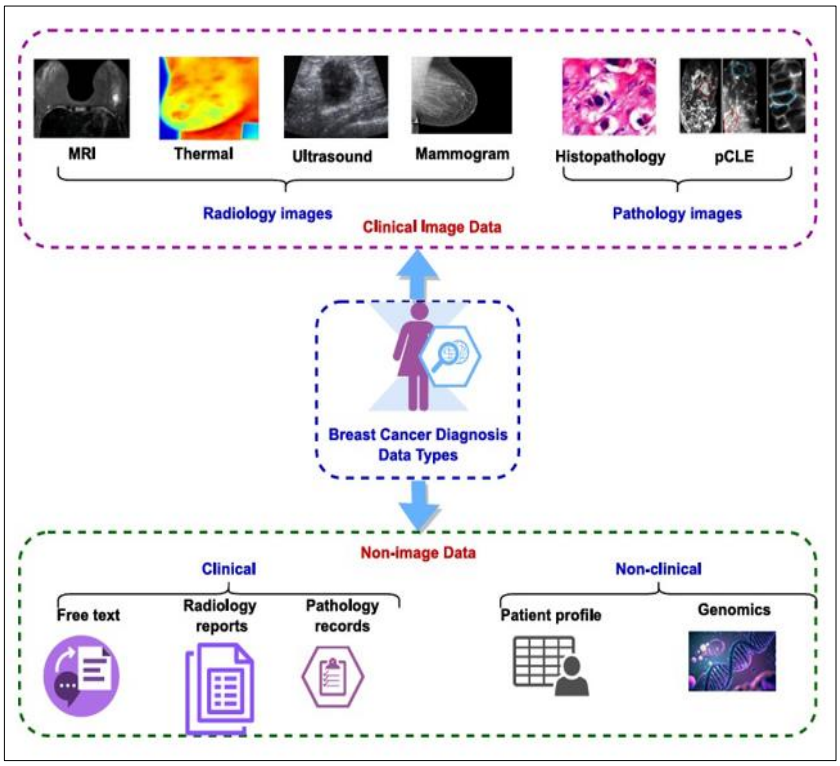

Figure 1. Types of breast cancer diagnosis data

126 HE-stained WSIs from 64 patients with triple-negative breast cancer and 62 patients with HER2-positive breast cancer, all of whom underwent neoadjuvant chemotherapy before surgery, are included in the IMPRESS dataset [44]. Additionally, it contains WSIs stained with immuno-histochemistry (IHC) and the associated scores, which are scanned at 20× magnification. The GTEx Project [45] provides histology images of normal human tissues, including breast (mammary) tissue, along with comprehensive gene expression data and sample metadata. The GTEx histology slides are available at high resolution (e.g., 20× magnification) and cover samples from both female and male donors.

134 patients with invasive breast cancer had 642 WSIs scanned at 20× magnification with resolutions of 0.25 and 0.5 µm/pixel in the CPTAC-BRCA dataset [46] compiled by the Clinical Proteomic Tumor Analysis Consortium. Proteomic, genetic, and clinical data are added to this dataset. The only publicly accessible breast histopathology WSI collection from Asia, the BCNB dataset [47], includes 1.058 WSIs from Chinese patients that were scanned with an Iscan Coreo scanner. The dataset contains extensive clinical information. Two pathologists annotated the tumor locations.

For a thorough assessment of multi-modal techniques, we chose datasets that provide thorough information on multiple modalities, including imaging, clinical records, and genetic data. High-quality datasets must

have accurate genetic data, lots of comprehensive and standardized clinical information, as well as high-resolution pictures. Secondly, it is also necessary to ensure the selection criteria of datasets by which they represent a variety of demographics. Criteria that should run by exclusion should weed out datasets that do not satisfy these requirements, e.g., datasets with poor or missing data. This meticulous screening process ensures that the study is representative of real clinical settings and generalizable to other patient groups [48]. As an example, TCGA-BRCA [42] was chosen because it has many types of cancer and quite complete genomic information, whereas structured datasets such as the EMR dataset cite54 contain complete data. Thus, in the context of precisely characterizing the inclusion and exclusion criteria, we want to guarantee the robustness and applicability of the multimodal method assessed in this work by precisely specifying these inclusion and exclusion criteria, such that the datasets we choose are complete, high-quality quality and representative across a number of clinical situations.

Table 1. Summary of datasets and methods employed in breast cancer detection studies

Dataset	Dataset Type	Method	Task	Modality
Proprietary	Proprietary	Unet3+ [49]	Segmentation	Ultrasound
CBIS-DDSM [50], Inbreast [51], Proprietary	Mixed (public + proprietary)	Yolo-based model [52]	Detection	Mammogram (DM)
TCGA [42]	Public	moBRCA-net [53]	Sub-type classification	Multi-omics
BreakHis [54]	Public	Hybrid CNN-LSTM [55]	Grade classification	Histopathology
Databiox [56]	Public	Ensemble CNN [28]	Grade classification	Histopathology
BUSI [57], Mini- DDSM [58]	Public	Hybrid CNN [59]	Detection	Mammogram, ultrasound images
Proprietary	Proprietary	KAMnet [60]	Detection	Ultrasound
Proprietary	Proprietary	Classifier-combined method [27]	Grade classification	MRI
WPBC	Public	Recurrence prediction [61]	Recurrence and metastasis	Clinical data
CBIS-DDSM [50], MIAS [62]	Public	Semantic segmentation [63]	Segmentation	Mammogram
BreakHis [54], IR Thermal Images [64]	Mixed (public + public)	EMDCOC [65]	Detection	Histopathology, IR thermal images
MIAS [62]	Public	Optimized LSTM with U-net segmentation [66]	Segmentation	Mammogram
TCGA [42]	Public	Multi-modal fusion [67]	Detection	WSI, gene expression
Ultrasound Image dataset [68], BUSI [57]	Public	DeepBreast CancerNet [69]	Detection	Ultrasound
TCGA [42]	Public	DSCCN [29]	Sub-type classification	Multi-omics
Proprietary	Proprietary	Prediction model for distant metastasis [70]	Recurrence and metastasis	Clinical data
BreakHis [54]	Public	Histogram K-Means segmentation [71]	Segmentation	Histopathology
Abbreviations and Terminology	—	LSTM: Long Short-Term Memory	Detection: lesion / cancer identification; Segmentation: lesion boundary delineation; Subtype / Grade classification: molecular / grade categorization	DM: Digital Mammography; IR: Infrared

4.7. Related Works

AI systems can be employed in different manners for DM or DBT based breast cancer screening programs, according to different implementing way according to the particular requirements and preferences of each screening program. In Table 2, this article presents the main approaches suggested for using AI in breast cancer screening in the scientific literature. Concurrent AI decision help during mammography interpretation is a popular strategy. For instance, Pacile et al. [72] showed that while radiologists' reading time per case rose by 9–14%, AI-enhanced readings improved their AUC from 76.9% to 79.7%. Similarly, Conant et al. [73] found that AI helps decrease radiologists' reading time by 53% while increasing their AUC from 0.80 to 0.895. AUC improvements with AI help were also seen in a number of other studies, including those by researchers in [74] and author in [47]. However, the effect on reading time varied, with some reporting quicker reading times and others finding no discernible change.

Using AI as a stand-alone secondary reading in screening procedures is another tactic. In an upcoming paired research, for example, Dembrower et al. [75] showed that the application of AI reduced the number of screening readings by 50% while increasing the cancer detection rate (CDR) by 4% and decreasing the recall rate by 4%. Other retrospective investigations, including those by [76, 77, and 78], provided similar findings, demonstrating lower recall rates and less workload, but with differing effects on CDR.

Additionally, AI may be used as a triage tool, with high-risk examinations being double-read and low-risk exams being single-read. For instance, writers in [79] found that using AI in this way reduced the number of reads by 44% while increasing CDR by 20% while maintaining the same recall rate. A similar strategy was also shown by the authors in [76], who showed the same CDR with a 35% reduction in reading burden and a 9% lower recall rate.

A more automated triage approach has also been suggested, in which only high-risk tests are double-read and low-risk exams are automatically classified as normal. According to writers in [80], this method reduced reading burden by 19% while maintaining the same sensitivity and recall rate. Similar methods were used by writers in et al. [81] for both DM and DBT, showing workload reductions of up to 72% and 17% lower recall rates. Furthermore, according to Sauthors in [82], AI triage reduced effort by 40% and recall rate by 25% while maintaining no inferior sensitivity.

As anticipated, the way AI is applied will determine its possible influence on breast cancer screening. While some solutions may boost sensitivity at the price of increased false positive rates, others may drastically reduce the effort without compromising sensitivity. Although Lang et al. [79 and 83] appear to be moving in the right direction, it is still unclear from empirical data if AI in screening might lead to concurrent gains in workload, sensitivity, and specificity.

It's crucial to remember that the many applications of AI in breast cancer screening can be coupled and are not exclusive of one another. Despite being one of the most studied subfields of radiological AI, there is still little proof of breast imaging AI's efficacy. A large number of research studies are retrospective in nature, examining just a small number of commercially accessible AI systems using data that is frequently not typical of screening programs throughout the world. Dataset enrichment, a dearth of research examining how radiologists' behavior may alter when AI is used, and the limited clinical significance of some findings—such as the aggressiveness or laziness of tumors found by screens—are other problems. Also, the range of possible applications for any AI system is limited by the unique characteristics and intended

purpose of any such a system. A classic example is that even if this is a part of a double reading scenario, an AI system that is cleared by the regulators for use as concurrent decision assisting equipment for mammography cannot be used as a stand-alone reader in the screening.

Table 2. Summary of AI strategies for breast cancer screening and their effectiveness

Publication	AI implementation strategy	Evaluation setting	Dataset / Modality	Effect on Screening / Performance	Effect on Workload
Pacile et al. [72]	Concurrent AI decision support	Reader-based evaluation	Screening mammography (DM)	AUC increased from 0.769 to 0.797	Reading time increased by 9–14%
Conant et al. [73]	Concurrent AI decision support	Reader-focused study (retrospective)	DBT	AUC increased from 0.80 to 0.895	Reading time reduced by 53%
Rodriguez-Ruiz et al. [74]	Concurrent AI decision support	Reader analysis experiment (retrospective)	Screening mammography (DM)	AUC increased from 0.87 to 0.89	No significant change in reading time
Van Winkel et al. [47]	Concurrent AI decision support	Multi-reader multi-case study (retrospective)	DBT	AUC increased from 0.83 to 0.86	Reading time reduced by 12%
Dembrower et al. [75]	AI as a stand-alone reader	Paired prospective Analysis	Screening mammography (DM)	CDR increased by 4%; recall rate reduced by 4%	Screening workload reduced by ~50%
Larsen et al. [76]	AI as a stand-alone reader	Retrospective performance analysis	Screening mammography (DM)	Comparable CDR reduction in recall rate	Screen readings reduced by ~50%
Sharma et al. [77]	AI as a stand-alone reader	Historical data assessment	Screening mammography (DM)	Non-inferior cancer detection and recall rates	Workload reduced by 30–45%
Leibig et al. [78]	AI as a stand-alone reader	Retrospective comparative review	Screening mammography (DM)	Lower CDR with increased recall rate	Screening requirements reduced by ~50%
Lång et al. [79]	AI as a triage tool	Randomized controlled experiment	Screening mammography (DM)	20% higher CDR trend with stable recall rate	44% fewer screen readings
Lång et al. [80]	AI triage for auto-normal labeling	Retrospective data analysis	Screening mammography (DM)	Maintained sensitivity and recall rate	19% fewer readings
Raya-Povedano et al. [81]	AI triage for auto-normal labeling	Historical review	Digital mammography (DM) and DBT	Maintained sensitivity with 17% lower recall rate (P < .001)	71% fewer DM and 72% fewer DBT readings
Lauritzen et al. [84]	AI triage for auto-normal labeling	Retrospective out-come evaluation	Screening mammography (DM)	Maintained sensitivity with 19% lower recall rate	Workload reduced by 63%
Dembrower et al. [85]	AI triage for auto-normal labeling	Retrospective impact analysis	Screening mammography (DM)	Potential gain of 71 additional cancers detected per 1,000 cases	60% fewer readings
Shoshan et al. [82]	AI triage for auto-normal labeling	Retrospective efficiency study	DBT	Maintained sensitivity with 25% lower recall rate (P = .002)	40% fewer DBT readings
Abbreviations and performance metrics				Recall rate: proportion of screened cases recalled for additional assessment	Sensitivity: true positive rate; Specificity: true negative rate

5. CONCLUSION AND RECOMMENDATIONS

The analysed literature demonstrates that artificial intelligence has significant potential to improve breast cancer screening and diagnostic processes, but also highlights limitations that prevent its regular clinical use. AI systems based on imaging and applied to mammography, ultrasound, MRI and histopathology are highly sensitive and display higher cancer detection rates, particularly when used to support concurrent decision-making or triage. However, these advantages are often accompanied by increased false-positive/recall rates, raising concerns about overdiagnosis and unnecessary follow-up. Most reported performance improvements are based on retrospective or enriched datasets that do not necessarily reflect the actual screening population in real-world situations, adding to dataset bias and reducing generalisability. Multimodal AI methods, which combine imaging with clinical or molecular data, are better at classifying subtypes and demonstrate better prognostic performance. However, their clinical implementation is limited by data scarcity, a lack of interoperability and the complexity of integrating data from multiple sources. Explainable AI algorithms, such as SHAP, Grad-CAM and LIME, enhance transparency and clinician confidence. However, their post-hoc disposition and sensitivity to model and data fluctuations can lead to unreliable and possibly misguided explanations. Furthermore, the lack of external and prospective validation, regulatory and implementation obstacles, and imbalances in training datasets continue to limit real-world deployment. These issues could be resolved through the implementation of standardised evaluation systems, varied and high-quality data, effective explainability policies, and future multicentre clinical trials to validate the reliable, fair, and clinically significant advantages of AI systems in breast cancer diagnosis.

Considerable advances have been made in the use of artificial intelligence (AI) in breast cancer screening and diagnosis. AI has demonstrated its ability to improve diagnostic accuracy, reduce interobserver variability and streamline radiological and pathological processes. AI systems can conveniently be used as co-decision support systems, independent readers or triage systems. This means they can be flexibly implemented depending on the needs and available resources of screening programmes, as outlined in this review. These applications demonstrate AI's potential to boost efficiency and provide more personalised patient management.

However, large-scale clinical translation cannot be achieved unless a number of crucial issues are resolved. Most of the available evidence is based on retrospective and single-centre studies, which may not reflect the heterogeneity of real-life screening populations, imaging protocols, and clinical practice. Subsequent studies should therefore focus on large-scale, prospective clinical trials and federated and multicentre datasets, so that their impact is more robust and representative of other demographic groups. Standardised evaluation metrics and reporting systems are needed to facilitate meaningful comparisons between AI systems, as well as to facilitate regulatory approval. Alongside explanations of AI and the clinician-in-the-loop approach, these metrics and systems should be integrated to enhance transparency, trust and clinical accountability. Addressing these methodological, technical and regulatory gaps is key to enabling AI to safely, fairly and sustainably assist radiologists and pathologists in the routine management of breast cancer, rather than replacing them.

CONFLICT OF INTEREST

The author(s) declare that they have no potential conflict of interest.

FINANCIAL DISCLOSURE

This research received no financial support.

DECLARATION OF ETHICAL STANDARDS

The authors of the article declare that the materials and methods used did not require ethics committee approval and/or regulatory approval.

REFERENCES

[1] Giaquinto, A.N., Sung, H., Miller, K.D., Kramer, J.L., Newman, L.A., Minihan, A., et al., (2022). Breast cancer statistics, CA Cancer J Clin, 72:524–541.

[2] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., et al., (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 71:209–249.

[3] Taylor, C., McGale, P., Probert, J., Broggio, J., Charman, J., Darby, S.C., et al., (2023). Breast cancer mortality in 500 000 women with early invasive breast cancer diagnosed in England, 1993–2015: population based observational cohort study. BMJ, 381:e074684.

[4] Marmot, M.G., Altman, D.G., Cameron, D.A., Dewar, J.A., Thompson, S.G., and Wilcox, M., (2013). The benefits and harms of breast cancer screening: an independent review. Br J Cancer, 108:2205–2240.

[5] The Royal College of Radiologists, (2022). RCR Clinical Radiology Workforce Census 2022.

[6] Metter, D.M., Colgan, T.J., Leung, S.T., Timmons, C.F., and Park, J.Y., (2019). Trends in the US and Canadian pathologist workforces from 2007 to 2017. JAMA Netw Open, 2:e194337.

[7] Connor, S.J., Lim, Y.Y., Tate, C., Entwistle, H., Morris, J., Whiteside, S., et al., (2012). A comparison of reading times in full-field digital mammography and digital breast tomosynthesis. Breast Cancer Res, 14:P26.

[8] Elmore, J.G., Longton, G.M., Carney, P.A., Geller, B.M., Onega, T., Tosteson, A.N., et al., (2015). Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA, 313:1122–1132.

[9] Acs, B., Fredriksson, I., Rönnlund, C., Hagerling, C., Ehinger, A., Kovács, A., et al., (2021). Variability in breast cancer biomarker assessment and the effect on oncological treatment decisions: a nationwide 5-year population-based study. Cancers (Basel), 13:1166.

[10] Fernandez, A.I., Liu, M., Bellizzi, A., Brock, J., Fadare, O., Hanley, K., et al., (2022). Examination of low ERBB2 protein expression in breast cancer tissue. JAMA Oncol, 8:1–4.

[11] Kim, S.H., Lee, E.H., Jun, J.K., Kim, Y.M., Chang, Y.W., Lee, J.H., et al., (2019). Interpretive performance and inter-observer agreement on digital mammography test sets. Korean J Radiol, 20:218–224.

[12] Whitney, J., Corredor, G., Janowczyk, A., Ganesan, S., Doyle, S., Tomaszewski, J., et al., (2018). Quantitative nuclear histomorphometry predicts Oncotype DX risk categories for early stage ER+ breast cancer. BMC Cancer, 18:610.

[13] Rajpurkar, P., Chen, E., Banerjee, O., and Topol, E.J., (2022). AI in health and medicine. Nat Med, 28:31–38.
[14] Kann, B.H., Hosny, A., and Aerts, H.J., (2021). Artificial intelligence for clinical oncology. Cancer Cell, 39:916–927.
[15] Niazi, M.K., Parwani, A.V., and Gurcan, M.N., (2019). Digital pathology and artificial intelligence. Lancet Oncol, 20:e253–e261.
[16] Hickman, S.E., Baxter, G.C., and Gilbert, F.J., (2021). Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. Br J Cancer, 125:15–22.
[17] Cardoso, F., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rubio, I.T., et al., (2019). Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol, 30:1194–1220.
[18] Gradishar, W.J., Moran, M.S., Abraham, J., Aft, R., Agnese, D., Allison, K. H., et al., (2022). Breast cancer, version 3.2022, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw, 20:691–722.
[19] Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M.O., Lehman, C.D., et al., (2007). American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin, 57:75–89.
[20] Tice, J.A., Miglioretti, D.L., Li, C.S., Vachon, C.M., Gard, C.C., and Kerlikowske, K., (2015). Breast density and benign breast disease: risk assessment to identify women at high risk of breast cancer. J Clin Oncol, 33:3137–3143.
[21] Gail, M.H., (2020). Choosing breast cancer risk models: importance of independent validation. J Natl Cancer Inst, 112:433–435.
[22] Holm, J., Li, J., Darabi, H., Eklund, M., Eriksson, M., Humphreys, K., et al., (2016). Associations of breast cancer risk prediction tools with tumor characteristics and metastasis. J Clin Oncol, 34:251–258.
[23] Gilbert, F.J., Tucker, L., Gillan, M.G., Willsher, P., Cooke, J., Duncan, K. A., et al., (2015). The TOMMY trial: a comparison of tomosynthesis with digital mammography in the UK NHS Breast Screening Programme – a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. Health Technol Assess, 19:i–ixxv, 1–136.
[24] Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., and Reede, D., (2017). Interpretive error in radiology. American Journal of Roentgenology, 208(4):739-749.
[25] Redondo, A., Comas, M., Macià, F., Ferrer, F., Murta-Nascimento, C., Maristany, M. T., et al., (2012). Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol, 85:1465–1470.
[26] Rai, H.M., (2024). Cancer detection and segmentation using machine learning and deep learning techniques: a review. Multimed Tools Appl, 83:27001–27035.
[27] Liu, Z., Lin, F., Huang, J., Wu, X., Wen, J., Wang, M., et al., (2023). A classifier-combined method for grading breast cancer based on Dempster-Shafer evidence theory. Quant Imaging Med Surg, 13:3288.
[28] Kumaraswamy, E., Kumar, S., and Sharma, M., (2023). An invasive ductal carcinomas breast cancer grade classification using an ensemble of convolutional neural networks. Diagnostics, 13:1977.

[29] Huang, Y., Zeng, P., and Zhong, C., (2024). Classifying breast cancer subtypes on multi-omics data via sparse canonical correlation analysis and deep learning. BMC Bioinformatics, 25:132.
[30] Guo, H., Li, M., Liu, H., Chen, X., Cheng, Z., Li, X., et al., (2024). Multi-threshold image segmentation based on an improved salp swarm algorithm: case study of breast cancer pathology images. Comput Biol Med, 168:107769.
[31] Rajoub, B., Qusa, H., Abdul-Rahman, H., and Mohamed, H., (2024). Segmentation of breast tissue structures in mammographic images. In: Artificial Intelligence Image Processing in Medical Imaging, pp. 115–146.
[32] Soliman, A., Li, Z., and Parwani, A.V., (2024). Artificial intelligence's impact on breast cancer pathology: a literature review. Diagn Pathol, 19:1–18.
[33] Gallagher, W.M., McCaffrey, C., Jahangir, C., Murphy, C., Burke, C., and Rahman, A., (2024). Artificial intelligence in digital histopathology for predicting patient prognosis and treatment efficacy in breast cancer. Expert Rev Mol Diagn, 24:363–377.
[34] Lundberg, S.M., and Lee, S.-I., (2017). A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, 30.
[35] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., (2017). Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.
[36] Ribeiro, M.T., Singh, S., and Guestrin, C., (2016). Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.
[37] Sweetlin, E.J., and Saudia, S., (2021). A review of machine learning algorithms on different breast cancer datasets. In: International Conference on Big Data, Machine Learning, and Applications, pp. 659–673. Springer.
[38] Heiliger, L., Sekuboyina, A., Menze, B., Egger, J., and Kleesiek, J., (2023). Beyond medical imaging – a review of multimodal deep learning in radiology. Authorea.
[39] Laokulrath, N., Gudi, M.A., Deb, R., Ellis, I.O., and Tan, P.H., (2024). Invasive breast cancer reporting guidelines: ICCR, CAP, RCPath, RCPA datasets and future directions. Diagn Histopathol, 30:87–99.
[40] Brancati, N., Anniciello, A.M., Pati, P., Riccio, D., Scognamiglio, G., Jaume, G., et al., (2022). BRACS: a dataset for breast carcinoma subtyping in HE histology images. Database, 2022:baac093.
[41] Aksac, A., Demetrick, D.J., Ozyer, T., and Alhajj, R., (2019). BReCaHaD: a dataset for breast cancer histopathological annotation and diagnosis. BMC Res Notes, 12:1–3.
[42] The Cancer Genome Atlas (TCGA), (2023). Genomic Data Commons Data Portal (GDC), TCGA-BRCA. Available at: https://portal.gdc.cancer.gov/projects/TCGA-BRCA (accessed July 07, 2023).
[43] Martel, A.L., Nofech-Mozes, S., Salama, S., Akbar, S., and Peikari, M., (2019). Assessment of residual breast cancer cellularity after neoadjuvant chemotherapy using digital pathology. Cancer Imaging Arch.
[44] Huang, Z., Shao, W., Han, Z., Alkashash, A.M., De la Sancha, C., Parwani, A.V., et al., (2023). Artificial intelligence reveals

features associated with breast cancer neoadjuvant chemotherapy responses from multi-stain histopathologic images. NPJ Precis Oncol, 7:14.
[45] The Genotype-Tissue Expression (GTEx) Project, (2023). GTEx portal. Available at: https://gtexportal.org/home/histologyPage (accessed July 07, 2023).
[46] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium, (2020). The Clinical Proteomic Tumor Analysis Consortium breast invasive carcinoma collection (CPTAC-BRCA). The Cancer Imaging Archive. Available at: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70227748 (accessed July 07, 2023).
[47] van Winkel, S.L., Rodriguez-Ruiz, A., Appelman, L., Gubern-Merida, A., Karssemeijer, N., Teuwen, J., Wanders, A.J., Sechopoulos, I., and Mann, R.M., (2021). Impact of artificial intelligence support on accuracy and reading time in breast tomosynthesis image interpretation: a multi-reader multi-case study. Eur Radiol, 31(11):8682–8691.
[48] Yan, R., Zhang, F., Rao, X., Lv, Z., Li, J., Zhang, L., et al., (2021). Richer fusion network for breast cancer classification based on multimodal data. BMC Med Inform Decis Mak, 21:1–15.
[49] Alam, T., Shia, W.C., Hsu, F.R., and Hassan, T., (2023). Improving breast cancer detection and diagnosis through semantic segmentation using the UNet3+ deep learning framework. Biomedicines, 11:1536.
[50] Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L., et al., (2017). Curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data, 4:1–9.
[51] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., and Cardoso, J.S., (2012). INbreast: toward a full-field digital mammographic database. Acad Radiol, 19:236–248.
[52] Prinzi, F., Insalaco, M., Orlando, A., Gaglio, S., and Vitabile, S., (2024). A YOLO-based model for breast cancer detection in mammograms. Cognit Comput, 16:107–120.
[53] Choi, J.M., and Chae, H., (2023). MoBRCa-Net: a breast cancer subtype classification framework based on multi-omics attention neural networks. BMC Bioinformatics, 24:169.
[54] Spanhol, F.A., Oliveira, L.S., Petitjean, C., and Heutte, L., (2015). A dataset for breast cancer histopathological image classification. IEEE Trans Biomed Eng, 63:1455–1462.
[55] Srikantamurthy, M.M., Rallabandi, V.S., Dudekula, D.B., Natarajan, S., and Park, J., (2023). Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Med Imaging, 23:19.
[56] DataBioX, (2024). DataBioX datasets. Available at: https://databiox.com/datasets/ (accessed June 02, 2024).
[57] Al-Dhabyani, W., Gomaa, M., Khaled, H., and Fahmy, A., (2020). Dataset of breast ultrasound images. Data Brief, 28:104863.
[58] Lekamlage, C.D., Afzal, F., Westerberg, E., and Cheddad, A., (2020). Mini-DDSM: mammography-based automatic age estimation. In: 2020 3rd International Conference on Digital Medicine and Image Processing, pp. 1–6. ACM.
[59] Sahu, A., Das, P.K., and Meher, S., (2023). High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. Biomed Signal Process Control, 80:104292.
[60] Guo, D., Lu, C., Chen, D., Yuan, J., Duan, Q., Xue, Z., et al., (2024). A multimodal breast cancer diagnosis method based on

knowledge-augmented deep learning. Biomed Signal Process Control, 90:105843.
[61] Hussein, M., Elnahas, M., and Keshk, A., (2024). A framework for predicting breast cancer recurrence. Expert Syst Appl, 240:122641.
[62] Kendall, E.J., Barnett, M.G., and Chytyk-Praznik, K., (2013). Automatic detection of anomalies in screening mammograms. BMC Med Imaging, 13:1–11.
[63] Ahmed, L., Iqbal, M.M., Aldabbas, H., Khalid, S., Saleem, Y., and Saeed, S., (2023). Images data practices for semantic segmentation of breast cancer using deep neural network. J Ambient Intell Humaniz Comput, 14:15227–1543.
[64] Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S., and Zerhouni, N., (2021). A CNN-based methodology for breast cancer diagnosis using thermal images. Comput Methods Biomech Biomed Eng Imaging Vis, 9:131–145.
[65] Parshionikar, S., and Bhattacharyya, D., (2024). An enhanced multi-scale deep convolutional orchard capsule neural network for multi-modal breast cancer detection. Healthc Anal, 5:100298.
[66] Sivamurugan, J. and Sureshkumar, G., (2023). Applying dual models on optimized LSTM with U-Net segmentation for breast cancer diagnosis using mammogram images. Artif Intell Med, 143:102626.
[67] Liu, H., Shi, Y., Li, A., and Wang, M., (2024). Multi-modal fusion network with intra and inter-modality attention for prognosis prediction in breast cancer. Comput Biol Med, 168:107796.
[68] Paulo, S., (2017). Breast ultrasound image. Mendeley Data.
[69] Raza, A., Ullah, N., Khan, J.A., Assam, M., Guzzo, A., and Aljuaid, H., (2023). DeepBreastCancerNet: a novel deep learning model for breast cancer detection using ultrasound images. Appl Sci, 13:2082.
[70] Murata, T., Yoshida, M., Shiino, S., Ogawa, A., Watase, C., Satomi, K., et al., (2023). A prediction model for distant metastasis after isolated locoregional recurrence of breast cancer. Breast Cancer Res Treat, 199:57–66.
[71] Sahu, Y., Tripathi, A., Gupta, R.K., Gautam, P., Pateriya, R.K., Gupta, A., et al., (2023). Computer aided diagnosis of breast cancer using histogram k-means segmentation technique. Multimed Tools Appl, 82:14055–14075.
[72] Pacile, S., Lopez, J., Chone, P., Bertinotti, T., Grouin, J.M., and Fillard, P., (2020). Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. Radiol Artif Intell, 2(6):e190208.
[73] Conant, E.F., Toledano, A.Y., Periaswamy, S., Fotin, S.V., Go, J., Boatsman, J.E., and Hoffmeister, J.W., (2019). Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. Radiol Artif Intell, 1(4):e180096.
[74] Rodríguez-Ruiz, A., Krupinski, E., Mordang, J.-J., Schilling, K., Heywang-Köbrunner, S.H., Sechopoulos, I., and Mann, R.M., (2019). Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology, 290(2):305–314.
[75] Dembrower, K., Crippa, A., Colon, E., Eklund, M., and Strand, F., (2023). Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit Health, 5:xxx–xxx.

[76] Larsen, M., Aglen, C.F., Hoff, S.R., Lund-Hanssen, H., and Hofvind, S., (2022). Possible strategies for use of artificial intelligence in screen-reading of mammograms, based on retrospective data from 122,969 screening examinations. Eur Radiol, 32(12):8238–8246.
[77] Sharma, N., Ng, A.Y., James, J.J., Khara, G., Ambrozay, C.C., Austin, C.C., Fox, G., Glocker, B., Heindl, A., et al., (2023). Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms. BMC Cancer, 23(1):1–13.
[78] Leibig, C., Brehmer, M., Bunk, S., Byng, D., Pinker, K., and Umutlu, L., (2022). Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. Lancet Digit Health, 4(7):e507–e519.
[79] Lång, K., Josefsson, V., Larsson, A.-M., Larsson, S., Hogberg, C., Sartor, H., Hofvind, S., Andersson, I., and Rosso, A., (2023). Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol, 24(8):936–944.
[80] Lång, K., Dustler, M., Dahlblom, V., Åkesson, A., Andersson, I., and Zackrisson, S., (2021). Identifying normal mammograms in a large screening population using artificial intelligence. Eur Radiol, 31:1687–1692.
[81] Raya-Povedano, J.L., Romero-Martín, S., Elías-Cabot, E., Gubern-Merida, A., Rodríguez-Ruiz, A., and Álvarez-Benito, M., (2021). AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: a retrospective evaluation. Radiology, 300(1):57–65.
[82] Shoshan, Y., Bakalo, R., Gilboa-Solomon, F., Ratner, V., Barkan, E., Ozery-Flato, M., Amit, M., Khapun, D., Ambinder, E.B., Oluyemi, E.T., et al., (2022). Artificial intelligence for reducing workload in breast cancer screening with digital breast tomosynthesis. Radiology, 303(1):69–77.
[83] Lång, K., (2024). Cancer detection in relation to type and stage in the randomised Mammography Screening with Artificial Intelligence trial (MASAI). In: European Congress of Radiology, Vienna, Austria.
[84] Lauritzen, A.D., Rodríguez-Ruiz, A., von Euler-Chelpin, M.C., Lynge, E., Vejborg, I., Nielsen, M., Karssemeijer, N., and Lillholm, M., (2022). An artificial intelligence–based mammography screening protocol for breast cancer: outcome and radiologist workload. Radiology, 304(1):41–49.
[85] Dembrower, K., Wåhlin, E., Liu, Y., Salim, M., Smith, K., Lindholm, P., Eklund, M., and Strand, F., (2020). Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit Health, 2(9):e468–e474.