Engineering Sciences
ISSN: 1308 7231
Article ID: 1A0501

Status : Review
Received: 05.07.2025
Accepted: 20.10.2025

Maral A. Mustafa
Graduate School of Natural and Applied Sciences, Department of Computer Engineering, Gazi University, 24833304401@gazi.edu.tr, Ankara-Türkiye
Production Mechanics Techniques, Kirkuk Technical Institute, Northern Technical University, maralanwer@ntu.edu.iq, Kirkuk-Iraq
O. Ayhan Erdem
Department of Computer Engineering, Faculty of Technology, Gazi University, ayerdem@gazi.edu.tr, Ankara-Türkiye
Esra Söğüt
Department of Computer Engineering, Faculty of Technology, 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 commonly diagnosed cancer in women all over the world, and its prevalence is constantly increasing despite significant advancements in the area of early diagnosis and individual treatment approaches. Nevertheless, present-day workflows in diagnostic interventions are struggling with problems such as overdiagnosis in populations with low risks, growing workloads among radiologists and pathologists, and inconsistencies in the interpretation of the findings of the imaging and pathological studies. In that regard, artificial intelligence (AI) has proven to be an effective solution to these drawbacks by enhancing image analysis, automating the working processes that consume a lot of labor, and facilitating clinical decision-making. This paper provides a narrative review of the recent AI implementation in breast cancer screening and diagnosis, including malignancy detection and classification, tumor segmentation, prediction of molecular subtype, and recurrence or metastatic risk. The data sources are analyzed both in imaging and non-imaging, which are mammography, ultrasound, magnetic resonance imaging (MRI), histopathology, clinical variables, and multi-modal data integration. Also, the reviewed articles identify explainable artificial intelligence (XAI) methods, including SHAP, Grad-CAM, and LIME, as central to improving the transparency, interpretability, and confidence clinicians have in AI-assisted systems. On the whole, the current evidence indicates that AI-based tools have the potential to increase the level of diagnostic accuracy, minimize inter-observer variability, and provide a personalized risk evaluation and treatment planning. However, there are still multiple obstacles to widespread clinical implementation such as heterogeneity of datasets, a lack of external and prospective validation, interpretability issues, and constraints based on real-world application. Future studies must, therefore, focus on the creation of more and better-quality data, standard assessment guidelines, solid explainability models, and future clinical trials to allow the safe, productive, and fair integration of AI into regular breast cancer care.

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

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 AI. Engineering Sciences, 20(4):129-147, DOI: 10.12739/NWSA.2025.20.4.1A0501.

1. INTRODUCTION

Breast cancer has been on a gradual rise and it is currently the most frequently diagnosed malignancy in women all over the world, outdoing lung cancer in incidence [1 and 2]. Although this burden is increasing, screening and the creation of individual treatment options have led to a dramatic drop in mortality related to breast cancer and better patient outcomes [1 and 3]. Nevertheless, there are still significant issues in the diagnostic pathway such as overdiagnosis in low-risk groups, increased strain on the radiology and pathology services, and inconsistency of image and specimen interpretation [4, 5, 6 and 7]. Moreover, lack of access to diagnostic tests, as well as high cost of advanced tests, still impede timely and fair care in most locales [8, 9, 10, 11 and 12]. The diagnostic process of breast cancer is multi-step, usually involving screening, assessment by imaging techniques, tissue biopsy, pathology, staging, molecular and biomarker profiling, and treatment planning, and in some cases, neoadjuvant therapy, surgery, and systemic adjuvant treatments [17 and 18]. However, the existing screening procedures fail to best consider diversity in personal risk, and they might expose the lower-risk groups to additional recall and therapy; in the United Kingdom, an autonomous analysis of randomized trials estimated a 19% risk of overdiagnosis with screening [4]. Despite the use of risk stratification tools to inform intensified surveillance, including annual MRI in women with a lifetime breast cancer risk of ≥20% [19], most models are based on non-routinely measured variables, have low predictive power (often AUC<0.7), and most tend to select those cancers, which have a better prognosis, thus restricting their population-level influence [20, 21 and 22]. At the same time, shortages in the workforce, as well as growing imaging volumes, further burden clinical services: it is estimated that the United Kingdom will be short of radiologists by 40% in 2027 [5], and the workload of a pathologist per practitioner in the United States has gone up by 41.73 in the last ten years [6].

These stresses are augmented by the time expenditure of sophisticated imaging including digital breast tomosynthesis (DBT)and MRI [23], the workload of the pathological procedures (e.g., the extra arrangement of slides) [7], and the consistency of agreement (between 75 and 88 percent) even in focused diagnostic settings [8 and 11]. Diagnostic proficiency in radiology is strongly influenced by clinicians' experience and training, which are known contributors to interpretive accuracy and error susceptibility [24]. However, substantial inter- and intra-reader variability persists in mammography interpretation, reflecting differences in training, experience, and interpretive approaches among radiologists [25]. Additionally, expensive and infrastructure-based tests, such as contrast-enhanced mammography, MRI, and gene tests such as Oncotype Dx are not accessible to all institutions and patients and may need 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 breast cancer diagnosis, enhance the interpretation of the images, automate time-consuming processes, and enabling predictive analytics in the field of radiology and pathology [13, 14, 15 and 16]. In line with this, this review summarizes the existing applications of AI in all major diagnostic tasks and data types to breast cancer care and describes the evidence base that supports their potential clinical utility.

2. RESEARCH SIGNIFICANCE

This review focuses on the role of artificial intelligence in addressing 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, alleviate clinician workload, and support personalized treatment planning. A particular emphasis is placed on radiology and pathology, where AI can assist in

detecting early-stage cancers, characterizing lesions, and predicting outcomes. By integrating image-based and non-image-based data, AI systems can support risk stratification, subtype classification, and recurrence prediction, helping clinicians make more informed decisions and optimize patient management. This review further contributes by highlighting the importance of explainable AI (XAI) approaches, which are essential for building trust, transparency, and accountability in AI-assisted medical decision-making. In high-stakes domains such as oncology, interpretability is a crucial factor for clinical adoption.

Highlights:

- It provides a comprehensive and up-to-date synthesis of artificial intelligence applications in breast cancer diagnosis, systematically outlining how AI addresses major clinical challenges such as overdiagnosis, interpretive variability, and 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, which are essential for safe integration of AI systems in oncology.
- It critically analyzes both imaging-based and non-imaging multimodal datasets, identifying current capabilities, limitations, and future research directions for AI-driven risk stratification, malignancy classification, early detection, and personalized treatment planning.

3. ANALYTICAL STUDY (LITERATURE REVIEW METHOD)

This work is designed as a comprehensive analytical review of the literature rather than an experimental bench or clinical trial study. This review is in narrative format with a synthesized and reproducible literature search strategy. Six large electronic databases, namely PubMed, IEEE Xplore, Scopus, Web of Science, ScienceDirect, and Google Scholar, were searched using combinations of key words, i.e. breast cancer, artificial intelligence, machine learning, deep learning, screening, segmentation, classification and recurrence prediction. This search was employed, using common variations of terms such as explainable, transparency, black box, understandable, and comprehensible. Peer-reviewed journal articles and conference papers utilizing AI- or ML-based techniques to diagnose or prognose breast cancer based on imaging, clinical, genetic, or multi-modal data were included, and the literature was reviewed primarily in the 2012-2024 period to give the methodological and historical background. Articles written in non-English, editorials, and studies that contained an insufficient amount of methodological description and articles that did not have a clear focus on the diagnostic or prognostic tasks in breast cancer were excluded. After the screening of the titles and the abstracts, full-text assessment was performed in order to verify relevance and quality, and to identify more studies through a snowballing strategy that relied on the list of references of the included articles.

4. FINDINGS AND DISCUSSIONS

4.1. Breast Cancer Diagnosis: An Overview

The modern methods of breast cancer diagnosis are more and more incorporating both the imaging and non-imaging data. These datasets can undergo machine learning (ML) algorithms to identify suspicious areas or abnormalities that can indicate the presence of tumors. These methods formulate clinically significant data at several points in the diagnostic process and have been found beneficial in enhancing the early detection of breast cancer [26].

One of the key diagnostic activities is malignancy classification that will define the presence of benign or malignant abnormalities and will directly affect further clinical management [27]. ML models process image-derived feature to support this process; these include features like shape, texture and intensive patterns. Trained on big and varied data these models are able to determine the probability of malignancy, thus helping clinicians make evidence-based decisions [28]. Moreover, breast cancer is biologically heterogeneous and may be subclassified into molecular subtypes according to given biomarkers, all linked to a different prognostic and therapeutic consequence. Viable subtype classification like triple-negative, HER2-positive and hormone receptor-positive breast cancers are, thus, an essential aspect of current diagnostic and treatment planning interventions [29]. Clinicians can develop more disciplined and selective treatment plans by classifying instances into different subgroups to make them available to the clinicians in a position to incorporate them into their treatment plans. In order to better predict the subtype and have a more optimal treatment plan, machine learning algorithms analyze genetic profiles, patterns of gene expression, and clinical history.

This process of dividing the pictorial image or area into segments or places of interests is called segmentation [26]. The imaging data segmentation is relevant in the diagnosis of breast cancer to identify whether there exists any suspicious lesion or a definition of tumor boundaries [30]. It is an important stage to define the extent of the tumor and its configuration and the basis of other work like the volume of tumor or the data obtained. Their performance is highly improved, which significantly helps in detecting breast cancer early because of the effectiveness of machine learning, particularly deep learning in cutting the breast lesion of the medical image [31].

In addition to this, another major aspect of breast cancer therapy and subsequent follow-up is a capacity to determine metastases formation and cancer recurrence estimation [32]. This is a type of work that entails an estimation of the possibilities of cancer developing again or further spreading in a distant part of the body. Since clinical markers are useful in measuring the risk of either relapse or metastasis, machine learning models, which by definition, demand the use of statistics in their algorithms, can handle multiple forms of data inputs in the form of imaging, genetic data, clinical summation, and electronic health records. Therefore, these predictions can be implemented successfully to tailor the further treatment of patients in an effort to avoid adverse effects and improve patient outcomes.

These activities ought to be combined and modeled together to come up with a better and more dynamic system of breast cancer diagnosis. Specifically, to enhance the accuracy of subtype classification, it is necessary to incorporate complementary processes like tumor segmentation and histological grading to give important structural and morphological data that are critical in accurate cancer characterization [33].

4.2. Explainable Artificial Intelligence (XAI)

The concept of interpretability in machine learning (ML) is the capability to comprehend the way a model acts; how it makes predictions. More importantly perhaps, this type of model did not particularly claim to provide specification as to why it did, but provided an organic view as to how they actually operated. Therefore, it could be assumed that not every interpretable model is completely explainable despite the interpretability being one of the main factors that maximize explainability. By interpretability in its turn, we also imply making forecasts regarding other potential situations and evaluating the modifications of some inputs. Considering both the obvious and the unobvious aspects, the purpose is not only to learn more deeply about the ML model but also to provide a more

detailed description of its activity.

Nonetheless, a number of AI models are not easily interpretable, and thus, the entire picture cannot be viewed behind the result. Based on this lack of transparency, it is easy to argue about their decision-making processes as a result of such reasons. Whereas ask and answer why is the most one can do using a model to determine how a model arrived at a decision based on the inputs, explainability goes one step higher in showing how a model reacts to changes in the inputs and how the output changes as a result of changes in inputs. This distinction is critical particularly in the medical institution where the decisions made are accompanied by the consequences of the lives of people who are the recipients of the given services. The case of a doctor being unable to prescribe their patients medication without knowing how the drugs work is a good example, that the doctor should use algorithms and not know how the recommendations are generated. Practitioners lose the openness to trust or use the AI technology especially when the practitioner is not able to ascertain the results produced by the technology. To be trustworthy and just AI approaches must be clarified and comprehended particularly in life and death fields like medicine, finance and law that have far reaching impacts.

There are two types of explainable AI that are post and intrinsic. The former is what is referred to as transparent models: they are inherently comprehensible and explainable. Conversely, post-hoc explainability techniques are of two types: model-specific approaches and model-agnostic approaches. Post-hoc or surrogate models are used to replicate the decision making of models the working of which would otherwise be incomprehensible. They are unique like the structure of a model and comprise saliency maps and Grad-CAM. Nonetheless, SHAP and LIME and other such model-agnostic explanations could be helpful and can be used to shed light on the prediction mechanisms of many models.

The methods of XAI may be divided into two broad groups, including global and local explanations. Global explainability attempts to describe the general behavior of the model as such that it exposes the general decision-making trends and feature dependencies of the model throughout the dataset. Conversely, local explainability is concerned with explaining why particular predictions have occurred, providing case-specific information on how particular inputs affect a single model output.

The XAI models which have been most actively used alongside ML and deep learning (DL) models in breast cancer studies are discussed in the following sections. The explanation of XAI models is then given, as well as the summary of the relevant studies done in the sphere of breast cancer.

4.2.1. SHapley Additive exPlanations (SHAP)

The concept of interpretability in ML is the capability to comprehend the way a model acts; how it makes predictions. More importantly perhaps, this type of model did not particularly claim to provide specification as to why it did, but provided an organic view as to how they actually operated. Therefore, it could be assumed that not every interpretable model is completely explainable despite the interpretability being one of the main factors that maximize explainability. By interpretability in its turn, we also imply making forecasts regarding other potential situations and evaluating the modifications of some inputs. Considering both the obvious and the unobvious aspects, the purpose is not only to learn more deeply about the ML model but also to provide a more detailed description of its activity.

SHAP [34] is an approach that was proposed by Lundberg and offers a powerful and interpretable framework used to understand the predictions of ML models assigning importance to each single feature in relation to a given output. Local and global interpretability: It allows the quantification of the contribution of each attribute to the overall

prediction that can be used to demystify complex models.

Based on ideas of game theory, namely, Shapley values, SHAP fairly allocates the prediction across features, similar to the way rewards are allocated among participants in a cooperative game.

The SHAP values can be expressed as follows:

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

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 addition of 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′∈ {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.

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

Grad-CAM [35] is a visual explanation device of a broad set of convolutional neural network (CNN) models and was initially introduced by Selvaraju et al. It operates by identifying and highlighting important regions in an input image that is useful in the model class predictions. This is achieved through development of a heatmap of the significant regions by computing the gradients of any target class with respect to the activations of the final convolutional layer. Grad-CAM is more versatile than CAM and is applicable with other CNN architectures with no design modifications being needed. Grad-CAM produces informative visualizations that are computed by calculating gradients and transforming the feature maps of the final convolutional layer by differentiating 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 is then calculated using 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 with the activation maps, and the ReLU function is added to it to create the Grad-Cam heatmap:
- $\text{Grad} - \text{CAM}_c = ReLU \sum_k (a^k A^k)$ (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.

4.2.3. Local Interpretable Model-Agnostic Explanations (LIME)

Ribeiro et al. [36] proposed LIME, an approach for generating locally faithful models and one which 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 nearby 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:

  $x \in R^d$ (9)

  $x' \in \{0,1\}^{d'}$ (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 $g \in G$ 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 $L(f, g, \pi_x)$, where $\pi_x(z)$ specifies the proximity of the instance z to x. LIME balances interpretability and fidelity by optimizing the following equation:

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

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

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.

4.3. 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 focusses 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, 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. Deeper understanding of the molecular and histological features of breast cancer is possible because to 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].

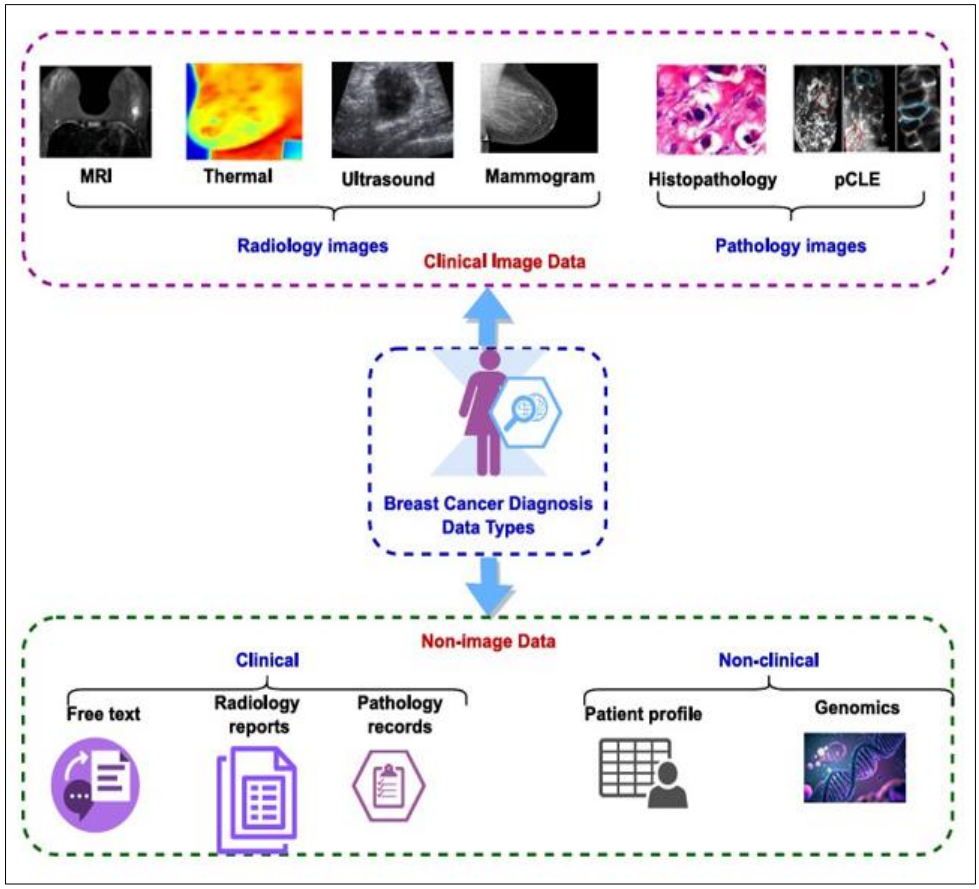

Figure 1. Types of breast cancer diagnosis data

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 prior to 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. In contrast, van Winkel et al. [47] investigated the use of AI as a simultaneous decision support system in DBT. In a multi-reader-multi-case retrospective study, it was shown that AI support significantly improved diagnostic performance, increased the AUC value, and also reduced the reading time of radiologists.

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]. For example, the TCGA-BRCA [42] was selected because of its comprehensive genomic characterization and inclusion of multiple breast cancer subtypes. In contrast, well-structured histopathology image datasets, such as those presented by Spanhol et al. [54] provide complete and standardized imaging data supporting robust model development. By clearly defining the inclusion and exclusion criteria, this study aims to ensure the robustness and applicability of the evaluated multimodal methods based on complete, high-quality datasets representing diverse clinical scenarios.

Table 1. Summary of datasets and methods employed in breast cancer 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]	Histopathological classification	Histopathology
DataBioX histopathology dataset [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]	Multiple 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]	Prediction	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

4.4. 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 requirement and preference of each screening program. In Table 2, this study 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 help decreased 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 van Winkel et al. [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 maintaining or slightly improving 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, Lång et al. [79] found that using AI in this way reduced the number of reads by 44% while maintaining a non-inferior cancer detection rate and comparable recall rates. 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 without compromising sensitivity and recall. 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 non-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 proportion of 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 biological behavior of screen-detected tumors, represent additional challenges. Also, the range of possible applications for any AI system is limited by the unique characteristics and intended purpose of any such system. A classic example is that even

if this is a part of a double reading scenario, an AI system which 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 Work-load
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)	Non-inferior CDR with preserved or reduced recall rate	Screening workload reduced by ~50%
Larsen et al. [76]	AI as a stand-alone reader	Retrospective performance analysis	Screening mammography (DM)	Comparable CDR with reduced 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)	Maintained or improved CDR with reduced recall rate	Screening requirements reduced by ~50%
Lång et al. [79]	AI as a triage tool	Randomized controlled experiment	Screening mammography (DM)	Non-inferior CDR with comparable 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)	Non-inferior 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)	Maintained cancer detection in simulation-based analysis	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 analyzed literature also demonstrates that artificial intelligence has significant potential to improve the process of screening and diagnostic tests of breast cancer, but also shows significant limitations that prevent its use in the clinic on a regular basis. AI systems based on imaging and applied to mammography, ultrasound, MRI, and histopathology have been shown to be highly sensitive and display higher cancer detection rates especially when utilized to support concurrent decision making or triage, although these advantages are often accompanied by increased false-positive/recall rates, which has concerns about overdiagnosis and unnecessary follow-up. Most of the reported performance improvements are based on retrospective or enriched datasets that do not necessarily reflect the actual screening population in the real-world situation, which adds to dataset bias and further generalizability. Multi-modal AI methods which combine imaging with clinical or molecular data are better at subtype classification and have better prognostic performance, but their clinical implementation is limited by data scarcity, lack of interoperability, and multi-source data integration complexity. Explainable AI algorithms such as SHAP, Grad-CAM and LIME are significant in enhancing transparency and clinician confidence, and their post-hoc disposition and sensitivity to model and data fluctuations can lead to unreliable and possibly misguided explanations. Moreover, the lack of external and prospective validation, regulatory and implementation obstacles and imbalances in training datasets are still limiting to real world deployment. These issues can be resolved by standardized evaluation systems, varied and quality data, effective explainability policies, and future multi-centre clinical trials that will validate that AI systems offer reliable, fair, and clinically significant advantages to breast cancer diagnosis.

The use of artificial intelligence in breast cancer screening and diagnosis has made considerable advances and has shown the capacity to enhance the quality of the diagnostic process, decrease the inter-observer variability, and simplify the clinical processes of radiology and pathology. AI systems can be conveniently used as co-decision-support systems, independent readers, or triage systems, which means that they can be flexibly implemented, depending on the needs of screening programs and their available resources as outlined in this review. All these applications highlight the possibility of AI as the efficiency booster and the ability to provide a more personalized management of patients.

Nevertheless, clinical translation cannot be done on a large scale unless a number of crucial issues are resolved. Most of the available evidence is based on retrospective studies and single-centre studies which may not reflect the heterogeneity of real-life screening populations, imaging protocols and clinical practice. Subsequent studies ought then to focus on the more important studies that are of large scale, prospective clinical trials, and federated and multi-center datasets so that their impact becomes more robust and representative of other demographic groups. Standardized evaluation metrics and reporting systems are needed to facilitate meaningful comparison among AI systems and also to facilitate regulatory approval. Along with the explanation of AI and clinician-in-the-loop, they should be integrated to enhance transparency, trust, and clinical accountability. These methodological, technical, and regulatory gaps are the key to allowing AI to play a safe, fair, and sustainable role of an assistant, but not a substitute, of radiologists and pathologists in the routine management of breast cancer.

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.