Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-20011-6

Deep learning approaches to detect breast cancer: a comprehensive review

Amir Mohammad Sharafaddini[1] · Kiana Kouhpah Esfahani[1] · Najme Mansouri[1]

Received: 9 April 2024 / Revised: 2 July 2024 / Accepted: 3 August 2024

Abstract
Detection and diagnosis of breast cancer have greatly benefited from advances in deep learning, addressing the critical problem of early detection and accurate diagnosis. This paper presents a review of 68 high-quality articles related to deep learning techniques applied to various imaging modalities including mammography, ultrasound, MRI, histopathology, and thermography published in 2022 and 2023. Additionally, this paper discusses and compares the deep learning approaches to detect breast cancer based on the dataset, input size of the model, model architecture, the proposed approach, targeted problem, and output performance. There is a primary concern about the variability in imaging data, which can lead to inconsistency in diagnosis. In this study, deep learning, particularly Convolutional Neural Networks (CNNs), was leveraged to enhance image accuracy and consistency across multiple imaging modes. CNNs have demonstrated enhanced sensitivity and specificity in mammography imaging, for example, by detecting microcalcifications and masses. According to studies based on MRIs, deep learning models are able to distinguish between different tissue types, aiding in the precise localization of tumors. Thermography is less common, but deep learning models can detect abnormal thermal patterns associated with malignancies. In addition, this paper addresses the issue of limited and imbalanced datasets, which often hamper deep learning models' performance. Data augmentation and transfer learning are explored as solutions to improve model robustness and generalizability. This paper provides a comprehensive review of breast cancer detection and diagnosis datasets, highlighting their significance and unique characteristics. Researchers can select appropriate resources for their diagnostic studies and model development by analyzing these datasets thoroughly. Even so, larger datasets and improved model interpretability remain challenges. We propose future research directions to address these issues, emphasizing multi-modal data integration and advanced algorithmic development.

Keywords Deep learning · Breast cancer · Convolution neural network · Transfer learning · Data augmentations

✉ Najme Mansouri
najme.mansouri@gmail.com

[1] Department of Computer Science, Shahid Bahonar University of Kerman, Kerman, Iran

Published online: 20 August 2024

Abbreviations

MRI	Magnetic Resonance Imaging
DLR	Deep Learning Radiomics
HER2	Human Epidermal Growth Factor Receptor 2
DFS	Disease-Free Survival
HR	Hazard Ratio
LDCT	Low-Dose Computed Tomography
ITR	Intra-Tumoral Region
PTR	Peri-Tumoral Region
AUROC	Area Under the Receiver Operating Characteristic
ALNM	Axillary Lymph Node Metastasis
SEN	Sensitivity
SPE	Specificity
ACC	Accuracy
PPV	Positive Predictive Value
NPV	Negative Predictive Value
SGD	Stochastic Gradient Descent
pCR	Pathologic Complete Response
NAST	Neoadjuvant Systemic Therapy
PEI	Pathological Enhancement Imaging
DT-KNN	Decision Tree-K-Nearest Neighbors
RF-KNN	Random Forest-K-Nearest Neighbors
MSB	Magnification-Specific Binary
MIB	Magnification-Independent Binary
MTRRE-Net	Multi-scale Two-fold Residual Recurrent Network
THG	Third Harmonic Generation
RANN-BCC	ResNet-based Attention Neural Network for Breast Cancer Classification
IDC	Invasive Ductal Carcinoma
MIL	Multiple Instance Learning
MTL	Multi-Task Learning
RMSE	Root Mean Square Error
CD	Chromogenic Detection
CDACS	CD with Automated Cell Segmentation
DCNN	Deep Convolutional Neural Network
TCGA	The Cancer Genome Atlas
BCSS	Breast Cancer-Specific Survival
ER	Estrogen Receptor
ERBB2	Receptor Tyrosine Kinase
BFMI	Bright Field Microscopic Imaging
AFMI	Autofluorescence Microscopic Imaging
OPMI	Optical Phase Microscopic Imaging
PLF	Patch-Level Fusion
DLF	Decision-Level Fusion
DMR	Database for Mastology Research
CADx	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
FPR	False Positive Rate

CBIS-DDSM	Curated Breast Imaging Subset of Digital Database for Screening Mammography
MIAS	Mammographic Image Analysis Society
IMPA	Improved Marine Predators Algorithm
OBL	Opposition-Based Learning
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
DDSM	Digital Database for Screening Mammography
ShCNN	Shepherd Convolutional Neural Network
IoT	Internet of Things
FACS	Feedback Artificial Crow Search
CSA	Crow Search Algorithm
FAT	Feedback Artificial Tree
TPR	True Positive Rate
OMLTS-DLCN	Optimal Multi-Level Thresholding-based Segmentation with DL enabled Capsule Network
OKMT-SGO	Optimal Kapur's based Multilevel Thresholding with Shell Game Optimization
BPNN	Back Propagation Neural Network
DL	Deep Learning
VGG	Visual Geometry Group
PAA	Patch Aggregation Architecture
Faster R-CNN	Faster Region-based Convolutional Neural Network
mAP	Mean Average Precision
JI	Jaccard Index
HRDO	Hybrid Dragon-Rider Optimization
FL	Federated Learning
E-RNN	Elman Recurrent Neural Network
FDR	False Discovery Rate
MCC	Matthews Correlation Coefficient
SVM	Support Vector Machine
DMD-CNN	Digital Mammogram Diagnostic Convolutional Neural Network
FPGA	Field-Programmable Gate Array
PCA	Principal Component Analysis
MLP	Multi-Layer Perceptron
MDS	Multidimensional Scaling
UMAP	Uniform Manifold Approximation and Projection
WDBC	Wisconsin Diagnostic Breast Cancer
LRP-NET	Longitudinal Risk Prediction Network
BUSI	Breast Ultrasound Image Dataset
ABUS	Automated Breast Ultrasound
DLN	Deep Learning Network
ASN	Automatic Segmentation Network
AP	Average Precision
NAC	Neoadjuvant Chemotherapy
TL	Transfer Learning
GA	Genetic Algorithm
YOLOv3-tiny	You Only Look Once version 3, tiny version
FPC	False Positive Count

QUS	Quantitative Ultrasound
TNBC	Triple Negative Breast Cancer
CNN-LSTM	Convolutional Neural Network-Long Short-Term Memory
BTEC-Net	Breast Tumor Ensemble Classification Network
RFS-UNet	Residual Feature Selection UNet (Ultrasound image)
BUSI	Breast Ultrasound Images
VGGNet	Visual Geometry Group Network
UCA	US-guided Co-Attention
US	Ultrasound
Resnet18	Residual Network 18
DLRP	Deep Learning Radiomic Pathway
CSOA-wKNN	Crow Search Optimization Algorithm with weighted K-Nearest Neighbors
FNR	False Negative Rate
CER	Classification Error Rate
AI	Artificial Intelligence
CAD	Computer-Aided Diagnostic
BC	Breast Cancer
ML	Machine Learning
CNNs	Convolutional Neural Networks
IARC	International Agency for Research on Cancer
CT	Computed Tomography
HI	Histopathology Images
WSIs	Whole Slide Images
GSA	Gravitational Search Algorithm
HHO	Harris Hawks Optimization
WOA	Whale Optimization Algorithm
SVM	Support Vector Machines
ReLU	Rectified Linear Unit
NLP	Natural Language Processing
CSOA	Crow-Search Optimization Algorithm
DFOA	Dragon-Fly Optimization Algorithm
PSO	Particle Swarm Optimization

1 Introduction

There are several diseases that are characterized by uncontrolled division and progression of sporadic cells. Through the lymphatic system and circulation framework, these bizarre cells can attack neighboring tissues and spread to other parts of the body. Medical image classification has greatly influenced diagnostic techniques and therapeutic interventions [1]. Traditional diagnostic procedures, such as colonoscopy, are difficult to perform accurately and take large amounts of time [1]. Diseases, like cancer, are also characterized by uncontrolled division and advancement of sporadic cells in the body. It is possible for these bizarre cells to attack adjoining tissues and spread to other parts of the body through the lymphatic system or circulation framework. It is the abnormal cell division that causes a mass to develop called a tumor. There are many factors that may cause cancer, including innate changes, common components, lifestyle choices, and certain diseases. Cancerous tumors (malignant) and benign tumors (benign) can both occur. Increasing incidences of cancer require precise and effective diagnostic tools to help medical experts make early diagnoses and treatment recommendations [2]. For example, the most common types of skin cancer caused by UV radiation are melanoma and nonmelanoma. Globally, skin cancer, especially melanoma, is on the rise. Skin cancer must be detected early for effective treatment and to improve patient outcomes. For the detection and classification of various cancers, including breast cancer, advanced machine learning techniques like deep learning have been successfully applied in medical imaging [3]. As a result of the use of deep learning models like the InceptionV3 coupled with support vector classifiers, timely treatment recommendations are provided to patients with various types of cancer [2]. In the same way, early diagnosis is crucial for effective treatment of breast cancer. Machine learning and artificial intelligence are enhancing breast cancer detection accuracy, increasing survival rates and improving outcomes. There are several types of cancer, including:

- Breast Cancer: The disease primarily affects women, but can also affect men's breast tissue.
- Lung Cancer: Cancer of the lungs caused by smoking or exposure to carcinogens.
- Colorectal Cancer: The disease usually occurs in the colon or rectum, and it is usually detectable early through screening.
- Prostate Cancer: Male prostate cancer is one of the most common cancers.
- Skin Cancer: Melanoma and nonmelanoma are the most common skin cancers caused by UV radiation.
- Leukemia: Bone marrow cancer and blood-forming cell cancer.
- Lymphoma: An immune system cancer that starts in the lymphatic system.
- Pancreatic Cancer: Diagnosed at an advanced stage when it develops in the pancreas.
- Ovarian Cancer: It affects only the ovaries and is often diagnosed only at an advanced stage.
- Brain Cancer: Brain tumors can develop in the brain itself or in the surrounding structures, resulting in a wide range of symptoms.

Global Burden of Disease Cancer Collaboration and the IARC (International Agency for Research on Cancer) have projected that cancer cases rose 28 percent between 2006 and 2016. In 2030, 2.7 million new cases of cancer are expected [4]. Figure 1 shows the projections for new cancer cases in the United States based on type [5].

In the United States, breast cancer ranks as the second most prevalent form of cancer. Breast Cancer (BC) is the malignant tumor that develops from breast cells. In most cases, BC begins in the milk producing ducts, or lobules, which drain milk to the nipple. When BC invades healthy tissues near the underarm lymph nodes, they have a path to other parts of the body. Cancer cells have spread beyond the original tumor at BC stage [6].

The importance of breast cancer lies in its impact on women's wellbeing and its questioning of their commitments. Cancer-related deaths are still a major cause of death, emphasizing the need for further study and improved treatments. The breast cancer mindfulness movement has set a standard for other cancers to follow. The importance of back communities, personalized care approaches, and treatment developments in healthcare are highlighted by its implementation. Its pertinence is highlighted by its financial burden, early discovery strategies, and advancements in behavioral mindfulness. The prevalence, health impacts, and role in forming cancer care make breast cancer a vital area of therapeutic research, open health activities, and constant support. Detecting cancer early, including breast cancer, has a profound impact on the delivery of healthcare and the understanding

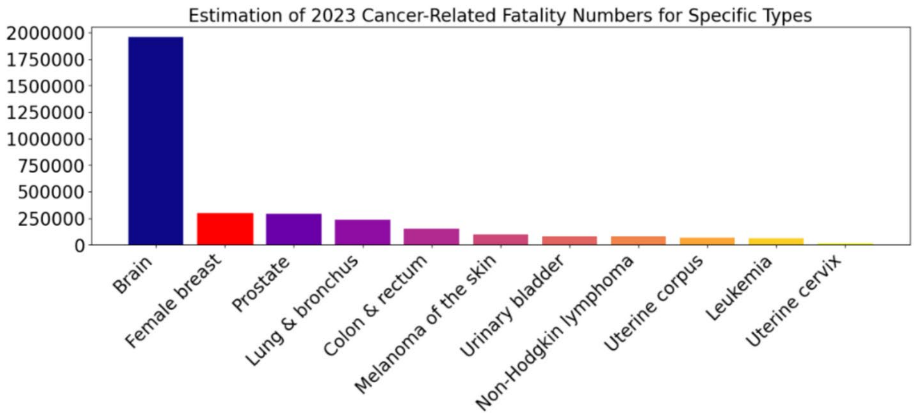


Fig. 1 Common cancer types projected to cause death in 2023

of results. Cancer detection at its earliest stages increases treatment success rates, reduces mortality, and reduces the need for forceful intervention. In early-stage breast cancer, five-year survival rates are around 99%. The use of less forceful medicines such as surgery, chemotherapy, and radiation led to quicker recuperation times and better corrections.

Artificial Intelligence (AI) has risen as a progressive drive within the field of medical diagnostics, outstandingly within the early location of cancers. Due to their tremendous potential to improve precision, proficiency, and unwavering quality, Machine Learning (ML) and Deep Learning (DL) have received much attention within this transformative scene. Deep Learning has made particularly significant strides in cancer discovery, especially in breast cancer. Machine Learning, which precedes Deep Learning, helps computers make forecasts based on information designs and enhance cancer location. Its algorithmic flexibility has enabled the investigation of broad datasets, leading to the identification of unpretentious characteristics characteristic of cancerous growths. Nevertheless, Machine Learning depends heavily on manual fine-tuning to achieve optimal results due to highlights' complexity and limited flexibility. Deep Learning uses neural networks with multiple layers to extract complex highlights from crude information without manual intervention. As a result, Deep Learning is capable of observing subtleties that are difficult to detect with conventional techniques. Deep Learning has since overcome Machine Learning's limitations, thereby improving cancer location specificity and affectability.

The powerful potential of Deep Learning is vividly demonstrated within the context of breast cancer. Detecting potential malignancies in mammograms can be difficult due to the complicated surfaces and designs. With its various levels of design, Deep Learning is uniquely capable of unraveling these complexities, perceiving small abnormalities that might be missed by humans. Because it can memorize enormous datasets, it can observe designs exhibiting breast cancer with astounding precision. Using Deep Learning models like Convolutional Neural Systems (CNNs), breast cancer location has exceeded past benchmarks in an outstanding manner. Using mammogram images, these models have shown an unmatched ability to identify potential cancerous injuries in the early stages of their development. As a result, they have an increased potential for early mediation and a move forward in understanding results.

The most outstanding manifestation of Deep Learning can be found in cancer location, which is based on Machine Learning. A smart technology called Deep Learning is changing how we identify breast cancer. Using computers, it can detect important features and patterns really accurately. It is possible for technology and medicine to change healthcare when they come together. It's amazing how deep learning is helping spot breast cancer.

1.1 Motivation

One of the most common and deadly cancers among women, breast cancer still poses a significant threat to millions of lives every year. Diagnostic imaging and therapeutic approaches have advanced in recent decades, but early and accurate diagnosis continues to pose a challenge. For successful treatment and improved patient survival rates, early detection is crucial. However, radiologists are not infallible when interpreting mammograms because of fatigue, differences in experience, and the subjective nature of visual assessment. Medical imaging analysis has been transformed by artificial intelligence (AI), particularly deep learning, in recent decades. Large datasets can be analyzed using deep learning algorithms, which mimic neural networks in the brain. As a result, they are particularly suitable for tasks that involve high-dimensional and intricate data, such as detecting subtle anomalies in medical images that indicate cancer. In addition to augmenting radiologists' diagnostic capabilities, deep learning in medical imaging can also improve breast cancer detection accuracy, efficiency, and consistency.

Deep learning methodologies play an important role in the detection of breast cancer in this comprehensive study. Over the past few years, deep learning has rapidly advanced and been applied to a variety of medical fields, with breast cancer research receiving particular attention. Data-driven deep learning promises to reduce human error and provide robust, reproducible breast cancer diagnostics. A detailed review of deep learning advancements in breast cancer detection is presented in this article. The aim of this review is to provide researchers, clinicians, and healthcare stakeholders with an understanding of deep learning's role in enhancing breast cancer diagnostics. We will explore various deep learning architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, discussing their strengths and limitations. In addition, we will discuss methods and techniques employed in preprocessing, training, and validating these models, emphasizing their contributions to improved accuracy and reliability.

This study will also address potential research directions for deep learning in breast cancer detection. The foundation of a trustworthiness and interpretability of AI-driven diagnostic systems will be enhanced by emerging trends like transfer learning, data integration across multiple modes, and explainable AI. We will also discuss the ethical and practical challenges of implementing these advanced AI solutions in clinical settings, including data privacy, algorithmic bias, and regulatory frameworks. It is crucial to analyze the transformative potential of deep learning in early breast cancer detection as artificial intelligence continues to advance and integrate into healthcare. In order to overcome this devastating disease, deep learning has the potential to improve early detection and save many lives.

1.2 Contribution

In this section, we depict the unmistakable commitments of this overview article to the advancing scene of breast cancer location and conclusion. By synthesizing an assorted cluster of investigate discoveries and experiences, we point to supply a comprehensive asset that propels our understanding of the essential role deep learning plays within the domain of restorative imaging and healthcare. Our commitments include a multifaceted investigation of breast cancer imaging modalities, a curated compilation of datasets, an in-depth investigation of neural arrange approaches, and a forward-looking viewpoint on the clinical suggestions of our discoveries. It aims to enhance the understanding of available datasets, facilitate better experimental design, and enable better deep learning models to be developed. These datasets highlight critical variations and challenges, such as imbalance between classes, data quality variability, and missing data. The main contributions of this paper are as follows:

- We offer up-to-date factual information on breast cancer types and mortality rates within the year 2023. By displaying Fig. 1, we contextualize the criticalness of making strides breast cancer location strategies and emphasize the pertinence of our study article in contributing to this imperative range of healthcare.
- In this survey article, we provide a comprehensive synthesis of various imaging modalities employed in the detection of breast cancer. We meticulously analyze each imaging technique, including Mammogram, CT scan, Ultrasound, MRI, and Histopathological imaging, elucidating their strengths and limitations. By consolidating a wide range of research findings, we offer a holistic view of how these modalities contribute to the early detection and diagnosis of breast cancer.
- A wide variety of breast cancer imaging techniques is included in our extensive collection of datasets. This compilation serves as a valuable resource for researchers and practitioners seeking benchmark datasets for evaluating and training deep learning models. By meticulously discussing the characteristics, sources, and annotations of these datasets, we facilitate an informed selection process for researchers embarking on breast cancer detection projects.
- This survey extensively explores the myriad ways neural networks have been leveraged to enhance breast cancer detection accuracy. We delve into the specifics of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and other deep learning architectures tailored for medical imaging analysis. By synthesizing research outcomes from various studies, we illustrate how these techniques harness the power of artificial intelligence to contribute significantly to accurate breast cancer diagnosis.
- It is important to emphasize the variety of augmentation strategies employed by different studies. Identify specific variations or combinations of techniques that have been found effective for diagnosing breast cancer. It could include discussing unique augmentation pipelines or incorporating domain expertise into augmentation.
- An analysis of recent breast cancer detection research yields new insights in this article. We discuss emerging trends, unresolved challenges, and potential avenues for future exploration by critically analyzing various studies. Readers gain a deeper understanding of breast cancer diagnosis through our unique perspective.

1.3 Paper structure

In order to provide a systematic exploration of the profound impact of deep learning on breast cancer detection, this survey article is organized into a structured framework. Each section is meticulously designed to guide readers through a comprehensive journey, depicted in Fig. 2, from understanding the fundamental principles of breast cancer imaging modalities to delving into the intricate landscape of deep learning techniques and their application.

Fig. 2 An illustrative roadmap for navigating deep learning in breast cancer diagnosis

2 Review methodology

This section outlines a structured methodology used to systematically review the existing literature on the use of deep learning techniques in detecting breast cancer. The following sections provide a detailed, step-by-step description of the methodology used in conducting this systematic review. The process emphasizes transparency, rigor, and compliance with established guidelines during the synthesis and analysis of evidence.

2.1 Planning the review

Creating research questions is the initial step in establishing the evaluation criteria. These meticulously crafted questions were used to conduct further searches across various data sources. The review method collects and identifies pertinent data for the proposed investigation. Articles are either considered or heavily discarded based on the evaluation process. The task selection by a single researcher could potentially introduce bias into the study. Therefore, this Systematic Literature Review was carried out by dividing the work among all the contributors of this paper. Each author drafted a document detailing their insights on the review process and shared it with the rest of the team. This cycle was repeated over a fixed period. Numerous online databases were thoroughly searched. Figure 3 illustrates the evaluation process.

2.2 Research questions

The systematic review that focuses on detecting breast cancer using deep learning techniques aims to address key research questions and objectives, striving for a comprehensive evaluation of the wide range of studies in this area. Its main goal is to thoroughly examine

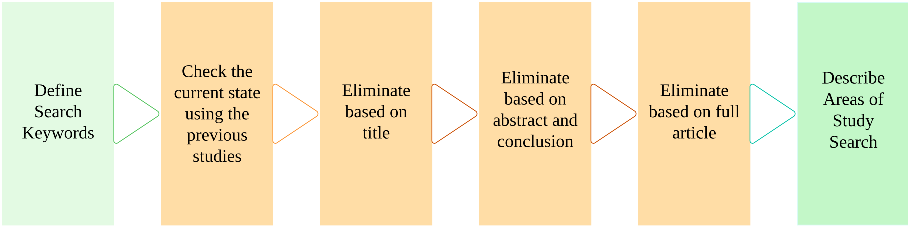


Fig. 3 Process of review methodology

the methodologies, advancements, and results in the field of using deep learning for breast cancer detection. Through a detailed analysis, this review aims to uncover crucial insights that could potentially enhance and improve the effectiveness of deep learning models in accurately detecting and diagnosing breast cancer, thereby making a significant contribution to the progress of medical diagnostics in this vital area. The research questions and objectives guiding the focus of this review are shown in Table 1.

2.3 Sources of information

To conduct a thorough review of the literature, it's crucial to perform an extensive search across various electronic sources. To increase the chances of finding relevant research publications, we have compiled a specific set of data sources as follows:

- Springer (www.springerlink.com)
- ACM Digital Library (www.acm.org/dl)
- IEEE Xplore (ieeexplore.ieee.org)
- ScienceDirect (www.sciencedirect.com)
- Taylor & Francis (www.taylorandfrancis.com)

2.4 Search criteria

Research studies published in 2022 and 2023 were examined to: (1) evaluate the use of various imaging modalities, (2) compare breast cancer (BC) imaging modalities, and (3) identify the most frequently cited and publicly available BC databases with different types of BC. The search criteria generally used in this research study consisted of the terms "Breast Cancer" AND "Deep Learning". We've selected papers that specifically focus on the intersection of "Breast Cancer" and "Deep Learning", choosing those where both keywords appear in their titles. This strict criterion was used to streamline the selection, given the vast number of papers that include these keywords in a broader context. Our goal is to present a curated collection that deeply explores the nuanced relationship between Breast Cancer and advancements in Deep Learning, offering a concentrated and insightful examination of this critical intersection.

2.5 Inclusion and exclusion criteria

The inclusion and exclusion criteria were set up to ensure the relevance and consistency in the selection of research studies for this investigation.

Table 1 Research questions, motivation, category, and mapping sections

No	Research Question	Motivation	Category	Mapping section
RQ1	How does Deep Learning compare with traditional Machine Learning in terms of precision, efficiency, and reliability in early breast cancer detection?	Analyze the potential of Deep Learning for detecting breast cancer with greater precision, efficiency, and reliability than traditional Machine Learning	Deep Learning Impact on Breast Cancer Detection	1. Introduction
RQ2	What advancements have Deep Learning made in the area of breast cancer detection in terms of precision, efficiency, and consistency?	Investigate the potential of Deep Learning to boost precision, efficiency, and consistency in breast cancer detection compared to traditional methods	Impact of Deep Learning in Breast Cancer Research	1.1. Motivation
RQ3	Using Springer, ACM Digital Library, IEEE Xplore, ScienceDirect, and Taylor & Francis as electronic sources, what is the contribution of this review to comprehensiveness and quality?	Analyze the impact and reliability of selected electronic sources in enhancing the comprehensiveness of the literature review	Electronic Sources in Literature Review	2.3. Sources of Information
RQ4	What are the strengths and weaknesses of different imaging modalities, such as mammography, ultrasound, MRI, histopathology, and thermography, in assisting in the accurate and early diagnosis of breast cancer?	Investigate the roles, strengths, and limitations of diverse breast cancer imaging modalities, emphasizing their collective contribution to accurate and timely detection of breast cancer	Breast Cancer Imaging Modalities	3.1. Different Imaging Approaches for Breast Cancer Detection
RQ5	Is there a solution to the limitations of qualitative image assessment for detecting breast cancer using advanced image analysis technologies, including machine learning algorithms, and what challenges and considerations must be addressed when these technologies are adopted in medical practice?	Consider the potential and challenges of advanced image analysis technologies, particularly machine learning algorithms, in improving consistency and accuracy in breast cancer detection	Advanced Image Analysis Technologies in Breast Cancer Detection	3.1.5. Thermography

Table 1 (continued)

No	Research Question	Motivation	Category	Mapping section
RQ6	What are the benefits and concerns of Computer-Assisted Diagnostic (CAD) systems for early detection of breast cancer?	Analyze the benefits and concerns raised in studies on AI-assisted CAD systems' impact on breast cancer early detection, and examine the impact of Moore's law on CAD development	Computer-Aided Diagnostic (CAD) Systems in Breast Cancer Detection	3.2. Computer-Aided Diagnostic (CAD)
RQ7	What role do Deep Learning (DL) and Machine Learning (ML) play in broader fields of artificial intelligence (AI)?	Explore the impact of technological epochs, including the transition to deep learning, and analyze the interrelationships and roles of Deep Learning (DL) and Machine Learning (ML) within artificial intelligence (AI)	Impact of Deep Learning on Technological Landscape	3.3 Deep Learning
RQ8	How does Convolutional Neural Network (CNN) leverage biomimetic principles and computational innovation to revolutionize computer vision, particularly in Medical Image Analysis (MIA) and the classification of breast cancer tumors, and which key architectural elements and activations are used by CNNs to accomplish this?	Analyze the architectural elements and activations employed in CNNs for enhancing performance as well as how they revolutionize computer vision, focusing on the use of CNNs in Medical Image Analysis (MIA) and breast cancer tumor classification	Impact of CNNs on Computer Vision	3.4. Convolutional Neural Networks (CNNs)

Table 1 (continued)

No	Research Question	Motivation	Category	Mapping section
RQ9	What is the role of transfer learning, especially when using pre-trained models, in improving the efficiency and effectiveness of machine learning models in scientific data analysis, and what are the key paradigms and applications of transfer learning, including instance-based, feature-based, and model-based transfers?	Study the role of transfer learning in scientific data analysis, with an emphasis on its ability to increase the efficiency and effectiveness of machine learning models. Transfer learning is explored in terms of its key paradigms and applications, including features, instances, and models	Impact of Transfer Learning in Scientific Data Analysis	3.5. Transfer Learning
RQ10	In what ways does data augmentation contribute to improving machine learning models, especially in the context of breast cancer detection, and how are key techniques involved in data augmentation, such as adding noise, cropping, flipping, scaling, brightening, rotating, translating, adjusting contrast, saturation, and enhancing colors, performed?	Analyze the role of data augmentation in improving machine learning models, focused on the detection of breast cancer. Examine various techniques for enhancing data, such as adding noise, cropping, flipping, scaling, brightness adjustment, rotation, translation, contrast adjustment, saturation, and color enhancement	Enhancing Machine Learning Models through Data Augmentation	3.6. Data Augmentation
RQ11	In terms of publications in breast cancer detection using deep learning, how has the landscape changed between 2022 and 2023 across leading academic publishers such as Springer, ScienceDirect, IEEE, Taylor & Francis, and ACM?	It is crucial to assess the dissemination of knowledge and identify potential shifts or focal points in breast cancer detection research by examining publication trends among leading academic publishers	Evolution of Publication Landscape in Breast Cancer Detection	7.1. Analyze based on the year and publisher

Table 1 (continued)

No	Research Question	Motivation	Category	Mapping section
RQ12	What challenges and opportunities does the integration of histopathological and radiological data present for enhancing diagnostic accuracy through deep learning models of breast cancer?	Study the potential benefits of combining histopathological and radiological data in deep learning models to improve breast cancer diagnosis accuracy	Integration of Histopathological and Radiological Data in Deep Learning Models	8. Open issues and Future research directions
RQ13	What are the most prevalent and variable characteristics of the primary breast cancer datasets that are used in deep learning research?	An accurate deep learning model depends heavily on the quality of the dataset. In order to ensure clean data, well-annotated data, and representative data, it is vital to understand the dataset	Data Quality and its Impact on Model Accuracy and Performance	6. Deep learning-based approaches to detect breast cancer

Inclusion criteria are:

- Publication Period: Only studies published in 2022 and 2023 were included to capture the most recent advancements and findings.
- Relevance to Breast Cancer: Studies that are directly related to breast cancer diagnostics, imaging modalities, and the application of machine learning and deep learning techniques in this context were included.
- Language: Only studies available in English were considered to ensure consistency in understanding and analysis.

Exclusion criteria are:

- Publication Date: Studies published before 2022 or after 2023 were excluded to concentrate on the latest developments in the field.
- Irrelevance: Studies that did not primarily focus on breast cancer diagnostics, imaging modalities, or the application of machine learning within this domain were excluded.
- Non-English Language: Studies not available in English were excluded to avoid potential language barriers in analysis and interpretation.

These criteria were rigorously applied during the selection process to ensure the validity and suitability of the studies included in this research.

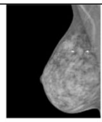
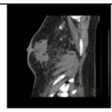
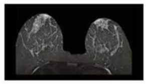
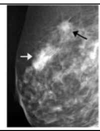
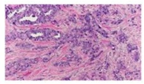
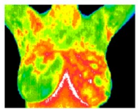
3 Background

Breast cancer is important because it affects women's wellbeing and questions their commitments. Despite advances in treatment and research, cancer-related deaths remain a leading cause of death. It is the breast cancer mindfulness movement that has set the standard for other cancers to follow. Medical image analysis models can be trained and retrained according to changes in image distribution using adaptive self-learning techniques [7]. Additionally, early detection of breast cancer using mammogram images reduces women's mortality rates and allows them to receive proper treatment, with recent advancements enhancing their effectiveness [8].

Women are mostly affected by this disease, but it can affect men's breast tissues as well. Breast cancer is one of the leading causes of death in women with an estimated annual incidence and mortality rate of over 2.1 million [9]. Screening methods like mammography increase the chance of successful treatment when detected early [9].

With Deep Learning, complex highlights can be extracted without manual intervention from crude information by using neural networks with multiple layers. As a result, Deep Learning is able to detect subtleties that conventional methods are unable to detect. In medical imaging, deep learning approaches are being used to improve detection and diagnosis of diseases such as rheumatoid arthritis using X-rays and magnetic resonance imaging [10]. By overcoming Machine Learning's limitations, Deep Learning has improved cancer location specificity and affectability. Among recent advancements in breast cancer classification frameworks, deep learning models have been incorporated for feature extraction and optimization, which significantly improves performance over traditional approaches [11].

Table 2. Breast tissue medical images

Screening Technique	Spatial Resolution	Advantages	Disadvantages	Image
Mammogram	Lower Spatial Resolution	Time efficient, requires lower average radiation dosage	Lower spatial resolution, uses X-rays, costly	
CT Scan	High Spatial Resolution	Precise, High spatial resolution	Uses X-rays, can cause allergic reactions, potentially costly	
MRI	High Spatial Resolution	No ionizing radiation, high sensitivity, good tissue differentiation	Costly, invasive procedure, low specificity	
Ultrasound	Moderate Spatial Resolution	Less expensive, uses sound waves, non-invasive	Low specificity, not able to detect all types of tumors	
Histopathological And Microscopic Imaging	Very High Spatial Resolution	Enables detailed cellular level analysis, crucial for definitive diagnosis	Requires tissue sample, invasive, time-consuming process	
Thermography	Low Spatial Resolution	Non-invasive, does not involve exposure to radiation	Limited in scope, it can only alert to changes that may need further investigation	

3.1 Different imaging approaches for breast *cancer* detection

It is vital to diagnose BC accurately and precisely in order to improve survivorship [12]. In breast cancer diagnosis, different imaging modalities play a different role, each with its own significance. There are several imaging modalities available, including mammography, ultrasound, histopathology, MRIs (Magnetic Resonance Imaging), and thermography. Table 2 illustrates these approaches in an elegant manner. These diverse approaches are more effective at understanding and diagnosing breast cancer.

3.1.1 Mammography

Mammography is the most common and time-tested method used to detect breast cancer, especially in women with no symptoms. It is the gold standard for identifying breast cancer in its earliest and most treatable stages through the use of this sophisticated imaging technology. X-ray images of the breast are a key component of this diagnostic prowess. The use of these images allows for a comprehensive examination and screening of breast tissue, employing low-energy X-rays [13]. Mammography harnesses the prowess of modern medical imaging to detect breast cancer even before outward symptoms manifest. In addition to improving patient outcomes, early detection significantly increases the likelihood of successful treatment options. A randomized controlled trial involving the broader population has firmly established mammography's effectiveness in saving lives. In these trials, mammography screening has been shown to reduce the incidence of breast cancer mortality [14]. Over time, mammography has evolved to offer even better diagnostic capabilities. In such an advancement, a three-dimensional imaging approach has been introduced known as 'Digital Tomosynthesis Mammography' that enables breast examinations to be more precise. Another pioneering technique, 'Contrast-Enhanced Digital Mammography' (CEDM), is also emerging. Through the intravenous infusion of an iodinated contrast agent along with traditional mammography procedures, CEDM increases the visibility of potential abnormalities [15]. Mammography plays an integral role in the battle against breast cancer due to these innovative methodologies.

3.1.2 Ultrasound

Ultrasound imaging is distinguished from various alternatives by its monochromatic display and lower resolution. The grayscale presentation of ultrasound is far from a drawback, despite its lack of color palette. More vibrant images might mask intricate details. Malignancies often show consistent patterns on ultrasound images. In these images, the visual narrative is often asymmetrical due to their irregular shapes. This area is closely examined by healthcare professionals since these irregularities may indicate malignancy. As a result of these irregularities, diagnosis of such illnesses is also challenging. The visual clarity of ultrasound images is often blurred, which differs from the sharp definition of high-resolution imaging techniques. Fuzziness may seem limiting, but it serves as a reminder of how complex medical imaging can be. These blurry images reflect the elusive nature of underlying pathologies in cases of malignancy. Additionally, vague margins in ultrasound images of malignant regions accentuate their intricate nature. They blend into surrounding tissue instead of having well-defined edges, creating uncertainty during the diagnostic process. Due to its sensitivity, ultrasound is able to capture the intricate interaction between malignant tissue and its surrounding environment. In contrast to mammography, ultrasound is excellent at distinguishing cysts from solid masses [16]. As opposed to more vibrant imaging techniques, ultrasound provides an in-depth view into the intricacies of the human body despite its monochromatic, low-resolution nature. An ultrasound image of a malignancy reveals irregular shapes, blurred contours, and indistinct margins, highlighting the inherent challenges and nuances of accurate diagnosis. Ultrasound images are intrinsically speckled, and this impacts their quality [17].

3.1.3 MRI

The sensitivity of Magnetic Resonance Imaging (MRI) is unparalleled among breast imaging methodologies. It reveals intricate details about breast lesions, including their shape, dimensions, and spatial orientation. The unmatched versatility and accuracy of MRI can be attributed to its ability to support multi-planar imaging and facilitate three-dimensional reconstruction [10]. However, MRI does come with some trade-offs despite its advantages. Although MRI provides a comprehensive picture of breast lesions, it is resource-intensive. There is a significant time commitment and costs associated with the procedure, potentially limiting its widespread use.

With mammography as its complement, MRI is a key component of breast cancer (BC) screening. As a result of its unprecedented sensitivity, it can detect subtle abnormalities that might elude other imaging techniques. With mammography and MRI combined, a comprehensive breast health evaluation is provided. Furthermore, MRI proves particularly useful for BC patients presenting with pulmonary symptoms beyond screening. The use of computed tomography (CT) images for preoperative staging has gained significant traction in such cases. The CT image provides a detailed view of the chest, which can aid healthcare professionals in understanding the extent of disease progression and the potential impact of surgical interventions. In addition to providing anatomical insight, CT images can also be predictive. Medical practitioners are able to discern lymph node metastasis likelihood by meticulously analyzing CT images. It can be used to guide treatment decisions and design personalized care pathways based on this predictive potential. Though Magnetic Resonance Imaging offers the highest level of sensitivity, its costs and time limitations limit its use. Medical professionals gain a deeper understanding of breast health through the synergy of mammography and MRI. CT imaging also provides predictive capabilities, increasing healthcare professionals' ability to fight breast cancer by increasing anatomical clarity simultaneously [18].

3.1.4 Histopathology

Breast cancer (BC) is an extremely difficult diagnosis to make if you don't have access to histopathology along with various other medical imaging modalities [19]. In the field of cancer diagnosis and treatment, histopathology plays an essential role due to its ability to capture phenotypic details. Histopathology Images (HI) are nevertheless subject to substantial limitations within BC's multi-classification, due to the inherent complexity of the samples they capture. Histopathology images show pronounced coherency among cancerous cells, one of the challenges. Due to the high level of similarity between cells, categorizing cells accurately can be difficult. In addition, HI faces a dual challenge of significant intraclass differences as well as relatively low interclass differences. On the basis of visual analysis, this complexity makes identifying different forms of cancer even more challenging.

When compared with images from different classes, histopathology images from the same class have greater resolution, more pronounced contrasts, and substantial visual disparities [20]. Because of this phenomenon, different types of BC can be difficult to distinguish because their images are so drastically different. It is because of this inherent variability that expert interpretation is necessary. Furthermore, gigapixel Whole Slide Images (WSIs) present their own set of challenges. WSIs with resolutions exceeding 1 GB each present formidable challenges for Deep Learning (DL) models. A huge volume of data makes processing WSIs a computationally demanding task. DL models struggle with the challenge of analyzing colossal images efficiently while maintaining depth of insight.

Despite its unrivaled status as the gold standard for BC diagnosis [19], Histopathology faces formidable obstacles during multi-classification applications. Images are complicated by cancerous cells' coherence, intraclass differences, and interclass dynamics. Furthermore, there are intrinsic differences in imaging characteristics within the same class that contribute to the complexity of BC classification. Gigapixel WSIs also present computational challenges, which test DL models' capabilities. It is difficult to diagnose breast cancer based on histopathology due to many challenges, but the pursuit of improved diagnostic accuracy remains a driving force.

3.1.5 Thermography

Another imaging modality used for diagnosing breast cancer (BC) is Breast Thermography, or Thermal Imaging. It uses heat patterns as a key indicator of potential breast abnormalities. It is based on the increased heat generation caused by malignant cells. Breast Thermography offers a number of advantages over its competition. The most prominent feature of this process is that it is non-invasive, painless, and non-contact. It prevents any discomfort that might occur when using traditional imaging techniques, thereby ensuring the safety of the thermographer as well as the patient. Due to its non-invasive nature, Breast Thermography is an attractive option for routine annual medical checkups [21].

Thermography plays a crucial role in early diagnosis of breast cancer when used in conjunction with mammography. Mammography is one of the most effective ways to detect breast cancer in young women, but it is limited by the dense breast tissue and low contrast of the images it produces. Thermography-based techniques are capable of diagnosing cancer 8 to 10 years before mammography at the early stages of development [22]. Thermal imaging has the potential to save lives because of this time advantage. Besides BC, Infrared Thermography can be used to detect a variety of cancers in the initial stages. The test is useful for detecting brain tumors, skin cancer, and more. Nevertheless, its significance goes beyond cancer detection. It can diagnose diseases such as liver disorders, diabetes, and ocular diseases, as well as emerging pathogens such as COVID-19 virus [23]. Infrared Thermography plays a pivotal role in holistic health assessment based on this multifaceted application. Essentially, Breast Thermography uses heat patterns to uncover potential breast anomalies in a non-invasive, patient-friendly manner [21]. It improves early detection of breast cancer when combined with mammography [22]. Infrared thermography's relevance to detection of cancer and disease [23] is bolstered by its broader application in modern medicine. Thermal imaging contributes to better patient outcomes by enabling earlier and more convenient diagnosis of conditions. The advantages and disadvantages of various breast cancer imaging modalities should be weighed after discussing their distinct characteristics and functionality. As shown in Table 3, each technique's strengths and limitations are summarized, making it easier to understand their respective roles in the diagnosis of breast cancer.

Medical experts like pathologists and radiologists must analyze medical images meticulously to detect breast cancer. However, qualitative image assessment has inherent limitations. There is a significant risk of subjectivity when there is a shortage of skilled pathologists. When evaluating numerous cells manually, concentration lapses can occur, leading to misdiagnoses. The monotony of this process makes it time-consuming and error-prone. These challenges can be addressed with advanced image analysis technologies. In addition to enhancing consistency and accuracy, automated methods, like machine learning algorithms, also reduce the need for human intervention. In addition to robust validation, technology adoption must address concerns about data security.

3.2 Computer-Aided Diagnostic (CAD)

In order to provide informed treatment for breast cancer, early detection is crucial. Through algorithms, computer-aided diagnostic systems simplify the detection and localization process. However, studies [24] highlight concerns as well as benefits. It is possible for CAD software to generate high false positive rates, which reduces sensitivity. CAD is inconsistent in its

Table 3 The pros and cons of different breast cancer imaging techniques

Diagnostic Imaging Approach	Advantages	Disadvantages
Mammograms	For the detection of breast cancer, mammography is a highly sensitive imaging technique that is particularly effective for detecting cancerous growths in breasts with more fatty tissue In such contexts, the device's sensitivity allows for early detection of subtle anomalies, leading to an accurate diagnosis	When breasts are dense and women are younger, mammograms are less sensitive. Mammography uses ionizing radiation, which may increase breast cancer risk Due to these factors, it cannot be used as a routine screening tool. As well as cost, there is also the issue of specificity, which can result in increased expenses in some cases. Breast compression is also required for accurate mammograms, causing discomfort and fear. Breast cancer screening and diagnosis are complicated by these factors
Histopathology Images (HI)	A wide array of cancer types can be diagnosed using histopathology images (HI), which provide a comprehensive and intricate picture of the disease. Medical diagnostics rely on their depth and precision to capture intricate details of cells and tissues	Due to irregular staining, lighting variations, and overlapping structures, analyzing histopathology images (HI) is challenging It is difficult to interpret these complexities accurately. Deep Learning (DL) models are also hindered by the large size of HI. These combined challenges must be overcome in order to extract meaningful insights from histopathology images
MRI	Women who are susceptible to risk factors will find this highly advantageous. It is particularly advantageous for women who are more likely to encounter adverse health conditions and other adverse circumstances	It is common for metallic implants to malfunction among patients. Due to their lower level of specificity, these devices are less accurate and more costly than ultrasounds and mammograms
Ultrasound	Ultrasounds are non-invasive and radiation-free. Patients prefer ultrasounds since they do not expose themselves to ionizing radiation due to this inherent characteristic. Furthermore, ultrasounds offer a more effective method of imaging for individuals with dense breast tissue	Ultrasound images, despite their utility, are often limited in resolution, producing poorer image quality than other imaging modalities. Moreover, ultrasounds may not completely cover the breast area during imaging procedures. It is important to combine multiple imaging techniques to ensure a thorough assessment of breast health
Thermal	A vital aspect of Breast Thermography is that it is safe, non-invasive, and painless. Therefore, it can be used routinely to detect breast cancer early. With its high diagnostic accuracy, it is especially useful for women with dense breast tissue	Temperature imaging is not used to diagnose breast cancer directly; instead, it is used to alert doctors of possible changes in breast tissue temperature. In thermography, abnormalities can be detected that indicate further investigation is needed. Typically, mammograms and biopsies confirm the presence of breast cancer or other underlying conditions

accuracy, and some cancerous lesions go undetected. In the development of CAD systems, Moore's law [25] has a significant impact. The ability to perform complex calculations on large datasets is improved by faster, more cost-effective hardware. CAD systems can enhance breast cancer diagnostics by integrating with medical imaging. Nevertheless, limitations highlighted in studies [24] emphasize continued system refinement. Moore's law [25] drives CAD development, enhancing its ability to handle large datasets.

The potential impact of AI-assisted Computer-Assisted Diagnosis (CAD) systems across various healthcare domains has garnered considerable attention [26, 27]. Although AI has been integrated into medical practice, substantial clinical benefits have yet to be realized. Due to this lack of practical applications, it can be difficult to assess AI's efficacy in the healthcare environment, since its promise often outpaces its actual implementation. The integration of AI into clinical practice still faces two major challenges:

- Generalization Challenges of Machine Learning and Deep Learning Models: Machine Learning (ML) and Deep Learning (DL) models are challenged by the intricate and multifaceted nature of complex medical datasets. It is often difficult for these models to generalize to intricate, real-world medical scenarios, despite their ability to recognize patterns. AI models must be able to adapt to patient populations, diseases, and treatment paradigms that are constantly changing in healthcare due to the dynamic data generated.
- Dataset Limitations and Ethical Considerations: High-quality labeled datasets are crucial to the successful development and deployment of AI models. Despite this, such datasets are still difficult to obtain, in part because of ethical and legal concerns about privacy and sharing. Data utilization is further restricted by emerging regulations, making it challenging for AI to tap its full potential for therapeutic advances.

In healthcare, AI has been largely proven in retrospective studies, which examine historical data retrospectively. However, since these studies are retrospective, selection biases are possible, which limits their applicability to prospective studies. In order to determine the true validity of a study, prospective studies are conducted and external validations are performed. The translation of AI-driven technologies from research environments to clinical practice requires these critical steps.

Integrating AI into clinical workflows must bridge the gap between research-driven potential and measurable clinical benefits. A prospective study validates the effectiveness of AI in real-time scenarios and helps refine models for improved generalization. For AI to be truly transformative in healthcare, it is crucial to address dataset challenges through innovative collaborations, follow ethical considerations, and navigate regulatory landscapes.

Even though AI-assisted CAD has tremendous potential in healthcare, its implementation is difficult due to technical obstacles such as model generalization, data availability, and ethical concerns. In order to transform AI from a theoretical potential into a practical reality, the healthcare community should pursue prospective studies and tackle these challenges actively.

3.3 Deep learning

Figure 4 visualizes the evolution of technology over time, delineating distinct epochs that shape the trajectory of progress. Foundational concepts and early computing paradigms were established during the inference period of 1950 to 1970. In the subsequent era from 1970 to 1980, an emphasis was placed on information processing and conceptual frameworks, marking the beginning of the knowledge era. There is an ascending

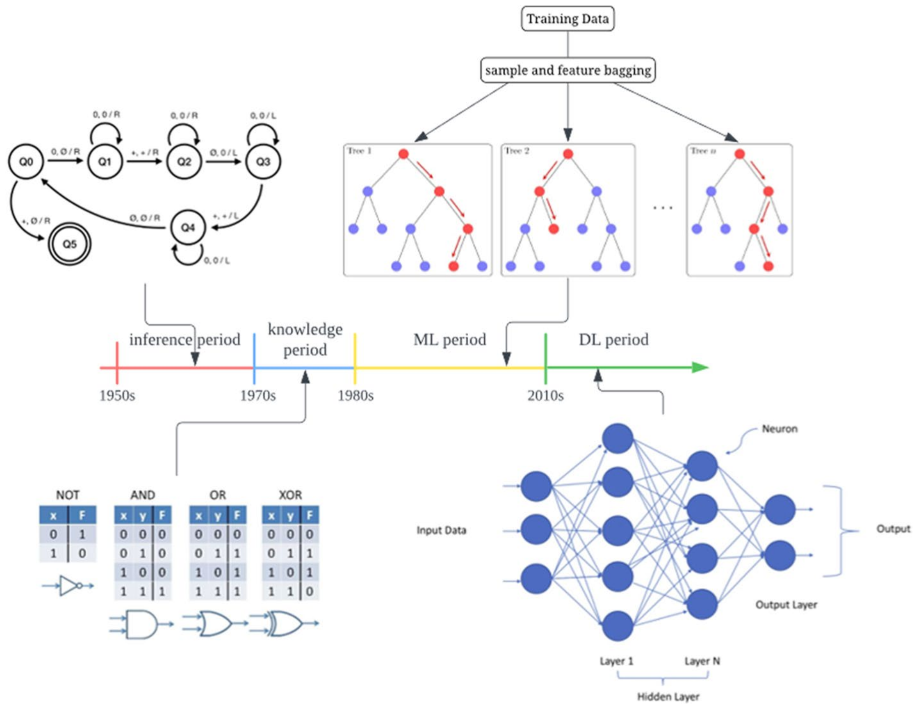


Fig. 4 Navigating technological epochs—tracing the evolution from inference to deep learning

trend in the graph, which marks the advent of machine learning (ML) in the 1980s. The intervening years from 1980 to 2011 witnessed the proliferation of algorithms and methodologies, culminating in the paradigm shift towards data-driven decisions. The last decade has been known as the deep learning period (DL), where vast datasets and sophisticated neural networks have helped push artificial intelligence forward to unprecedented levels. Figure 4 demonstrates the dynamic eras that have shaped the technological landscape, serving as a visual narrative that complements this paper's more thematic discourse.

Machine Learning (ML) and Deep Learning (DL) stand as pivotal cornerstones in the realm of artificial intelligence (AI), facilitating the development of intelligent systems capable of learning from data and making informed decisions. Artificial intelligence uses machine learning to make predictions and make decisions from data patterns without explicit programming. Figure 5 illustrates the relationship between deep learning, machine learning, and artificial intelligence. Artificial intelligence encompasses both deep learning and machine learning, both of which contribute to the broader field of artificial intelligence.

A neural network with interconnected layers is used for Deep Learning (DL) to enhance ML capabilities. Machine learning automates the process of extracting hierarchies of features from raw data by mimicking brain function. Due to its depth and complexity, Deep Learning enhanced the performance of AI systems. In artificial intelligence (AI), Deep Learning (DL) and Machine Learning (ML) utilize data to predict or make decisions. The complexity and depth of the two algorithms for feature extraction

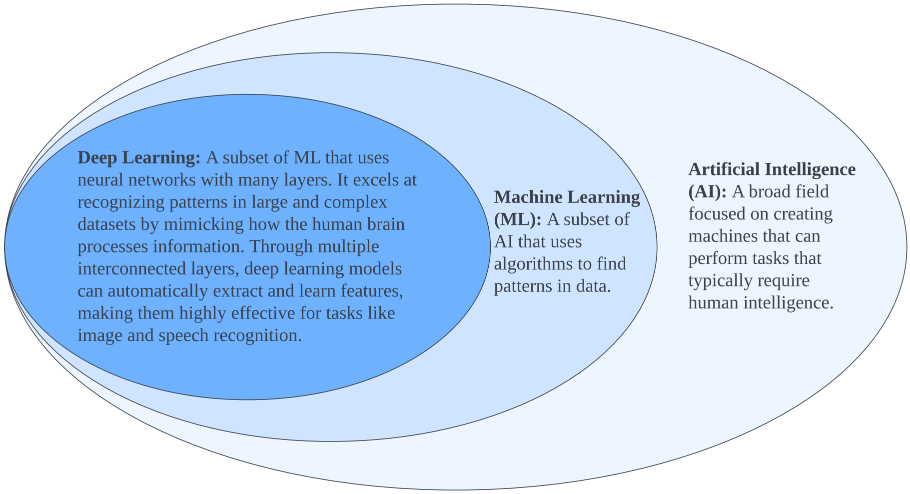


Fig. 5 Deep learning, machine learning, and artificial intelligence interrelate

and representation make them fundamentally different. Machine learning algorithms use features extracted from raw input data as inputs for learning. It is the quality and relevance of the features that directly influence the performance of the model at this point.

Machine Learning (ML): There are many different ways to apply machine learning techniques, including both traditional and sophisticated algorithms. Traditional machine learning involves domain experts identifying and engineering relevant features from data through manual or semi-manual processes. Based on these features, algorithms such as decision trees, support vector machines, random forests, or logistic regression are formulated. Feature selection plays a significant role in model success. Machine learning includes both traditional and more sophisticated algorithms. Traditionally, feature extraction has been a manual or semi-manual process, requiring domain experts to identify and engineer relevant features from data. Decision trees, support vector machines, random forests, and logistic regression algorithms use these features as inputs. A great deal depends on the quality of the features selected for these models.

Deep Learning (DL): Artificial neural networks are used in deep learning to automatically learn hierarchical representations of data using artificial neural networks. Feature engineering can be eliminated by using deep learning algorithms to learn features and representations from raw data. It is particularly useful when dealing with unstructured and complex data, such as images, audio, and text. Deep learning involves feature extraction as an integral part of the learning process [28]. Each layer in a deep neural network learns to represent the input data at different levels. Initial layers learned simple features like edges and corners, while deeper layers learned more complex features. In several domains, including image recognition, natural language processing, and speech synthesis, this ability to automatically learn relevant features has led to significant advances.

Figure 6 shows the different approach of DL and ML.

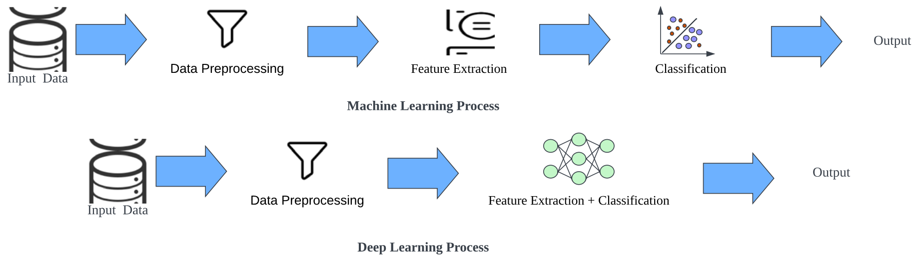


Fig. 6 ML and DL behavior according to feature extraction

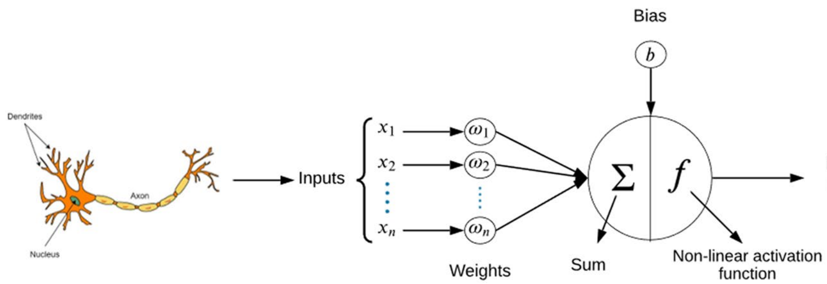


Fig. 7 A simple artificial neuron (σ is a non-linear activation function)

3.4 Convolutional Neural Networks (CNNs) in the realm of computer vision

Artificial Neural Networks (ANNs), including Convolutional Neural Networks (CNNs), have revolutionized computer vision. The ANNs are modeled after the intricate operations of human neural networks, resulting in a revolution in the field of malignant tumor detection. In Fig. 7, artificial neurons interact to create a cohesive structure.

The "input layer" of this architecture serves as the gateway for external information to permeate the neural network. The next step is to unfurl the "dense layers," capable of accommodating the intricate nuances of data processing. Figure 8 illustrates how each neuron within these layers serves as a fundamental unit of computation in computation. The neurons behind these layers are adept at aggregating and processing information inherited from preceding layers, a process characterized by weighted multiplication of inputs and bias terms.

As a result of this cognitive construct, insights from previous hidden layers are combined into a cohesive output layer that represents how the network interprets input data. A comparison of the projected output with the expected output results in a quantification of an error metric that quantifies the discrepancy between actual and projected output. A critical component of refining predictive prowess is the training process. By adjusting synaptic weights, the dynamic process minimizes detected errors by using the principle of backpropagation. To narrow the gap between projected and actual results, the "Gradient Descent Algorithm" systematically aligns the network with actual data patterns. Figures 7 and 8 illustrate how computational innovation and biomimetic principles have propelled the frontiers of scientific discovery and fundamentally changed the landscape of artificial intelligence. Computer vision has been revolutionized by CNNs, which take this concept one step further, simulating human brain functions in processing complex data.

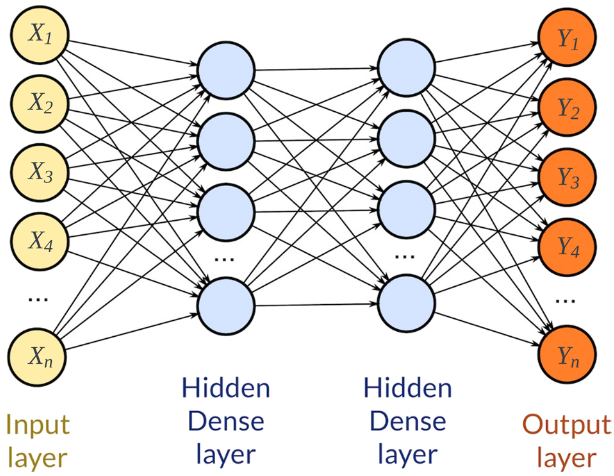


Fig. 8 Artificial neural network architectures

Human brains use CNNs to process visual information. Neurons in the human visual cortex respond only to local visual stimuli. It is possible to replicate the concept with CNNs, which use convolutional layers for detecting local features, pooling layers for preserving important data, and fully connected layers for predicting the future.

In Medical Image Analysis (MIA) and classification, CNN is a highly preferred neural network [29]. There are many applications of CNNs, including computer vision [30], face recognition [31], natural language processing (NLP) [32], audio and video processing [33]. There are two key benefits of CNN architectures: parameter sharing and sparse connections [34]. The basic architecture of a Convolutional Neural Network (CNN) consists of multiple layers that employ filters or kernels, followed by a pooling layer for down sampling. ReLU, Sigmoid, Tanh, and Softmax are activated by the Fully Connected (Dense) layers.

Within the regions covered by the Convolutional layers, the input image is processed through mathematical (Convolutional) operations. Each convolutional layer is activated by a ReLU activation function in order to overcome the 'Vanishing Gradient Problem'. It primarily focuses on dimensionality reduction by using the 'Sliding Window' concept. In general, three methods of pooling exist: maximum pooling, average pooling, and sum pooling. Maximum pooling is the most commonly used method. Fully connected layers (FC) establish connections between each node within a layer. Convolutional and pooling layers are primarily classified by dense layers. The ultimate dense layer generates probability values between 0 and 1 using the 'Softmax' activation function for individual artificial neurons. The architecture of a model has a significant impact on enhancing performance of various applications. As shown in Fig. 9, convolutional neural networks are used to categorize breast cancer tumors.

3.5 Transfer learning

Transfer learning is a machine learning technique with great potential for advancing scientific data analysis. Pre-trained models are employed to enhance the learning process and predict performance on new tasks. As a result of leveraging insights from one domain, transfer learning can accelerate learning in another, even when data distributions differ. In essence, transfer learning improves a target domain's understanding

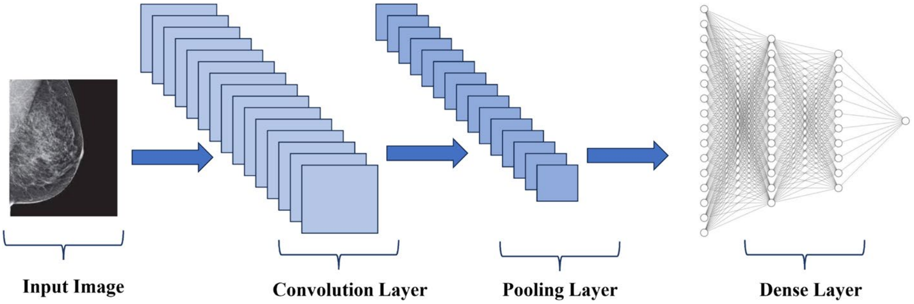


Fig. 9 A basic convolutional neural network for classifying mammogram images

by using knowledge extracted from a source domain. In addition, the benefit is even stronger when labeled data is limited, allowing for a faster convergence of the model during training. There are multiple paradigms for applying transfer learning, such as instance-based transfer, feature-based transfer, and model-based transfer.

Instance-based transfers use the same data instances or examples in both the source and target domains. With feature-based transfer, learned representations (features) are transferred from one domain to another, allowing for better generalization. An advantage of model-based transfer is that it uses pre-trained models, such as neural networks, to enhance the efficiency and effectiveness of the target domain's model.

The LeNet architecture introduced by LeCun et al. [35] sparked the evolution of Convolutional Neural Networks (CNNs). The early CNNs are limited to handwritten numeral classification, which made them less generalizable. As a result of Krizhevsky et al.'s [36] "AlexNet", the landscape has changed significantly. A deeper model and advanced parameter optimization techniques were utilized in this architecture to push the boundaries of image recognition and classification. At the time, hardware limitations limited CNNs' learning capacity, requiring "AlexNet" to be trained on two GPUs simultaneously.

Overfitting challenges are introduced by the pursuit of greater network depth. To overcome these issues, the authors implemented Local Response Normalization (LRN) and overlapping subsampling techniques. VGGNet [37] emerged as a multilayered architecture that surpassed AlexNet by 19 layers. By using compact 3×3 filters, it achieved the effects of larger filters like 5×5 or 7×7, reducing parameter count and complexity. Nevertheless, it was hampered by the extensive use of 140 million parameters.

Inception-V1 [38] or GoogleNet [39] aimed at balancing performance and computational cost. Feature extraction has been revolutionized by the "Inception block"'s merge, transform, and split functions. Despite its heterogeneous topology and representation inefficiencies, "Inception-V1" also introduced auxiliary learners for faster convergence. In "ResNet" architectures, residual connections are used to mitigate vanishing gradient problems. The ImageNet [40] dataset provided an opportunity to demonstrate the versatility of various ResNet models [41], including ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152. Inception-V3 and Inception-V4 [42] incorporated asymmetric filter sizes and refined "Inception-V1" concept. Inception-ResNet combines residual connections for a departure from filter concatenation.

MobileNet [43] is an innovative depthwise separable convolution method introduced by Sandler et al. In this scheme, spatial and depth-wise filtering are efficiently decoupled, thus drastically reducing computations and maintaining a high level of performance. With the lightweight architecture, image tasks could be accelerated on the device. Inverted residual blocks and linear bottlenecks accompanied "MobileNetV2 [44]" to enhance the original concept. It is an attractive choice for a variety of resource-constrained scenarios due to its superior efficiency and its use of fewer parameters. "MobileNetV3 [45]" continued the evolution with tailored architectures for different efficiency and accuracy trade-offs. MobileNetV3 expanded the horizons of efficient CNNs by integrating advanced design elements like h-swish activation functions. MobileNet's journey illustrates the importance of versatile models that can handle a variety of computational requirements. Model scaling is revolutionized by "EfficientNet [46]. EfficientNet uses compound scaling to achieve optimal balance between width, depth, and resolution. Using this approach, the network is systematically scaled up in multiple dimensions to increase performance and efficiency.

With EfficientNet's compound scaling method, depth, width, and resolution of the network are uniformly scaled using a single scaling coefficient. Networks created using this technique are both computationally efficient and representationally powerful. By reducing parameters and computing costs, it achieves state-of-the-art performance on various benchmarks.

This architecture uses depthwise separable convolutions, inverted residual blocks, and a carefully considered scaling schedule. EfficientNet models are able to achieve impressive accuracy in image classification and other computer vision tasks while being resource-efficient. In neural architecture design, EfficientNet serves as a benchmark for balancing performance and efficiency in deep neural networks. As a result, it demonstrates how careful scaling across dimensions can result in improved models that can be deployed across different platforms and devices. Scientific data analysis can be transformed by transfer learning. Utilizing prior knowledge from related fields enhances learning, fosters better predictive performance, and accelerates scientific discoveries across many fields. In modern scientific research, transfer learning techniques continue to refine their applications despite challenges, making them an important asset.

3.6 Data augmentation

Various data transformations are applied to existing data samples in order to artificially increase the diversity and quantity of training data. Machine learning models can be improved through data augmentation by exposing them to a broader range of variations in the data [47]. This section describes various augmentation techniques in detail:

- Adding Noise: The purpose of adding noise to existing data is to deliberately add random variation to it. In this way, real-world fluctuations can be simulated and overfitting reduced, resulting in a more robust model. It is possible to add noise to text data in various ways, such as with Gaussian noise for images. After augmented data is provided, the model is able to learn from a broader spectrum of examples, improving its generalization and performance on new data.
- Cropping: In data augmentation, cropping involves removing unwanted portions of an image. In this way, the model is more able to cope with changes in object orientation and scale. The random cropping process extracts different sections of an image, introducing diversity and improving the model's ability to identify objects from different viewpoints. In training datasets that lack diversity in object placement or have images of varying dimensions, cropping is particularly useful.

- Flipping: Data augmentation techniques such as flipping involve mirroring an image or data vertically or horizontally. Flipping is mostly used in image classification and object detection to expand datasets and introduce variations that enhance robust feature learning. Thus, this method effectively counteracts the bias associated with object orientation at the outset, increasing generalizability across diverse real-world scenarios.
- Scaling: The process of rescaling involves changing the size of an image or input data. As a result, tasks like image classification and object detection can be performed. Changing the scale increases dataset diversity, helping the model learn a broader range of features. The ability to handle various scales in real-world scenarios is enhanced by rescaling, as it anti-biases biases due to specific sizes in the original data.
- Brightness: The brightness level of an image or input data can be adjusted as part of data augmentation. It is particularly relevant to image analysis and object recognition since it introduces variations in illumination conditions into the dataset. In a real-world situation, the model learns to recognize objects under different illumination settings by modifying brightness.
- Rotation: Images are rotated using a specific angle as a data augmentation technique. Image classification and object detection are particularly well suited to rotation, as it allows the model to learn features from different perspectives. It addresses biases that may be associated with specific object orientations in the initial data, enhancing the model's ability to adapt to various angles in real-world situations.
- Translation: The translation process involves shifting images or input data horizontally or vertically. Images and objects can be recognized using this technique. By translating the dataset, the model is able to learn features from different positions as a result of added diversity. It is particularly useful in counteracting bias associated with specific object placements in the initial dataset, improving the model's ability to generalize across different scenarios.
- Contrast: Data augmentation technique contrast adjustment involves modifying the contrast level of images or input data. By introducing varying contrast levels, this approach improves image classification and object detection. Models can be trained to differentiate objects under different visual conditions by varying contrast, which leads to improved performance and adaptability.
- Saturation: Color saturation is the process of altering the intensity of colors in an image or input data. Images can be classified and objects can be detected using this technique. In order to identify objects in various color conditions, the model learns how to manipulate saturation levels. In real-world situations, this approach is especially effective at overcoming biases caused by specific color representations in the original dataset.
- Color augmentation: Color augmentation involves modifying the color properties of images or input data using various techniques. Images and objects can be recognized using this approach. As the model learns features under different visual conditions due to color changes, the dataset becomes more diverse. It enhances the model's ability to adapt to different color scenarios in real-world settings by mitigating biases caused by specific color representations in the initial data.

Models of breast cancer detection benefit greatly from data augmentation. Mammographic images are rotated, flipped, and colored to represent real-world scenarios [48]. The model is enhanced to recognize cancerous features under changing lighting conditions, orientations, and colors. Due to its improved generalization, the model can more accurately and reliably diagnose breast cancer in various cases. Figure 10 illustrates data augmentation techniques.

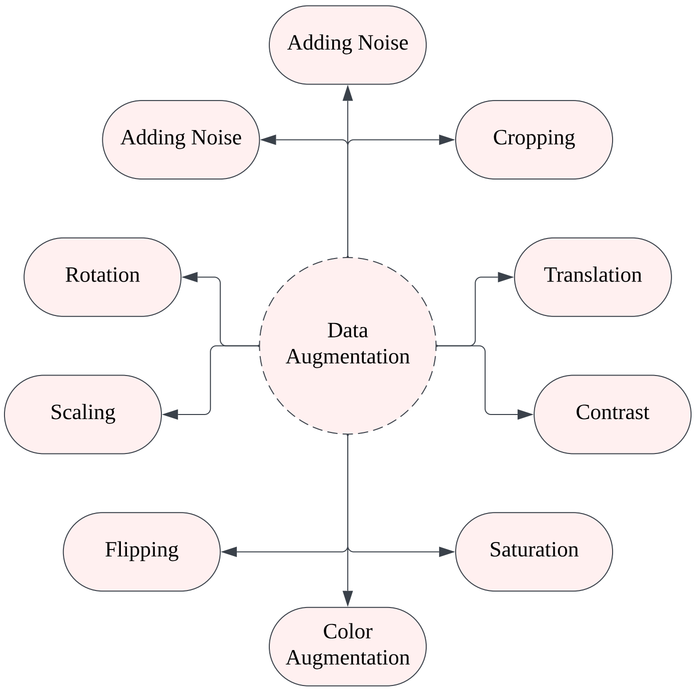


Fig. 10 Image data augmentation techniques

3.7 Hybrid methods in deep learning and optimization algorithms

Deep learning (DL) combined with optimization algorithms, such as metaheuristics, represents a promising research area in artificial intelligence. Hybrid methods leverage the strengths of both domains to enhance machine learning performance and robustness. This section explores the synergy between deep learning and optimization techniques in breast cancer detection, highlighting their applications and contributions.

3.7.1 Enhancing deep learning with metaheuristics

In optimization, metaheuristics refer to high-level procedures that can identify, generate, or select a heuristic that can provide a satisfactory solution. In combination with DL, these algorithms can improve model accuracy and generalization significantly.

Optimizing Neural Network: Optimizing neural network architectures is another key application of hybrid methods. Genetic algorithms and particle swarm optimization can automatically search for the best combination of layers, nodes, and connections for a breast cancer detection model, leading to more efficient and powerful results [49]. With these optimized architectures, medical images can capture complex patterns, improving diagnostic accuracy. Furthermore, metaheuristic algorithms can be used to enhance the performance of neural networks to detect breast cancer [50].

Hyperparameter Tuning: Additionally, hybrid methods excel at hyperparameter tuning, an essential step in the training of deep learning models. The use of metaheuristics allows researchers to identify the optimal set of hyperparameters, including learning rates, batch sizes, and regularization parameters. To fine-tune these parameters, techniques such as ant colony optimization (ACO) [51] and simulated annealing [52] have been successfully used.

Feature Selection and Extraction: The use of hybrid methods is also effective in the selection and extraction of features in addition to optimizing architectures and hyperparameters. The use of metaheuristics can reduce dimensionality and improve the efficiency of deep learning models when applied to large datasets. Medical images can be used to distinguish between benign and malignant features in breast cancer detection. The use of metaheuristic algorithms in feature selection for medical image analysis has been demonstrated, for instance, in a GA-based hierarchical feature selection approach [53].

3.7.2 Applications in breast *cancer* detection

Breast cancer detection using hybrid deep learning and optimization algorithms has yielded remarkable results. Through the combination of deep learning and metaheuristics, these methods can enhance detection accuracy and provide more reliable diagnostic tools. Hybrid methods have demonstrated their effectiveness in a variety of medical imaging tasks, as demonstrated in [54] and [55]. This area is also benefitting from the Enhanced Firefly Algorithm for Constrained Numerical Optimization [56], which provides improved optimization techniques for enhancing DL models' performance in breast cancer detection.

Improving Diagnostic Accuracy: The use of hybrid methods has improved breast cancer detection models' accuracy. These methods optimize feature selection, model parameters, and training processes to make DL models more effective for detecting malignant patterns in medical images such as mammograms, ultrasounds, and MRIs. It is critical for patient outcomes to have improved diagnostic accuracy because it not only aids in early detection, but also reduces false positives and false negatives.

Enhancing Model Robustness and Generalization: Metaheuristics are useful in developing models that are not only accurate but also robust to variations in data. Medical applications require robustness because patient data variability can significantly impact diagnostic performance. It is easier to generalize hybrid models across different datasets and patient populations. Adaptability ensures consistently good performance in diverse clinical settings, increasing the reliability and applicability of diagnostic tools.

Case Studies and Practical Implementations: Hybrid methods have been implemented in real-life situations to detect breast cancer. DL and optimization techniques in hybrid systems outperform traditional methods, increasing diagnoses' accuracy and speed. The integration of PSO with convolutional neural networks (CNNs) has improved identifiers and categorization of breast lesions.

Integration with Healthcare Systems: The integration of hybrid models into existing healthcare systems can have a significant impact on workflow and decision-making. These models help healthcare professionals diagnose patients accurately and expeditiously, which leads to improved treatment outcomes. In addition, hybrid methods can reduce the burden on radiologists and improve healthcare efficiency by streamlining the diagnostic process. Deep learning and optimization algorithms, especially metaheuristics, offer significant advancements in breast cancer detection. Diagnostic accuracy, model robustness, and workflow efficiency are all improved by these hybrid methods. In addition to optimizing neural network architectures, hyperparameters, and feature selection processes, these approaches provide more reliable and efficient diagnostic tools. The implementation of these hybrid methods will lead to better outcomes for patients and more effective healthcare solutions with continued research and development.

4 Explainability of deep learning models in breast *cancer*

Deep learning (DL) has shown promising results in the diagnosis and prognosis of breast cancer, significantly improving the accuracy and efficiency of medical imaging and diagnostic tasks. In spite of this, clinical adoption and trust are hindered due to the black-box nature of these models. To address these issues, Explainable AI (XAI) aims to make the decision-making process of deep learning models more transparent and understandable for medical practitioners.

Explainability in healthcare is essential for several reasons, including ensuring trust in the model's predictions, enabling understanding of complex medical conditions, and enabling regulatory compliance. According to Wani et al. [57], explainable AI integration in the Internet of Medical Things (IoMT) presents a variety of techniques, challenges, and opportunities. They described how XAI can enhance the interpretability of models and maintain high performance while unraveling the complexities of AI-driven healthcare applications. Saraswat et al. [58] also discussed how XAI can bridge the gap between AI models and human expertise in Healthcare 5.0. The authors discussed the opportunities and challenges of implementing XAI in healthcare, including improving diagnostic accuracy, patient trust, and regulatory compliance, while addressing issues related to model complexity and data privacy.

In order to diagnose breast cancer, medical imaging data such as mammograms, ultrasounds, and MRI scans are often analyzed. In recent years, deep learning models, particularly convolutional neural networks (CNNs), have been widely used to detect tumors, segment them, and classify them. In spite of their success, these models may not be easy to interpret, which can hinder their clinical application. It has been discovered that a number of XAI techniques can be used to enhance the transparency of DL models when diagnosing breast cancer.

Rajpal et al. [59] suggested the XAI-MethylMarker approach can be used to discover biomarkers of breast cancer subtypes using methylation data. In this two-stage framework, autoencoders are used for dimensionality reduction, feed-forward neural networks are used for classification, and then a biomarker discovery algorithm is used to identify 52 biomarkers. With fivefold cross-validation, they identified clinically relevant biomarkers associated with druggable genes, prognostic outcomes, and enriched pathways associated with breast cancer subtypes.

- Saliency Maps: A saliency map indicates which regions in an image the model considers important. During the analysis of a mammogram, a saliency map can reveal which areas contributed most to the classification of the region as malignant or benign.
- Class Activation Maps (CAMs) and Grad-CAM: Visual explanations are provided by heatmaps that identify an image's important regions corresponding to specific classes. As a result, radiologists are able to gain a better understanding of which parts of the image influenced the prediction of the model.
- Layer-wise Relevance Propagation (LRP): Each input feature contributes to the prediction of a neural network using LRP. LRP can be used to determine the features (e.g., shape, texture) of a tumor that led to a diagnosis of breast cancer.

Fig. 11 Workflow of Explainable AI (XAI) in deep learning models for breast cancer diagnosis

- SHAP (SHapley Additive exPlanations): For each feature, SHAP values are used to explain the prediction. Model output can be influenced by patient characteristics and imaging features.

The application of XAI to breast cancer and healthcare in general remains challenging, despite its advancements:

- Scalability: Medical imaging datasets tend to be large and computationally intensive, making XAI methods difficult to scale.
- Interpretability vs. Accuracy Trade-off: Maintaining accuracy and making a model more interpretable are often at odds. Developing XAI solutions that balance these two aspects is crucial.
- Domain Expertise Integration: In order for XAI to be effective, medical experts and AI researchers must work together.
- Regulatory and Ethical Considerations: In healthcare, XAI must comply with regulatory standards and address ethical concerns, such as patient privacy.

In conclusion, explainable AI can enhance the clinical applicability and reliability of deep learning models for breast cancer. XAI can improve healthcare diagnostics and patient outcomes by ensuring the decision-making process of these models is transparent. The integration of explainable AI (XAI) with deep learning models for breast cancer must be illustrated in detail, from data collection to clinical validation. Figure 11 illustrates each crucial step and the techniques used to ensure model accuracy and interpretability.

Data collection and preprocessing are the first steps in the process. In this initial stage, medical imaging data such as mammograms, ultrasounds, and MRI scans are collected along with methylation data. Deep learning models rely heavily on these datasets for training and validation. Data collection leads to two primary branches based on the type of data. There are three main branches of medical imaging data: mammograms, ultrasounds, and

MRIs. An ideal deep learning model for image analysis is a Convolutional Neural Network (CNN). The CNN is capable of identifying patterns and features in medical images that are indicative of breast cancer. It is the second branch that deals with biomarkers for breast cancer derived from methylation data. Autoencoders and Feed-Forward Neural Networks are used for this type of data. The autoencoder reduces the dimensions of the data, while the feed-forward neural network makes predictions using the compressed representation.

Explainability techniques are used to interpret the predictions once the models have been trained. It is crucial to gain insight into the model's decision-making process and ensure that clinicians can understand the results. Visual explanations are provided using Grad-CAM, while feature attribution is handled by LIME/SHAP. Grad-CAM generates heatmaps to illustrate the regions in medical images that contribute to the model's predictions, helping to visualize the areas in which the CNN concentrates when making a diagnosis. It is possible to attribute the model's predictions to individual features in the methylation data using LIME and SHAP, which helps to understand the contribution of each feature to the final prediction. Interpretation and validation are the final stages of the workflow. Translating model predictions into clinically actionable insights is crucial. Heatmaps are created to illustrate which areas influence the model's predictions, feature importance plots are created for methylation data to highlight the most significant features influencing the predictions, and clinical validation of biomarkers for breast cancer diagnosis and prognosis is conducted. Figure 11 illustrates how this comprehensive workflow integrates different types of data, deep learning models, and explainability techniques to improve AI's clinical application. This structured approach is designed to ensure that artificial intelligence advances facilitate meaningful, interpretable insights for healthcare professionals.

This workflow describes the entire process of using XAI to diagnose breast cancer, from data collection to clinical validation, ensuring accuracy as well as interpretability. Through enabling AI systems to make their decision-making processes more transparent, XAI can improve patient outcomes and advance medical diagnostics.

5 Related surveys

Recently, there's been a lot of interest in using deep learning for breast cancer research. Deep learning is good at analyzing complex patterns in large sets of data, which can help us understand and detect breast cancer early. Deep learning is used in different ways for breast cancer research, like looking at medical images (such as mammograms) to find suspicious areas, studying the genes related to breast cancer, and combining patient records with molecular information for better diagnosis and predictions. In recent years, many survey works have been reported. In this section, we'll talk about existing survey papers that summarize what we know about using deep learning for breast cancer. These surveys give a good overview of the methods, challenges, and progress in this area. We'll focus on eight surveys that cover various aspects of deep learning applications in breast cancer research, from analyzing images to understanding the genetic aspects. These surveys help us see what we've achieved so far and where we need to focus our efforts in the future. on deep learning that are summarized as follows.

Loizidou et al. [60] presented a comprehensive review of computer-aided breast cancer detection in mammography. They explored the effectiveness of conventional feature-based machine learning and deep learning algorithms, focusing on the detection and classification of micro-calcifications and masses. The authors critically analyzed popular open-access mammography datasets and emphasized the importance of best practices for algorithm validation and reporting metrics. The review concludes with insights into the potential impact of CAD algorithms on reducing biopsies and error rates, while acknowledging challenges like the lack of explainability in machine learning results. The authors recommended further research and validation with an increased amount of clinical data to enhance the clinical utility of Computer-Aided Diagnosis (CAD) systems.

Sahu et al. [61] provided a detailed examination of recent advancements in the realm of artificial intelligence (AI)-based breast cancer detection using mammograms. The authors focused on the evolving landscape of machine learning (ML) and deep learning (DL) techniques, delving into their applications for accurate breast cancer diagnosis. They categorized mammogram-based breast cancer detection techniques within a structured framework, offering a comprehensive overview of the methodologies employed. The paper extensively discusses the significance of DL-based features over hand-crafted features and highlights the preference for transfer learning in scenarios with limited datasets.

Radak et al. [62] conducted a thorough investigation into the application of machine learning (ML) and deep learning (DL) techniques for breast cancer diagnosis and classification, emphasizing their significance in addressing the global health concern of breast cancer. The paper, published in June 2023, delves into the pivotal role of computer-aided diagnosis (CAD) in enhancing the precision of medical condition identification, with a specific focus on breast cancer. The authors highlighted the efficacy of CAD in early and rapid breast cancer diagnosis, particularly through the integration of CAD into mammography analysis. This integration is demonstrated to improve diagnostic accuracy, facilitating prompt medical interventions and potentially enhancing patient outcomes. The paper underscores the significance of various diagnostic methods for breast cancer, including biopsy, mammography, magnetic resonance imaging (MRI), and ultrasound. The authors underscored the significance of integrating these diagnostic methods to enhance precision and facilitate prompt medical interventions.

Thakur et al. [63] presented a systematic review of machine and deep learning techniques for breast cancer identification and classification through medical image modalities. The paper, published by Springer Nature, outlines ten research questions covering image modalities, datasets, pre-processing, segmentation, and classification techniques. The authors reviewed papers and book chapters from 2010 to 2021, focusing on digital mammograms and public datasets in 57% of the studies. Common techniques include noise removal, data augmentation, scaling, and image normalization to address inconsistencies. Segmentation methods include thresholding, region-based, edge-based, clustering-based, and deep learning techniques, with SVM and CNN variants as prevalent classifiers.

Chugh et al. [64] presented a survey on machine learning (ML) and deep learning (DL) applications in breast cancer diagnosis. Published by Springer Nature, the paper emphasizes the importance of early diagnosis for survivability. The authors conducted a review of recently developed computer-aided diagnosis (CAD) systems, comparing Machine Learning (ML) and Deep Learning (DL) techniques with traditional methods. They discussed the technical details, advantages, and disadvantages of each model, and addressed unresolved issues and research gaps in the field. The survey acknowledges the emergence of artificial intelligence technologies assisting radiologists in medical image analysis. ML and DL, as subsets of AI, have demonstrated promising results in breast cancer diagnosis, providing improved efficiency and cost reductions. The paper examines various classifiers of ML and DL approaches, with DL surpassing traditional ML in diagnosing breast carcinoma, particularly with extensive datasets.

Table 4 Technical comparison between related surveys

Ref	Paper selection process	Taxonomy	Future works and open issues	Year analysis	Publisher analysis	Valuation metric analysis	Data pre-processing techniques	Graphical Representation
[60]	✓	x	x	x	x	x	x	✓
[61]	x	✓	✓	x	x	x	x	✓
[62]	x	✓	✓	x	x	✓	x	✓
[63]	✓	✓	✓	✓	✓	✓	✓	✓
[64]	x	✓	✓	x	x	✓	x	✓
[65]	x	x	x	x	x	x	x	✓
[66]	x	✓	✓	x	x	✓	✓	✓
[67]	✓	✓	x	x	x	✓	✓	✓

Yadavendra and Satish [65] conducted a comparative study on breast cancer tumor classification, focusing on the application of classical machine learning (ML) methods and a deep learning (DL) approach. It addresses the growing importance of accurate and timely breast cancer diagnosis. Early detection is highlighted as a critical factor in improving outcomes and reducing the impact of the disease on patients. In their study, the authors explore different ML methods, including logistic regression, random forest, support vector classifier (SVC), AdaBoost classifier, bagging classifier, voting classifier, and the DL method Xception. The research employs a well-established dataset of breast histopathology images, consisting of over two lakhs (200,000) color patches, each sized 50×50 and scanned at a resolution of $40\times$.

Li et al. [66] explored the use of machine learning (ML) and deep learning (DL) techniques to improve patient outcomes by detecting lung cancer early. This study examines the landscape of deep learning applications in breast cancer diagnosis, with a particular focus on augmentations that improve early detection accuracy. The purpose of this review is to explore the advancements in computer-aided diagnosis (CAD) systems designed specifically for breast cancer and to provide a detailed comparison between ML, DL, and conventional diagnostic approaches. By exploring the technical intricacies of CAD systems, we shed light on their potential to revolutionize breast cancer diagnosis. Their survey paper offers clinicians a nuanced perspective on both ML and DL models, bridging theoretical aspects with practical implications.

Wen et al. [67] explored recent developments in machine learning (ML) and deep learning (DL) for breast cancer diagnosis, focusing on applications of AI to computer-aided diagnosis. The researchers used a systematic approach to evaluate 166 papers covering ML and DL for breast cancer detection. The review included analysis of frequently used datasets, preprocessing techniques, and classification methods. In addition, the authors provided insights into their effectiveness and limitations through a variety of methods and challenges. ML and DL applications in breast cancer diagnosis were analyzed comprehensively for insights into the current state of the art.

Table 4 serves as a comprehensive summary, presenting insights into taxonomy, paper selection processes, future works, open issues, fundamental aspects of deep learning techniques, addressed problems, and diverse analyses. Distinguishing itself from prior surveys, this paper delves specifically into the realm of deep learning applications in breast cancer diagnosis, providing a thorough classification of all covered papers. Numerous shortcomings observed in previous works have motivated the undertaking of this research:

- Temporal Coverage: Unlike many earlier reviews focusing on papers published before 2020, this study considers the latest research up to 2023, ensuring a more up-to-date perspective on advancements in deep learning for breast cancer diagnosis.
- Overlooking Key Topics: Numerous papers in the field often overlook crucial topics related to deep learning in breast cancer, such as diverse architectures, data preprocessing techniques, and model evaluation methodologies. This survey aims to bridge these gaps.
- Focused Approach: Previous works often lacked a specific focus on deep learning methods for breast cancer diagnosis, often concentrating on specific subtopics. In contrast, this paper centers on the overarching theme of deep learning applications in breast cancer.
- Technical Depth: Technical explanations and analyses of deep learning approaches were often insufficient in prior reviews. This study extensively examines various deep learning techniques in breast cancer diagnosis, providing a deeper understanding of their intricacies.
- Visual Aids for Comprehension: While graphical representation was absent in many prior reviews, this study employs visual aids to enhance comprehension, facilitating a more accessible understanding of complex methodologies.
- Systematic Classification: This study categorizes reviewed papers into relevant subcategories, offering a systematic classification that aids in a more detailed and nuanced comparison.
- Rigorous Paper Selection Process: Existing works often lack a systematic paper selection process, while this study rigorously outlines its methodology, ensuring transparency and reliability in the selection of literature.
- Comprehensive Analysis: A comparison of different evaluation metrics and datasets is often overlooked in previous works, and this paper provides a comprehensive analysis of these aspects, offering readers a more nuanced understanding of the performance metrics employed.
- Applications and Challenges: Analyzing the specific applications and challenges addressed by researchers adds a comparative perspective, aiding readers and researchers in choosing suitable techniques based on their specific objectives and constraints.

Given these shortcomings in existing literature, this survey paper is dedicated to deepening our understanding of the landscape of deep learning applications in breast cancer diagnosis, addressing gaps, and providing a comprehensive overview of recent advancements in this critical area of medical research. In the context of deep learning applications for breast cancer, this survey paper conducts a detailed comparison of various technical and non-technical metrics, shedding light on the evolution of research in the field.

6 Deep learning-based approaches to detect breast *cancer*

This section discusses advanced deep learning techniques that have significantly impacted breast cancer detection. The focus of this paper is to explore how deep learning methodologies can enhance diagnosis through a variety of imaging types, including convolutional neural networks (CNNs) and optimization algorithms.

6.1 Mammographic imaging

In this subsection, deep-learning approaches are used to detect breast cancer using mammography imaging. The purpose of this paper is to discuss how deep learning techniques have been applied to mammographic data to improve the accuracy and efficiency of breast cancer diagnosis. As Table 5 represents Kerschke et al. [68] proposed a retrospective approach for evaluating breast cancer screening performance. Women 50–69 years of age underwent 2257 full-field digital mammography screenings in 2011–2013. Using deep learning AI, each recalled lesion was scored (0–95), indicating its likelihood of being cancerous. In contrast to human readers, the sensitivity of AI on lesion level and non-FPRs for women were estimated as functions of the classification cutoff. This study aims to test whether an AI system based on deep learning can distinguish benign from malignant abnormalities in mammograms. In this way, the AI system improved the discrimination between benign and malignant lesions, thus increasing its positive predictive value while reducing its sensitivity. It is important to consider whether the loss of sensitivity can be traded for a reduction in false-positive recall, as those early-stage cancers may have relevance to improving survival. AI-assisted recall decisions seem to benefit the masses, especially since the system achieved comparable sensitivity for readers with less false positives.

According to Houssein et al. [69], a deep learning architecture based on a marine predator algorithm can detect breast cancer faster. Based on hybrid CNNs, an improved optimization algorithm, and transfer learning, this paper proposes a novel classification model to help radiologists diagnose breast cancer efficiently. They improved the MPA by using an opposition-based learning strategy to address its inherent weaknesses. The improved marine predators algorithm (IMPA) is used to determine the optimal hyperparameters for CNN architectures. A CNN model called ResNet50 (residual network) is used in the proposed method. An IMPA-ResNet50 architecture is created by combining this model with the IMPA algorithm. CBIS-DDSM curated breast imaging subset and MIAS mammographic datasets are used for analysis. Various approaches have been compared with the proposed model. In comparison to the state-of-the-art approaches, the proposed model performed better. In addition to GSA-ResNet50, HHO-ResNet50, WOA-ResNet50, and MPA-ResNet50, the counterpart algorithms are also hybrids with the ResNet50 architecture. IMPA-ResNet50 has achieved a better performance than its counterparts, according to the results.

Qu et al. [70] proposed deep learning on digital mammography for expert-level diagnosis accuracy in breast cancer detection. Comparing DL models trained on historical mammograms with only image-level pathology labels with experienced radiologists, they performed surprisingly well. With 5979 historical exams obtained before September 2017 with biopsy-verified pathology, DenseNet was trained and cross-validated, and tested with 1194 newly acquired cases after that. DL predictions generated higher ROCs than radiologists' ratings in both cross-validation and test sets. This study aims to

Table 5 Deep learning-based approaches to breast cancer detection in Mammogram images

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[68]	Full-field digital mammography screenings (2257 mammography screenings)	Nan	CNN	CNN (Transpara) for automated breast cancer detection	Evaluation of AI and human detection of breast cancer in mammograms	Proportion of False-Positives (FPR) Reduction: 62.0%, Non-False Positive Rate (non-FPR): 11.1%, Sensitivity: 91.1%, Positive Predictive Value of Recall (PPV-1): 16.5%, Non-FPR for Mass-Related Lesions: 36.7%, Sensitivity for Mass-Related Lesions: 97.1%
[69]	Mammographic image analysis society (MIAS) and CBIS-DDSM (Curated Breast Imaging Subset of DDSM)	Nan	IMPA-ResNet50 hybrid architecture	Improved Marine Predators Algorithm with ResNet50 and OBL strategy for hyperparameter optimization	Breast Cancer Classification	CBIS-DDSM (Accuracy: 98.32% Sensitivity: 96.61% Specificity: 98.56% Precision: 98.68% F-score: 97.65%), MIAS (Accuracy: 98.88% Sensitivity: 97.61% Specificity: 98.40% Precision: 98.30% F-score: 97.10%)
[70]	Digital mammography collected	1024×832	DenseNet-121	Dataset preparation, model training (AlexNet, VGG, ResNet, DenseNet), performance evaluation	Breast Cancer Classification	Cross-validation AUCs: 0.953, Test Set Mean AUC: 0.940, sensitivity: 0.9, specificity: 0.85
[71]	DDSM and CBIS-DDSM	299×299	Ensemble models with VGG16, InceptionV3, and VGG19	Pre-processing, Ensemble learning using VGG16, InceptionV3, VGG19, Neural network classification	Breast Cancer Classification	Accuracy: 98.10, Sensitivity: 97.01, Specificity: 99.12, Precision, Recall and F1 score: 0.98, 0.98, 0.98

Table 5 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[72]	MIAS And DDSM	1024 × 1024	ShCNN (IoT-based system)	IoT simulation, FACS, Fitness measure, Pre-processing, Feature extraction, Data augmentation, Breast cancer classification	Breast cancer classification method using an Internet of Things (IoT)–based smart healthcare system	MIAS (TPR: 99.45%, Sensitivity: 96.10%, Specificity: 91.80%, Accuracy: 91.56%), DDSM (TPR: 97.43%, Sensitivity: 90.57%, Specificity: 88.31%, Accuracy: 90.17%)
[73]	Mini-MIAS (Mammographic Image Analysis Society Mini Mammogram Database) and DDSM	Nan	OMLTS-DLCN	Preprocessing, Segmentation (OKMT-SGO), Feature extraction (CapsNet), Classification (BPNN)	Breast cancer diagnosis model using Capsule Networks (CapsNet) on mammogram images	Mini-MIAS: Accuracy—98.50%, Sensitivity—98.46%, Specificity—99.08%, F-measure—98.91% DDSM: Accuracy—97.55%, Sensitivity—97.579%, Specificity—98.217%, F-measure—97.753%
[74]	BCDR01 (Breast Cancer Digital Repository consists of patients with at least one breast cancer)	227 × 227	Modified DL techniques	Preprocessing, Convolution, ReLU layers, Batch normalization, Max pooling, Fully connected layers, Soft-max	Diagnosing tumor location in breast cancer using modified deep learning techniques, especially CNNs, on mammogram images	Nan

Table 5 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[75]	INbreast and MIAS	224×224	MobileNetV2	Image preprocessing, Feature extraction, Model development	Efficient deep learning-based anomaly detection for early breast cancer detection	INbreast: Normal vs. Cancer (AUC: 94.25%, F1 Score: 91.14%, Accuracy: 91.84%), Normal vs. Abnormal (AUC: 89.79%, F1 Score: 87.17%, Accuracy: 88.44%); MIAS: Normal vs. Cancer (AUC: 97.36%, F1 Score: 92.15%, Accuracy: 91.76%), Normal vs. Abnormal (AUC: 81.40%, F1 Score: 72.59%, Accuracy: 75.60%)
[76]	MIAS and INbreast and WDBC (Wisconsin Diagnosis Breast Cancer)	224×224	CSOA-wKNN	Feature extraction using metaheuristic algorithms (PSO, DFOA, CSOA), wKNN classification WDBC dataset: Feature vectors directly used for classification with wKNN	Automatic Classification of Mammograms (Benign/Malignant) and Breast Masses (Benign/Malignant)	Accuracy: MIAS: 74.78%, INbreast: 73.45%, WDBC: 95.61% Sensitivity: MIAS: 70.59%, INbreast: 71.93%, WDBC: 93.87% Specificity: MIAS: 78.13%, INbreast: 75%, WDBC: 96.64% F1-Score: MIAS: 71.29%, INbreast: 69.07%, WDBC: 95.50% FPR: MIAS: 21.87%, INbreast: 25%, WDBC: 3.36% CER: MIAS: 25.22%, INbreast: 26.549%, WDBC: 4.394%

Table 5 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[77]	Collected Mammography Data (2321 females, 9284 images)	512 × 512	BSNet	Mammography Transformation, Feature Extraction, Decision Layer	Classification of Breast Cancer Subtypes (HR +, HR-, Benign) based on Mammography Images	AUC: 0.89 (Test Set), 0.92 (External Validation Set) Micro AUC: 0.892, Macro AUC: 0.884 Benign AUC: 0.972, HR + AUC: 0.900, HR- AUC: 0.764 Accuracy: 0.777, Precision: 0.760, Recall: 0.768, F1-score: 0.777
[78]	Collected Mammograms (200 patients)	224 × 224	LRP-NET	Deep Learning CNN (VGG16-based LRP-NET) for Longitudinal Risk Prediction	Prediction of Breast Cancer Risk Based on Longitudinal Mammograms	AUC: 0.67, Accuracy: 0.61 for Breast Cancer Risk
[79]	CBIS-DDSM and INbreast and MIAS	1333 × 800	PAA algorithm and Faster R-CNN	Data Preprocessing, First-Stage: Detector + Two-Branch ROI Detector, Third-Stage: Classifier	Breast Cancer Detection in Mammography Images	Precision: 0.829, Recall: 0.831, Specificity: 0.828, AUC: 0.911 (mAP of 0.651 for Three-Stage PAA with Post-Processing)
[80]	DDSM and MIAS	1024 × 1024	Hybrid CNN with Radon transform	Radon transform, Data augmentation, Hybrid CNN, Classification, Mathematical Morphological-Based Segmentation	Detection and Diagnosis of Breast Cancer Using Hybrid CNN Architecture	DDSM: Se—97.91%, Sp—97.83%, Acc—98.44%, JI—98.57% MIAS: Se—98%, Sp—98.66%, Acc—99.17%, JI—98.07%

Table 5 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[81]	CBIS-DDSM	Nan	HRDO-FL + E-RNN	Federated learning, Feature extraction, Proposed HDRO, ERNN-based classification	Breast Cancer Detection	Accuracy: 95.73%, Sensitivity: 95.75%, Specificity: 95.70%, Precision: 95.78%, FPR: 4.30%, FNR: 4.25%, NPV: 95.70%, FDR: 4.22%, F1-Score: 95.76%, MCC: 91.45%
[82]	MIAS	224 × 224	GoogleLeNet architecture with SVM and Decision Tree and Random Forest or Naive Bayes	Image Enhancement, Segmentation, Feature Extraction, Classification	Breast Cancer Detection and Classification (Benign vs. Malignant)	Accuracy: 99.12%, Sensitivity: 99.89%, Specificity: 98.45%, Dice Coefficient: 82.15%, Jaccard Coefficient: 89.11%
[83]	No name Mammogram breast lesions (BI-RADS)	64 × 64	DMD-CNN	Pre-processing, DMD-CNN model, Classification output	the early detection of breast cancer, using deep-learning and FPGA techniques	Accuracy: 98.2%, Precision: 0.99, Recall: 0.98
[84]	CBIS-DDSM	227 × 227	Ensemble of AlexNet and ResNet-50 and ResNet-101 and DenseNet-201	Suspected region extraction, Ensemble learning, Shallow classifier, Fusion	Detection of breast cancer focusing on Suspected Nodule Regions (SNRs)	Accuracy: 94%

Table 5 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[85]	CBIS-DDSM	224 × 224	Fisher Information Networks CNN	Dataset Preparation, Feature Extraction, Dimensionality Reduction (PCA), MLP Model, FI Metric Calculation, Distance Calculation, Gaussian Radial Kernel, MDS, Visualizing Latent Space, 'Patient-Like-Me' Approach, Comparison to UMAP	Improving breast cancer diagnosis using advanced ML techniques, enhancing explainability, and reducing overdiagnosis/missed diagnoses	Test set accuracy: 72%, Background class accuracy: 89.6%
[86]	Collected DBT images (230 image pairs (115 M/N pairs and 115 N/N pairs))	256 × 256	VGG16 and Resnet50 and DenseNet121 and Xception	Bootstrap sampling, Random under-sampling, DL model	Predicting stromal invasion in breast cancer using DBT images	AUC: 0.75
[87]	Mini-DDSM	AlexNet: 227 × 227, VGG16/ResNet50: 224 × 224	CNN with/without transfer learning (AlexNet and VGG16 and ResNet50)	Image preprocessing (resizing, rescaling, data augmentation), use of CNN architectures (AlexNet, VGG16, ResNet50) for image classification, transfer learning with models pretrained on ImageNet and fine-tuned on Mini-DDSM	Classification of Benign and Malignant Breast Tumors in Mammogram Images	Specificity: 0.9194, Sensitivity: 0.9646, AUC: 0.8652, Accuracy: 65.89%

Table 5 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[88]	Collected DBT images (230 image pairs (115 M/N pairs and 115 N/N pairs))	256×256	BilAD (modified from Xception CNN)	Bilateral asymmetrical detection using a modified Xception CNN model for bilateral breast images, Unilateral CNN model for single breast images	Diagnosing breast cancer via examining bilateral breast tissue asymmetry in DBT images	Accuracy: 0.84, Sensitivity: 0.73, Specificity: 0.93, AUC-ROC: 0.90

demonstrate that deep learning can be trained using big-data resources to create expert-level breast cancer diagnosis models. According to their study, a well-defined deep neural network can be trained on historical mammogram-pathology pairs to accurately detect breast cancer. Furthermore, their study demonstrated that mammography-specific neural networks can be further optimized. Neuronal networks should provide radiologists with knowledge instead of them providing it themselves.

The ensemble model proposed by Nemade et al. [71] uses deep learning to classify breast cancers. A single pretrained model or an ensemble of such models is used for Transfer Learning (TL). In preprocessing, datasets are resized and normalized, and imbalances are corrected. The use of CNN models, including the well-known VGG16, InceptionV3, and VGG19, facilitates robust ensemble results. The primary objective is to develop an efficient Computer-Assisted Diagnosis (CAD) model to categorize breast masses. In order to boost the performance of deep models, an ensemble classifier was created by stacking previously trained models on extensive datasets. The key strategies used include balancing skewed data, TL, and early stopping. This research is notable for its efforts to address dataset imbalances to avoid model bias, leverage well-established CNN models and enhance their performance with ensemble classifiers. Two mammography datasets were evaluated: DDSM and CBIS-DDSM. Three pre-trained convolutional neural network (CNN) models were used as base classifiers to enhance breast lesion classification from mammography scans. In total, two ensemble models were trained. For classification, Ensemble Model 1 uses a linear meta-learner in the form of logistic regression, and Ensemble Model 2 uses a neural network. As a result, the proposed models outperformed existing state-of-the-art systems in terms of breast cancer classification performance. Thus, the proposed models may provide medical professionals with a useful tool for accurately classifying breast lesions. By comparing the proposed system with existing methods, it was demonstrated to be robust and superior. Nevertheless, the study used only fixed-size images that had already been curated, which limited its results.

Majji et al. [72] explored smart IoT applications in breast cancer detection. A feedback artificial crowd search-based Shepherd Convolutional Neural Network (ShCNN) is used in this research as an IoT-based smart healthcare system. As a first step, the FACS method is used to determine the best routes. To determine the optimal route, this method takes into account factors like distance, energy, and latency. After merging the CSA and FAT, the FACS is developed. The system categorizes breast cancer at the base station once routing is complete. The pre-processed mammography images are analyzed for features such as area, mean, and contrast. Breast cancer is classified by ShCNN, a FACS algorithm that enhances the image quality via data augmentation. An evaluation of this method is based on the amount of energy it consumes, the time it takes, and the accuracy of the results. Results showed a high True Positive Rate (TPR) of 99.45%, a sensitivity of 96.10%, and a 91.56% accuracy. IoT can be used in modern healthcare to provide a reliable method for breast cancer classification.

Kavitha et al. [73] suggested that deep learning-based capsule neural networks could be used to diagnose breast cancer from mammogram images. OMLTS-DLCN is a new model for early breast cancer detection utilizing multi-level thresholds and DL enabled capsule networks. The segmentation model incorporates Adaptive Fuzzy Median Filtering for noise removal, Optimal Kapur's Multilevel Thresholding with Shell Game Optimization for noise removal. For feature extraction, a Capsule Network is used, and for classification, a Back-Propagation Neural Network is used. OMLTS-DLCN's accuracy rates on Mini-MIAS and DDSM benchmark datasets were 98.50% and 97.55%, respectively.

Ghoushchi et al. [74] reported that deep learning could be used to predict where tumors are located in breast cancer. In this study, modified deep learning (DL) techniques are leveraged to pinpoint cancer tumor locations using machine learning. This study analyzes resized and segmented BCDRD01 data. To address imbalances between classes, the first experiments employed a simple architecture with a weighted function. The Gabor filter and visualization of the images helped pinpoint the breast tissue's location. The subsequent experiments used more complex network architectures, including a VGG (9 layers, 2.9 million parameters) and a 10 layer, 0.9 million parameters network. Breast cancer lesions can be distinguished well using convolutional neural networks (CNNs).

Alloqmani et al. [75] proposed using deep learning to detect anomalies in breast cancer. This paper presents a deep learning framework for detecting benign and malignant breast abnormalities using only normal data. A major challenge in the medical domain is the imbalance of data. It consists of two stages: (1) image pre-processing, and (2) MobileNetV2 pre-trained feature extraction, followed by single-layer perceptron. Breast cancer diagnosis models were developed on the basis of different datasets. To reduce manual labor and potential diagnostic errors, the project's main objective was to tackle the prevalent issue of data imbalance in the medical sector. It also addresses the limitations of earlier studies using patch-based methods. Earlier studies have been able to process entire images due to the current framework. Tests of this approach were conducted on two public datasets: INbreast and MIAS. The INbreast dataset showed an AUC of 94.25% for cancer detection and 89.79% for abnormalities. Cancer detection AUC values for MIAS were 97.36% and abnormality detection AUC values were 81.40%.

For breast cancer severity classification, Chakravarthy et al. [76] proposed a metaheuristic weighted k-nearest neighbor algorithm that uses deep learning. This study aims to improve the classification of breast cancer, one of the most prevalent and deadly types of cancer in women. This research integrates deep learning, metaheuristics, and classification algorithms to increase the survival rate and enable early diagnosis. In the MIAS, INbreast, and WDBC databases, digital mammogram images are used to classify breast cancer severity. By using three nature-inspired algorithms (i.e., particle swarm optimization (PSO), dragon-fly optimization algorithm (DFOA), and crow-search optimization algorithm (CSOA)), the weighted k-nearest neighbor (wKNN) algorithm's classification performance was enhanced using transfer learning for feature extraction.

Zhang et al. [77] developed an image-based weakly supervised deep learning framework to detect breast cancer in women with HR status. It focuses on the development of a deep learning framework, called BSNet, for non-invasively diagnosing the hormone receptor (HR) status of breast cancer. Breast cancer treatment is largely determined by the hormone receptor status. Breast cancer can be detected with mammography, but its molecular status cannot be determined without a pathological biopsy. For easy diagnosis of breast cancer patients' HR status, a web server has been developed in order to make the model more accessible to medical professionals. The purpose of this tool is to provide precision medical care without invasive procedures such as punctures. Positive results have been found in the study. The average AUC for BSNet on the test and external validation sets, respectively, was 0.89 and 0.92, exceeding those of other baseline models. To refine the model for larger clinical applications, however, more work needs to be done.

Dadsetan et al. [78] proposed using deep learning to predict breast cancer risk. Several deep learning architectures are presented in this research, including LRP-NET, a novel architecture. Specifically, it is designed to detect changes in breast tissue over the course of multiple negative/benign mammograms. Case–control studies are used to predict breast cancer risk in the near future. The imaging tool is based on clinical knowledge and uses four consecutive mammograms to capture bilateral breast tissue changes. In this study, the LRP-NET model is used to examine the relationship between spatial–temporal changes in breast tissue across longitudinal normal screening mammograms and cancer risk. Four consecutive mammographic examinations are used to extract these spatiotemporal features. LRP-NET performed better than other models when it came to predicting breast cancer risk.

Jiang et al. [79] developed a three-stage deep learning framework to detect and categorize breast cancer in mammograms using the PAA algorithm. The primary goal of this study is to enhance the precision and efficiency of medical diagnoses, thereby easing physicians' workloads. Initially, PAA and Faster R-CNN are compared for their detection capabilities for breast lesions. The single-stage PAA algorithm was then extended into a two-stage model utilizing cascading two-branch ROI detectors. In addition, the post-processing algorithm was adaptively adjusted based on breast density classification. For mammograms, the method achieved impressive classification outcomes by combining the ROI classifier with the image classifier in the third stage. This method identifies lesion types, pinpoints lesion locations, and categorizes entire images on the basis of experimental results. However, the research relied solely on input from one perspective. The literature suggests that multi-views and multiple inputs can significantly enhance classification and detection. This model will be streamlined and multi-viewed in the future, according to the researchers. Additionally, the attention module will be used to emphasize the interrelationship between lesions across different views, thus enriching the features of identical lesions.

Raaj [80] developed an artificial intelligence architecture for detecting and diagnosing breast cancer. Classifying mammogram images as normal, benign, or malignant is done using Convolutional Neural Networks (CNN). Radon transform, data augmentation module, and hybrid CNN architecture are included in the proposed system. A radon transform converts each spatial pixel in a source mammogram image into a time–frequency variation image. The existing dataset has been enhanced to better detect breast cancer. Using the hybrid CNN architecture, the data-augmented images are classified into three types. A mathematical morphological-based segmentation algorithm is used to segment cancer pixels. Data from the Digital Database for Screening Mammography (DDSM) and the Mammographic Image Analysis Society (MIAS) is used to estimate a deep learning architecture's performance efficiency. Experimental results for both open access datasets are compared to similar recent studies. According to the experimental results in this article, the methodology clearly segments abnormal mammogram images into cancer regions.

Kumbhare et al. [81] proposed using federated learning to detect breast cancer through heuristic-based deep learning. To enhance accuracy and precision of breast cancer diagnosis, the study introduces an intelligent heuristic-based deep learning framework based on Fuzzy Logic (FL). The HyperEye Medical System uses this model for real-time applications. FL is commonly used in medical diagnosis systems for processing and storing mammogram images. Enhanced Recurrent Neural Networks (E-RNNs) address gradient vanishing and overfitting issues by using DenseNet for feature extraction. HDRO algorithms enhance the efficiency and reliability of RNNs. An evaluation of the proposed HRDO-FL + E-RNN model revealed improved accuracy compared to FL + CNN, DenseNet, Ensemble, and RNN methods. Detecting breast cancer in an earlier stage has both the potential to reduce mortality rates and raise awareness among women, thanks to a new approach to early detection.

Ramesh et al. [82] developed a deep learning architecture to categorize and segment breast cancer. Through deep-learning algorithms, benign from malignant tumors were segmented carefully. By replacing expert annotations and exhaustive pathology reports with this technological leap, physicians are able to detect malignancies with greater precision. GoogleLeNet architecture is used as the framework for this innovation, which powers segmentation. A machine learning classifier, such as SVM, Decision Tree, Random Forest, or Naive Bayes, takes control of the segmented data. By combining these processes, the study simplifies the diagnostic procedure and improves cancer classification accuracy and reliability. They found that their work provided better accuracy, Jaccard coefficient, sensitivity, and specificity than conventional architectures. Various machine learning architectures benefit from the proposed method because its segmentation accuracy is 99.12%. Due to this, our methodology has proved very beneficial when applied to the medical field.

Maria et al. [83] proposed hyper-resolution machine learning and FPGA techniques for real-time BI-RADS breast cancer classifier deployment. The second most common cancer in women worldwide is breast cancer. It is important to develop better diagnostic tools since surgical, radiation, and medication treatments have high success rates, especially when detected early. The study uses a custom Digital Mammogram Diagnostic Convolutional Neural Network (DMD-CNN) model to categorize mammogram breast lesions, focusing on BI-RADS criteria. This work uses a PYNQ-based Artix 7 FPGA to accelerate the DMD-CNN model, a first in breast cancer diagnostics. This model is more accurate than existing methods by 98.2%. The current model has an accuracy increase of 4% and a recognition rate of 96% over the current model. Using k-fold cross-validation as well as extensive mammography dataset testing, the model's efficiency was further confirmed. In addition, the FPGA hardware acceleration processed around 91 images per second, significantly faster than GPU and CPU methods, and consumed only 3.12 Watts.

Hekal et al. [84] reported that a deep learning ensemble system is capable of detecting early breast cancer. An ensemble deep learning system is presented in the study for the early detection of breast cancer. Unlike traditional ensemble learning, this system analyzes only suspected nodule regions (SNRs) instead of entire images. In order to detect even small nodules, dynamic thresholding is used; it adjusts according to the image's specifics. Four Convolutional Neural Networks (CNNs) are used in this ensemble: AlexNet, ResNet-50, ResNet-101, and DenseNet-201. In order to determine whether an image is benign or malignant, a binary Support Vector Machine (SVM) is applied after each CNN. Final decisions are made based on the ensemble CNN's outputs and the four CNN's training accuracy. Tests on the public CBIS-DDSM dataset showed that the system was able to categorize malignant and benign mass nodules with a 94% accuracy. Ensemble systems showed superior accuracy and benefits over other methods using the same dataset.

Ortega Martorell et al. [85] characterized and visualized breast cancer patients using deep learning. Through the use of Fisher Information Networks, a deep learning model was used to extract features from mammograms for a deeper understanding of these features. It provides a visual representation of mammogram features that help organize patients based on their similarities, opening the door to a "patient-like-me" approach. By comparing new patient data to existing cases with similar characteristics, clinicians can offer more personalized diagnoses. The visual analysis of the CBIS-DDSM dataset demonstrated significant diagnostic potential. In addition to facilitating a more intuitive understanding for clinicians, this method increases breast cancer diagnosis accuracy. The result is more precise clinical decisions made based on comparisons with closely matched existing patient data, thus reducing overdiagnosis.

Shimokawa et al. [86] reported that stromal invasion in digital breast tomosynthesis could be predicted using a deep learning model. This study used digital breast tomography (DBT) to develop a DL algorithm for predicting stromal invasion. Initially, 499 patients with suspected breast cancer underwent DBT between March 1 and August 31, 2019, ranging in age from 29 to 90 years. A further analysis was conducted on 140 patients who later underwent surgery for breast cancer. Patients were classified into two groups based on pathological reports: non-invasive cancer (20 patients) and invasive cancer (120 patients). Four deep learning models were used to define non-invasive and invasive cancers, namely VGG16, Resnet50, DenseNet121, and Xception. AUC was used to measure the diagnostic capability of the models. Accordingly, the AUC values for the four models were 0.56, 0.67, 0.71, and 0.75. The DL model can predict stromal invasion in breast cancer patients using DBT images.

Mohapatra et al. [87] evaluated deep learning models based on histopathological mammogram images. Using a convolutional neural network (CNN), the authors investigated how breast cancer can be detected from mammograms. A number of CNN architectures were evaluated, including AlexNet, VGG16, and ResNet50. A few of these models used transfer learning with pre-trained weights, while others were trained from scratch. Using the mini-DDSM dataset, a freely accessible resource, the classifiers were trained and tested. Medical datasets have limited samples, so overfitting is possible. To address this problem, rotation, and zooming techniques were applied. The validation strategy adopted a 90:10 split. To conclude, CNNs, especially with transfer learning, can be used to identify benign, cancerous, and normal mammogram images. Nevertheless, training approaches and architectures can have significant effects on performance.

Shimokawa et al. [88] proposed an imaging deep learning model to detect breast cancer using bilateral asymmetrical detection (BilAD). Bilateral asymmetry detection (BilAD) is a deep learning method that uses asymmetry in breast tissue to detect breast cancer. It involved 115 breast cancer patients with pathological confirmation. The original images of each patient were paired with those of normal areas and malignant tumors. In order to distinguish between these image pairs, the BilAD model was modified from the Xception convolutional neural network (CNN). The BilAD model, based on the asymmetry of bilateral breast tissue, is more effective than a unilateral CNN model (uCNN) at detecting breast cancer.

6.1.1 Datasets for mammographic imaging

This section provides an overview of several pivotal mammographic imaging datasets used to detect and diagnose breast cancer. There is a wide range of datasets, different image quality, and detailed annotations to choose from, making them an invaluable resource for developing and validating diagnostic algorithms, as well as developing medical imaging technology. A description of prominent mammography datasets is presented here, along with the highest levels of accuracy achieved by algorithms derived from these datasets.

- In the **Digital Database for Screening Mammography (DDSM)** [89], digitized film-screen mammograms are displayed with ground truth and other information. This resource provides a large set of mammograms in digital format for researchers to evaluate and compare the performance of computer-aided detection (CAD) algorithms. The database was completed in the fall of 1999. It contains 2620 mammography screening exams with four views. New software tools have been added since then, simplifying the extraction of image data to other file formats, enabling easier access to the ground truth data, and simplifying the evaluation of CAD algorithms. A retrospective evaluation of cancer cases is often the starting point for the evaluation of a CAD algorithm. This type of preliminary evaluation saves time and money compared to a prospective clinical evaluation. It is possible to digitize mammograms from a mammography center's case files and use those for a retrospective CAD performance evaluation, but this task is time-consuming and expensive. The DDSM data can be used to avoid this expense. It is possible to learn much more about CAD methods by evaluating performance when investigators use a common database rather than their own data. It is possible to compare the relative strengths of different algorithms by using the same data, performance measures, and test methodologies. Thus, new or combined approaches to the problem may be developed, resulting in superior results. Mammograms were obtained from Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital, and Washington University of St. Louis School of Medicine for the DDSM. In each case, four different standard views (medio-latral oblique, cranio caudal, and lateral) were digitized. Each case was assigned to a volume based on its severity. The normal volumes contain mammograms from screening exams that were reported as normal and had a normal screening exam four years later. This volume contains exams that had a noteworthy abnormality but did not require additional work-up. These benign volumes contain cases in which something suspicious was found and the patient was recalled for additional testing that proved benign. Histologically proven cancer cases are included in cancer volumes. The volumes may contain cases with less severe findings as well as cases with more severe findings, depending on the volume to which the case is assigned. An expert radiologist specified the ACR breast density for every case in the DDSM, including the patient's age, the screening exam date, and the date when the mammograms were digitized. All volumes other than the normal volume contain pixel-level abnormality markings. Mammograms have been automatically cropped to remove much of the background (non-breast tissue) after digitization. As a next step, they were manually processed to darken (digitally zero) pixels in regions that contained patient identifiers, and then stored in truly lossless compressed files. Ground truth data was entered into computer-readable format using custom software that allowed one to draw free-form digital curves of the radiologist-identified ground truth regions. The highest accuracy achieved by algorithms in this review utilizing this dataset is 98.44%.

- The **CBIS-DDSM (Curated Breast Imaging Subset of DDSM)** [90] is a refined version of the Digital Database for Screening Mammography (DDSM), created to enhance the evaluation of computer-aided detection (CADe) and diagnosis (CADx) systems in breast cancer research. By providing a standardized and comprehensive resource for medical imaging researchers, this dataset addresses the shortcomings of the original DDSM. There are 753 cases with calcifications and 891 cases with masses, which total 2,620 scanned film mammograms. In order to facilitate various analytical tasks, the images were decompressed and re-annotated by trained mammographers to ensure accurate region-of-interest (ROI) segmentations. The CBIS-DDSM preserves all the original information by converting images from an obsolete lossless JPEG format to 16-bit grayscale TIFF and then to DICOM. This dataset features precise ROI segmentation, especially for mass cases, achieved through a modified local level set framework, improving the accuracy of CADe and CADx algorithms. In addition, 339 images with unclear or incorrect annotations were removed in order to maintain the dataset's quality and reliability. BI-RADS categories are used to divide the dataset into training and testing sets for comprehensive evaluation. Additionally, the CBIS-DDSM includes extensive metadata in CSV format, including patient identification, breast density, view type, mass characteristics, calcification details, BI-RADS assessment, pathology, and subtlety rating. A more accurate diagnostic tool can be developed using this metadata. As a result of the dataset, researchers are able to analyze both full mammography images and cropped images of abnormalities. In order to ensure compatibility with modern systems and to increase accessibility to the dataset, the tools used for image processing have been updated and made available on GitHub. In order to validate the dataset, hand-drawn outlines were compared with automated segmentations by an experienced radiologist, demonstrating a significant improvement in segmentation accuracy over the DDSM annotations. During the quality control process, researchers ensured that the dataset was reliable, providing a reliable resource for developing new methods for detecting and diagnosing breast cancer. By providing a meticulously curated and standardized dataset, the CBIS-DDSM offers a significant advancement in mammography research. The highest accuracy achieved by algorithms in this review utilizing this dataset is 98.32%.

- The **Mammographic Image Analysis Society (MIAS)** [91] consists of 322 digitized images with original resolutions of 1024 × 1024 pixels. MIAS is a UK-based research group. The dataset contains 207 normal images, and 115 aberrant images. Several sections of the dataset are divided into training and testing sections. Specifically, 31% of the samples (100 images) are used for training, and 69% (222 images) are used for testing, with no data augmentation applied. The MIAS database also includes ground truth annotations for the abnormalities that appear in the images, which is an invaluable resource for developing diagnostic algorithms. The highest accuracy achieved by algorithms in this review utilizing this dataset is 99.17%.
- The **INbreast** [92] contains full-field digital mammograms (FFDM), created specifically to support the development of computer-aided detection (CAD) and diagnosis systems in breast cancer research. Annotations will be provided to make the dataset both useful for clinical practice and research. Using a MammoNovation Siemens FFDM machine with an amorphous selenium solid-state detector with a pixel size of 70 microns and a contrast resolution of 14 bits, the dataset was collected at CHSJ's Breast Center in Porto. DICOM format ensures the images and associated metadata are preserved, while all personal information is anonymized. INbreast contains 115 cases and 410 images. In 90 cases, there are two views per breast (craniocaudal and mediolateral oblique), while 25 cases involve women who had a mastectomy. Mammograms with normal findings, masses, calcifications, architectural distortions, and asymmetries are included in the dataset. In orthogonal views, masses appear as three-dimensional structures with convex borders, while calcifications are classified as benign (larger and smoother) or malignant (small and finer). Mammographic distortions appear as star-shaped interruptions in the normal pattern, and breast asymmetries reflect differences in volume or parenchyma. OsiriX software is used to annotate the dataset meticulously, performed by a specialist and validated by another. Data on lesion types, locations, and contours are stored in XML files. Each XML file contains information such as the number of regions of interest, the area, and center coordinates. The MLO view includes annotation types such as asymmetries, calcifications, microcalcification clusters, masses, distortions, and spiculated regions. For CAD system evaluation, this thorough annotation process ensures high-quality ground truth data. The highest accuracy achieved by algorithms in this review utilizing this dataset is 91.84%.

The Mammographic Image Analysis Society (MIAS) database is the most commonly used dataset in mammography imaging research. Figure 12 provides a detailed overview of the process involved in creating this dataset. The following section describes various models and their performance results using this dataset.

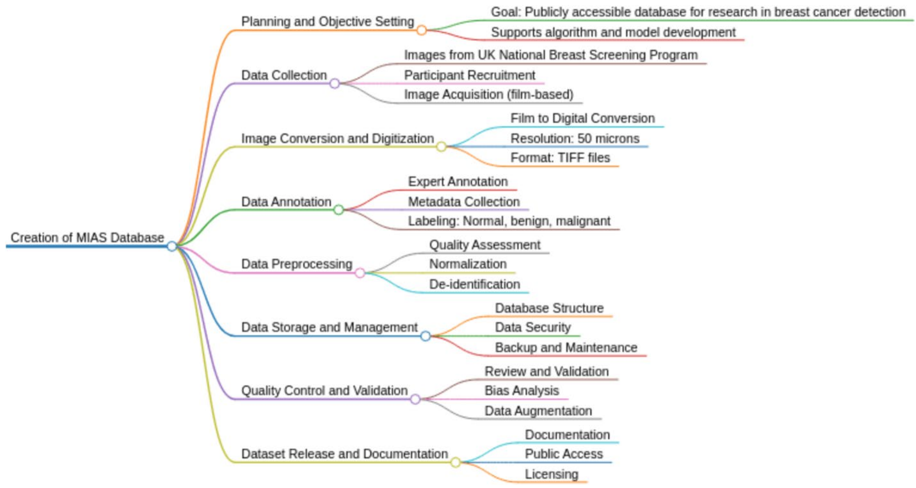


Fig. 12 Showing the process of generating the MIAS dataset

6.1.2 Model accuracy comparison on the MIAS dataset

State-of-the-art (SOTA) models have been used for mammographic imaging for breast cancer detection using the MIAS dataset, demonstrating a range of accuracy. From traditional machine learning techniques to sophisticated deep learning architectures, these models span a wide range of capabilities. Hybrid models that incorporate advanced techniques and deep learning architectures have shown superior performance in handling the complexity of mammographic images. Raaj [80] developed a hybrid CNN with Radon transform that achieved 99.17% accuracy. The Radon transform captures essential features from mammograms, which, combined with the deep learning capabilities of the CNN, allows robust feature extraction and classification. Similarly, the GoogleLeNet architecture combined with machine learning classifiers like SVM, Decision Tree, Random Forest, or Naive Bayes, as developed by Ramesh et al. [82], also achieves high accuracy (99.12%), underscoring the effectiveness of deep feature extraction followed by traditional classification methods. Alloqmani et al. [75] proposed MobileNetV2 with pre-trained feature extraction and a single-layer perceptron that achieved 91.76% accuracy. However, MobileNetV2's single-layer perceptron may not be sufficient to capture the complex patterns found in mammography images, resulting in lower performance. Chakravarthy et al. [76] proposed the CSOA-wKNN, which has an accuracy of 74.78%. In addition, traditional algorithms cannot handle the intricacies of medical image data, even when combined with other methods. Therefore, sophisticated feature extraction and classification techniques are critical to achieve high accuracy in mammographic imaging as a result of the performance disparity among these models. For breast cancer detection using the MIAS dataset, state-of-the-art hybrid models and deep learning architectures are essential to capturing complex patterns and nuances. Figure 13 shows the different model accuracies for the MIAS dataset.

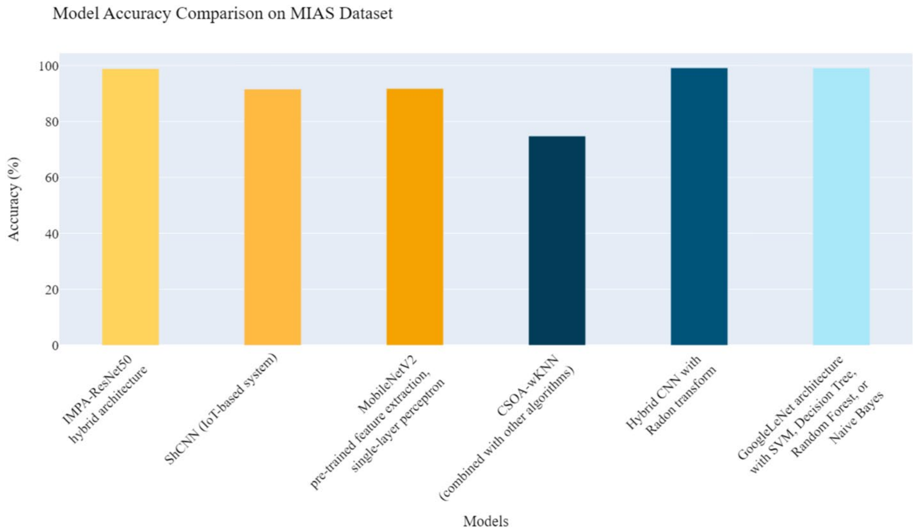


Fig. 13 Comparison of model accuracy on the MIAS dataset

6.1.3 Model accuracy comparison on the CBIS-DDSM dataset

With the CBIS-DDSM dataset, several state-of-the-art models demonstrate diverse performance levels, underscoring the complexity and varying effectiveness of different deep learning architectures. Houssein et al. [69] developed a hybrid architecture combining IMPA with ResNet50, which achieves an impressive accuracy of 98.32%. Feature extraction and representation are enhanced by ResNet50's integration with IMPA since it mitigates the vanishing gradient problem and offers robustness against the vanishing gradient problem. Ensemble models that combine VGG16, InceptionV3, and VGG19 demonstrate 98.1% accuracy. Nemade et al. [71] proposed that the strength of this ensemble approach lies in the complementability of each individual model: InceptionV3 is renowned for its ability to capture hierarchical patterns thanks to its deep yet straightforward architecture, while VGG16 and VGG19 are renowned for their deep but straightforward architectures. According to Hekal et al. [84], the ensemble of AlexNet, ResNet-50, ResNet-101, and DenseNet-201 records an accuracy of 94%. The inclusion of advanced networks such as DenseNet-201, which improve gradient flow and reuses features extensively, might dilute the overall performance, resulting in suboptimal integration of learned features, despite the inclusion of AlexNet, an older and less sophisticated model. Other models like the Fisher Information Networks CNN, as proposed by Ortega Martorell et al. [85], and HRDO-FL + E-RNN, as proposed by Kumbhare et al. [81], report accuracies of 89.6% and 95.73%, respectively. Although Fisher Information Networks' CNN may be effective in specific contexts, it may not be able to capture the intricate patterns necessary for high-accuracy mammography. As a result, HRDO-FL + E-RNN performs better but still falls short of the top performers, and this may be a consequence of its limited ability to deal with spatial dependencies or to integrate sequential information. In summary, the varying performance of these models emphasizes the importance of advanced architectures such as ResNet and efficient ensemble techniques in achieving superior accuracy in mammography imaging, whereas older

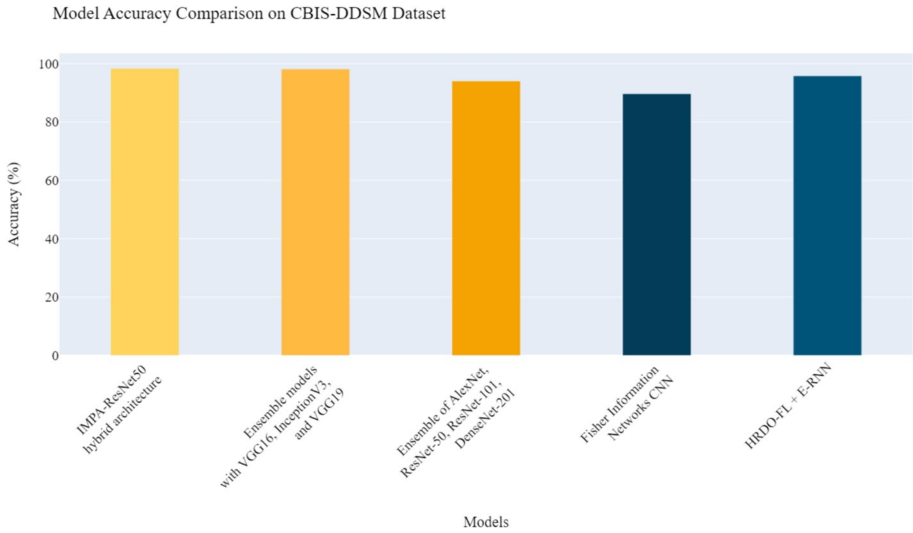


Fig. 14 Comparison of model accuracy on the CBIS-DDSM dataset

models or those with less optimized feature extraction capabilities tend to lag behind. Figure 14 shows the different model accuracies for the CBIS-DDSM dataset.

6.2 Ultrasound imaging

It emphasizes the use of deep learning techniques in ultrasound imaging as a method of detecting breast cancer. Several studies have demonstrated how deep learning can improve the accuracy and efficiency of breast cancer detection using ultrasound imaging by harnessing the capabilities of deep learning. As Table 6 represents Wang et al. [93] proposed using a deep learning network incorporating an automatic segmentation network to diagnose breast cancer by automated breast ultrasound. In this study, 769 breast tumors were enrolled in training and test sets at 600 versus 169. With new DLNs (Resent v2, ResNet50 v2, ResNet101 v2), morphological information was extracted from breast tumors by adding a novel ASN to the traditional ResNet networks. It calculated accuracy, sensitivity, specificity, positive and negative predictive values, area under the receiver operating characteristic curve (AUC), and average precision (AP). Two radiologists with differing experience levels were compared with novel DLNs. In order to obtain morphological information on breast lesions, the researchers developed a new segmentation network. To create novel DLNs, the new segmentation network was added to the traditional classification network. Breast cancer detection by ABUS benefited from DLNs with novel diagnostic characteristics as measured by AUC and AP.

Yu et al. [94] developed a model for predicting neoadjuvant chemotherapy response using deep learning radiomics. Six hundred and three patients who had NAC between January 2018 and June 2021 across three different institutions were retrospectively included in the study. The trained DCNNs were validated in testing cohorts (n = 183), which n refers to the number of patients, using ultrasound images taken before and after treatment. The image-only model structure was found to perform the best based on predicted performance. Additionally, independent clinical-pathologic variables were used to construct

Table 6 Revolutionizing Ultrasound-based breast cancer detection with deep learning methods

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[93]	collected	224×224	DLNs (ResNet34 v2 and ResNet50 v2 and ResNet101 v2) with ASN	DLNs trained on manually marked breast lesion volumes, ASN for shape information extraction	Breast Lesion Diagnosis through ABUS	Accuracy: 78.11%, Specificity: 76.81%, Sensitivity: 85.00%, AUC: 0.85, AP: 0.90, F1: 0.82
[94]	Gathered from 3 institutions Ultrasound images	224×224	Deep learning radiomics (ResNet 34 and ResNet 50 and VGG16 and DenseNet 101) with SVM	Feature extraction and prediction of pathological response to NAC in breast cancer, incorporating clinical parameters and Grad-CAMs	Prediction of Pathological Response to NAC in Breast Cancer Patients	ResNet50: AUC 0.879, Accuracy 82.5%; Integrated DLR Model: AUC 0.962 (training), 0.939 (validation)
[95]	BUSI dataset (Breast Ultrasound Images Dataset) and Breast Ultrasound Image dataset	64×64	HMB-DLGAHA (CNN models and Transfer Learning (TL) and Genetic Algorithms (GA))	Hybrid approach combining CNNs with Genetic Algorithms for hyperparameter optimization	Breast Cancer Classification using Ultrasound Images	Loss: 0.2101, Accuracy: 92.80%, F1-Score: 0.9293, Precision: 92.55%, Recall: 96.71%, Specificity: 0.9869, AUC: 85.17%
[96]	Collected	256×256	YOLOv3-tiny	Ground Truth, Feature extraction & classification, Model training, Detection, ROI identification, Image processing	the effectiveness of a deep learning-based computer-aided detection (CADe) system for breast ultrasound	AUC: 0.7726, Sensitivity: 95.4%, Specificity: 86.6%, PPV: 78.1%, NPV: 97.4%, FPC: 0.43
[97]	Collected	224×244	CNN (VGG-based architecture)	Pre-processing, VGG-19, Fully connected layers, Sigmoid activation	the development of a deep-learning system for the automatic identification of triple-negative breast cancer (TNBC) solely from ultrasound images	AUC: 0.86, Accuracy: 85%, Sensitivity: 86%, Specificity: 86%, F1-score: 0.74

Table 6 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[98]	Not Reported	512×512	Residual networks and Residual attention networks	Preprocessing, Feature and predictive networks demonstration, Residual module, Attention module	Predictive model using QUS and machine learning for personalizing treatment in locally advanced breast cancer	Accuracy: 88%, Specificity: 92.5%, Sensitivity: 70%, Loss: 0.16, AUC-ROC: 0.86 (95% confidence interval)
[99]	Collected	99×99	CNN-LSTM	Semi-Automated Bimodal CAD for Breast Cancer Detection	Detection and Classification of Breast Cancer	Unimodal-Mammogram: Acc: 97.29%, Se: 93.54%, Sp: 100%, AUC: 0.96, MCC: 94.54% (before feature selection), Acc: 97.16%, Se: 93.23%, Sp: 100%, AUC: 0.96, MCC: 94.29% (after feature selection); Unimodal-Ultrasound: Acc: 98.58%, Se: 96.62%, Sp: 100%, AUC: 0.98, MCC: 97.13% (before), Acc: 98.84%, Se: 97.23%, Sp: 100%, AUC: 0.98, MCC: 97.64% (after); Dual modality approach: Acc: 99.35%, Se: 98.46%, Sp: 100%, AUC: 0.99, MCC: 98.68% (both before and after feature selection)

Table 6 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[100]	BUSI and UDIAT (Ultrasound Diagnosis of Abnormalities in Tissues)	224×224	BTEC-Net (Classification) and RFS-UNet (Segmentation)	Image Resizing, Classification (BTEC-Net), Segmentation (RFS-UNet)	Improve the Accuracy and Reliability of Breast Tumor Segmentation in Ultrasound Images	Accuracy: 99.487%, Recall: 99.846%, Precision: 99.538%, F1 Score: 99.692%
[101]	Collected	256×256	DeepRec subsystem	Automated Diagnostic System with R-S-A Procedure, including DeepRec, DeepCls, and DeepAti	Automated Diagnosis of Breast Cancer	Accuracy: 94.51%, Sensitivity: 97.20%, Specificity: 94.89%, AUC: 0.982
[102]	Collected	Nan	VGGNet-based CNN	Fog computing-based system for breast cancer detection from ultrasound images, utilizing edge, fog, and cloud layers with DL models	Early Detection of Breast Cancer Masses from Ultrasound Images	Training Accuracy: 98.5%; Specificity: 0.9783; Sensitivity: 1.0; Precision, recall, f1 score, and support for normal are 0.96, 1.0, 0.98, and 23 respectively. Precision, recall, f1 score, and support for malignant are 1.0, 0.98, 0.99, and 46 respectively
[103]	Collected	Nan	Resnet18 and Attention Model	Multi-modal DLRP with Co-Attention Module	Classification of Breast Cancers into Luminal and Non-Luminal Subtypes	AUC: 0.929 (internal validation), 0.900 (external test set). Sensitivity: 0.940 (internal), 0.892 (external). Specificity: 0.643 (internal), 0.812 (external). PPV: 0.949–0.957. NPV: ~0.600

an integrated DLR model. The DeLong method was used to compare the areas under the curve of these models and the AUCs of two radiologists. DLR was used to evaluate the therapy response to NAC in advanced breast cancer patients and the resulting model was validated using independent multicenter data. In this way, the response to NAC can be evaluated early and therapy planning can be guided accordingly. A DLR-based prediction model for NAC response was constructed based on pretreatment US imaging and significant clinical factors. The DLR model performed exceptionally well in a validation set. Integrated DLR models can also predict NAC efficacy in advance and adjust therapy regimens in time for people who may not respond well to NAC.

Balaha et al. [95] reported that a hybrid deep learning and genetic algorithms approach (HMB-DLGAHA) is suitable for early breast cancer detection. This study will introduce a comprehensive hybrid methodology for learning, detecting, classifying, and identifying breast ultrasound images. A combination of Convolutional Neural Network (CNN) models, Transfer Learning (TL), and Genetic Algorithms (GA) is used in this innovative approach. CNN models play a pivotal role in both the learning and parameter optimization phases of this framework. Alternatively, genetic algorithms can be utilized to optimize hyperparameters by fine-tuning them. The authors aim (1) to simplify the proposed 'HMB1-BUSI' CNN architecture in comparison with pre-trained CNNs such as VGG19 and ResNet, and (2) to demonstrate its generalization to various datasets.

Fujioka et al. [96] proposed using a deep learning-based computer system to detect breast cancer using ultrasound. In this study, a deep learning-based computer-aided detection system was tested for breast ultrasound accuracy. Furthermore, a real-time YOLOv3-tiny model was trained to detect lesions using an expanded set of training images. Twenty-eight readers evaluated test images with and without CADe. Results of the study showed that the CADe system significantly improved lesion detection, with an area under the curve (AUC) of 0.7726 compared to 0.6314 without it. CADe's deep learning-based approach improves readers' interpretation of breast ultrasounds, increasing the accuracy of breast cancer screening and diagnosis.

Boulenger et al. [97] proposed an automated method for triple negative breast cancer detection from ultrasound images based on deep learning. The purpose of this study was to develop an automated system that could identify triple-negative breast cancer (TNBC) from ultrasound images by utilizing deep learning. An analysis of 145 patients and 831 images from Peking Union College Hospital was conducted over a period of one year. Immunohistochemical (IHC) results were used for determining cellular subtypes. In order to predict TNBC, a CNN with a VGG-based architecture was used. This approach is unique in that it is non-invasive and automated. This method does not rely on biopsy information or manually defined features like traditional methods. Clinicians can use this system to devise treatment plans and assess prognoses. Modeling results showed AUC of 0.86, 85% accuracy, 86% sensitivity, 86% specificity, and an F1-score of 0.74. Different molecular subtype groups showed distinct differences in the model's features, indicating more accurate treatment selection.

Taleghamar et al. [98] proposed using multi parametric ultrasound images to predict chemotherapy response to breast cancer. Quantitative ultrasound (QUS) multi-parametric images were analyzed before treatment to predict how breast cancer responds to neoadjuvant chemotherapy (NAC). QUS images of breast tumors were generated using data from 181 patients scheduled for NAC and subsequent surgery. NAC's effectiveness was evaluated post-surgery using standard clinical and pathological assessments. To extract optimal feature maps from these images, two deep convolutional neural network (DCNN) structures were explored: residual networks and residual attention networks. It is possible to extract these features from both the tumor core and its margin. Based on RAN architecture, the model extracted features from both tumor cores and margins with 88% accuracy. In a decade-long survival analysis, responders and non-responders were found to have significant survival differences based both on their pre-treatment and post-treatment predictions. DCNNs that are attention-focused may provide early indications of breast cancer treatment response.

Atrey et al. [99] proposed the use of a hybrid deep learning approach to classify breast cancer using mammography and ultrasound. This study investigates the feasibility of combining mammography and ultrasound to enhance breast cancer (BC) diagnosis, two tools commonly used together by radiologists. It introduces a bimodal Computer-Assisted Diagnosis (CAD) algorithm based on CNN-LSTM deep learning. While the Convolutional Neural Network (CNN) is used for feature extraction, the Long Short-Term Memory (LSTM) handles classification. The most relevant features are retained by applying a statistical significance analysis to the extracted features. An ultrasound and mammography dataset were used to test the proposed method in real-time. AUC and accuracy of this bimodal CAD approach are better than those of unimodal systems. Furthermore, the bimodal system performed better than the Support Vector Machine (SVM) classifier. Through a bimodal DL approach, the study demonstrates the benefits of integrating mammography and ultrasound features in BC diagnosis.

Cho et al. [100] proposed a method based on deep learning to segment ultrasound images for breast cancer diagnosis. An image segmentation network was used to identify tumor regions in breast ultrasound images by the authors using a multistage segmentation algorithm. The network consists of two main stages: classification and segmentation. To classify input images as normal or abnormal, the BTEC-Net model was used. The fusion of two subnetworks led to the creation of a feature network. The pixel values of the normal image were all set to zero, significantly reducing false positive (FP) errors. When images are classified as abnormal, the RFS-UNet model is used to segment the tumor regions. BUSI and UDIAT datasets were used to test the method's efficacy. A comparison of the proposed method with traditional techniques showed that the proposed method was more accurate than the former. Moreover, the Grad-CAM analysis confirmed the model's focus on the relevant breast region for accurate segmentation, with relevant features extracted.

Qi et al. [101] suggested using deep learning to detect breast cancer in ultrasound images. Using ultrasonography images for breast cancer diagnosis, the authors introduced an automated system intended to improve the efficiency and reliability of breast cancer screenings. Mobile phones are used in this innovative system to detect malignant changes in ultrasound reports. Three subsystems are based on deep neural networks. It utilized three extensive annotated ultrasonography breast image datasets to train and assess its performance. As a result of its similar performance to human experts, the system shows real-world clinical potential. Nevertheless, it remains difficult to interpret.

According to Welhenge [102], deep learning-based fog computing can detect breast cancer. VGGNet-based convolutional neural networks (CNNs) are used in this study for accurate cancer detection from ultrasound images. A three-tiered architecture is used to solve deep learning's data storage problems: edge devices capture data, fog devices process it, and cloud devices store the results. As a result of this approach, not only is diagnosis accuracy improved, but data management is also improved. A VGGNet-based CNN was used in this study to achieve 98.5% training accuracy in detecting breast cancer.

According to Huang et al. [103], deep learning radiopathomics can distinguish luminal and non-luminal tumors in breast cancers at an early stage. This multicentre study examined data from three cohorts collected between January 2019 and August 2021 from a prospective study (ChiCTR1900027497). A total of 1809 ultrasound imaging and 603 H&E-stained whole slide images were collected from 603 patients with early-stage breast cancer for the study. For US images, a Resnet18 model pre-trained on ImageNet was used. For WSI images, an attention model based on multi-instance learning was used. UCA is a unique US-guided Co-Attention module that fuses US and WSI features. The Deep Learning Radiomics-Pathomics (DLRP) model was built using 1467 US images and 489 WSIs from 489 patients. As part of the model, three sets of features were incorporated: the deep learning US feature, the deep learning WSIs feature, and the UCA-fused feature. To test the diagnostic abilities of the DLRP model, 342 US images and 114 WSIs from 114 patients were validated, followed by 270 US images and 90 WSIs from 90 patients. Additionally, the study compared the diagnostic efficacy of DLRP against deep learning radiomics and deep learning pathomics models.

6.2.1 Datasets for ultrasound imaging

This section provides an overview of several pivotal ultrasound imaging datasets used to detect and diagnose breast cancer. There is a wide range of datasets, different image quality, and detailed annotations to choose from, making them an invaluable resource for developing and validating diagnostic algorithms, as well as developing medical imaging technology. A description of prominent ultrasound datasets is presented here, along with the highest levels of accuracy achieved by algorithms derived from these datasets.

- The Breast Ultrasound Images (BUSI) [104] dataset contains a comprehensive collection of breast ultrasound images intended to facilitate classification, detection, and segmentation of breast cancer. This dataset was compiled from 600 female patients in Cairo, Egypt, who visited Baheya Hospital for Early Detection & Treatment of Women's Cancer in 2018. After preprocessing to remove unimportant information, the dataset contains 780 images from an original collection of 1100 images. There are 133, 487, and 210 images in each of these classes, which are categorized into normal, benign, and malignant. These images were captured using LOGIQ E9 ultrasound systems, which are known for their high resolution of 1280×1024 pixels and their capability in radiology, cardiac, and vascular applications. Transducers used in the imaging process operated at frequencies between 1 and 5 MHz on a Matrix linear probe ML6-15-D. These images were initially stored in DICOM format, but were later converted to PNG format using a DICOM converter application. Deep learning models can be trained on images with an average size of 500×500 pixels. Data preprocessing was crucial to ensuring its quality and utility. This process involved removing duplicate images and cropping images to remove unused boundaries, which could interfere with classification. An expert team of radiologists from Baheya Hospital meticulously reviewed and corrected any incorrect annotations. For segmentation tasks, the refined dataset was organized into folders according to the three categories, with each image corresponding to a specific category, and a ground truth mask image for each category. Additionally, DICOM images were converted to PNG format and image names were annotated. Machine learning tasks were enhanced by ground truth annotations. The boundary of breast masses was delineated using Matlab freehand segmentation. To signify its purpose, each image has a corresponding mask image whose name is appended with "_mask". In addition, the detailed annotation process ensures that the dataset can be effectively used to train models to detect and segment breast cancer lesions with high accuracy. For

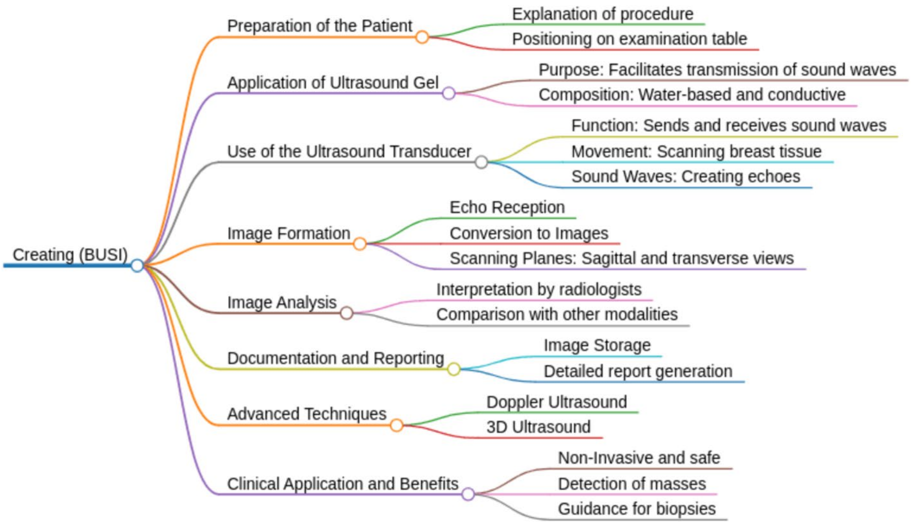


Fig. 15 Showing the process of generating the BUSI dataset

this research, patient confidentiality and informed consent were strictly adhered to. The highest accuracy achieved by algorithms in this review utilizing this dataset is 99.487%.

The most famous used dataset in ultrasound imaging research is the Breast Ultrasound Images (BUSI) database. Figure 15 provides a detailed overview of the process involved in creating this dataset.

6.3 MRI and CT imaging

The purpose of this subsection is to explore the application of MRI (Magnetic Resonance Imaging) and CT (Computerized Tomography) imaging for breast cancer detection, particularly the effect of deep learning methodologies on the diagnostic process for breast cancer detection. According to research in this area, advanced imaging techniques may improve accuracy and precision in breast cancer diagnosis. As Table 7 signifies Guo and colleagues [105] used MRI-based deep learning radiomics to identify low HER2 HER2 status in breast cancer patients in order to predict disease-free survival. Two institutions retrospectively recruited 481 breast cancer patients who underwent preoperative MRI. Radiomic features and DSFR features were extracted from segmented tumors separately to construct models. After averaging the output probabilities of both models, the DLR model was constructed to assess HER2 status. To analyze disease-free survival (DFS) in HER2-low-positive patients, a Kaplan–Meier survival analysis was performed. DFS was further investigated using a multivariate Cox proportional hazard model. DLR models showed AUCs of 0.868 for training cohorts and 0.763 for validation cohorts for HER2-positive and HER2-overexpressing patients, respectively. This study aims to develop a noninvasive MRI-based DLR method for the assessment of HER2-positive breast cancer and to investigate how prediction scores affect the prognosis of patients. DLR predicted disease-free survival in patients with tumors with different HER2 expression, and the prediction result was a significant independent predictor of DFS. In both the training and validation cohorts,

Table 7 Elevating breast cancer detection in MRI and CT scan imaging through deep learning innovations

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[105]	collected	Nan	MRI-based Deep Learning Radiomics (DLR)	DLR model for predicting HER2 status using radiomics and deep learning	Prediction of HER2 Status in Breast Cancer Patients	DLR model performance: HER2-negative vs. HER2-overexpressing (AUCs: 0.868 training, 0.763 validation), HER2-low-positive vs. HER2-zero (AUCs: 0.855 training, 0.750 validation). Significant predictors of DFS: DLR score (HR, 0.175; $p=0.024$) and lesion size (HR, 1.043; $p=0.009$)
[106]	collected	256×256	ResNet18-based Deep Learning Model with ITR and PTR features	Feature extraction from intra-tumoral (ITR) and peri-tumoral (PTR) regions, multi-input ResNet18 model, feature fusion	Classification of Lung Lesions (BCLM and PLC) using LDCT Images (PLC) using LDCT images	AUROC: 0.913, Accuracy: 0.920, Sensitivity: 0.825, Specificity: 0.883
[107]	collected	224×224	DLRN (radiomics nomogram)	ROI segmentation, Feature extraction, Model development: Radiomics signature, deep learning signature, clinical data integration	the prediction of axillary lymph node metastasis (ALNM) in breast cancer patients using a deep learning radiomics nomogram (DLRN)	AUC: 0.893, SEN: 0.853, SPE: 0.808, ACC: 0.833, PPV: 0.852, NPV: 0.808

Table 7 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[108]	collected	512×512	RetinaNet-based network	Training with Chest CT axial images, bounding boxes marked around breast cancers, and image augmentation	Detection of Breast Cancer on Chest CT Scans	Internal Test Set: 96.5% detection rate, 13.5 FPs/case. External Test Set: 96.1% detection rate, 15.6 FPs/case. Sensitivities: 0.3 threshold—92.0% (Internal), 93.0% (External); 0.4 threshold—88.5% (Internal), 90.7% (External). Lesion Size-Based Sensitivities: ≤ 2 cm—92.0% (Internal), 91.5% (External); > 2 to ≤ 5 cm—100% (Internal), 98.3% (External); > 5 cm—100% in both sets
[109]	collected	Nan	nn-Unet with SGD Classifiers	Pipeline for processing CT scans, including autocontouring of normal tissues, laterality classification, autocontouring of target volumes, and lumpectomy/mastectomy classification	Classification of Laterality and Lumpectomy/Mastectomy in Breast Cancer Patients	AUC: 0.93; Accuracy: 93.75%
[110]	collected	80×80	CNN DL model for DCE MRI and diffusion-weighted imaging	Preprocessing, Development of 3D DL network using baseline and post-NAST MRI scans, Incorporating PEI maps and b800 DWI	Predicting pathologic complete response (pCR) to neoadjuvant systemic therapy (NAST) in TNBC using DL on multiparametric MRI	AUC: 0.86 ± 0.03, Accuracy: $77 \pm 4\%$, Sensitivity: $77 \pm 10\%$, Specificity: $77 \pm 10\%$, PPV: $73 \pm 8\%$, NPV: $82 \pm 6\%$

DLR demonstrated high and stable accuracy in predicting HER2 status. Using the DLR model, radiologists and clinicians will be able to better identify HER2-low-positive breast cancers and guide patient treatment accordingly.

Li et al. [106] developed a method for predicting lung metastases using imaging characteristics associated with intratumoral and peri-tumoral tissue. This study retrospectively collected LDCT images of 100 breast cancer patients with lung lesions. In the study, 60 cases of BCLM and 40 cases of PLC were included. This study combined residual convolutions with traditional radiomics features by using a ResNet18-based deep learning model. Using a multi-region strategy, this model incorporated both ITR and PTR features. Through random assignment of training and validation sets, they conducted three-fold cross-validation. With deep learning and radiomics combined with ITRs and PTRs, the model aimed to improve BCLM and PLC diagnoses. In order to improve classification, they integrated radiomics features and deep learning features into a feature fusion model. BCLM and PLC can be diagnosed by considering both intra-tumoral and peri-tumoral image features. AUC could be significantly improved by combining radiomics features with deep learning features. With this model, precision medicine can be achieved through accurate classification of BCLM and PLC, as well as assistance with clinical diagnosis.

Zhang et al. [107] predicted axillary lymph node metastasis using a deep learning radiomics nomogram derived from multiphase computed tomography. DLRN is designed to predict axillary lymph node metastasis (ALNM) in breast cancer patients. Invasive breast cancer patients with nonspecific symptoms were studied. A non-linear support vector machine was used to create radiomics and deep learning signatures for CT scans. DLRN was evaluated for performance and used independent predictors. Based on the results, radiomics signatures, deep learning signatures, and clinical N stage, which N stands for Node, were all independent predictors. With a receiver operator characteristic curve area of 0.893, DLRN accurately predicted ALNM in the validation set. Also, the DLRN demonstrated higher clinical utility than other predictors. Consequently, the DLRN provides valuable insight into individualized breast cancer treatment.

Koh et al. [108] reported that deep learning is helpful for detecting breast cancer on chest computed tomography. In this study, a deep learning algorithm was assessed and validated specifically for detecting breast cancers on chest CT scans. This study investigated the performance of the RetinaNet-based network, a deep learning algorithm previously demonstrated to detect breast masses on mammography, on chest CT images in order to detect breast cancer. Previous studies have shown that RetinaNet rivals traditional two-stage object detectors in mammography. Besides being evaluated for its ability to detect breast cancer on chest CT scans, a deep learning model was also tested for its robustness to external and internal validations. By using these methods, the researchers sought to determine the comprehensive applicability and accuracy of RetinaNet. Retrospective analysis included 1170 preoperative chest CT scans taken after breast cancer diagnosis. Three categories of scans were conducted: algorithm development (1070), internal testing (100), and external testing (100). These datasets were used to train and evaluate a deep learning model based on the RetinaNet framework. This study demonstrated that deep learning algorithms have the potential to detect breast cancer early on chest CT scans across two test sets.

Bargalló et al. [109] classified laterality and mastectomy/lumpectomy for breast cancer patients to improve deep learning auto segmentation performance. The study offers a methodical pipeline for CT imaging of breast cancer that uses a trained nn-Unet model to auto-contour standard tissue structures. Through auto-contouring, it determines whether a patient needs a lumpectomy or mastectomy. Classifiers using stochastic gradient descent consistently outperformed Random Forest Classifiers. Due to the classification's high accuracy, precise network contouring can be achieved over generic models, reducing target delineation inaccuracies. Furthermore, the research highlights the importance of considering unique anatomies or procedures when classifying institutions. Despite the proposed system's effectiveness in classifying breast cancer patients, there is still room for improvement in workflow elements. It can effectively distinguish between patients with breast cancer on left or right sides, as well as between those who had lumpectomies versus mastectomies. In both cases, the Random Forest algorithm was found to be the most optimal. In the primary test set, the Random Forest classifier was able to discriminate between breast and lumpectomy surgeries, and it also had an AUC of 9.93 in the secondary sample of 16 patients, with a 93.75% accuracy rate. The proposed method showed high accuracy based on laterality and surgical procedure, the proposed method showed high accuracy and efficiency in classifying breast cancer patients.

Zhou et al. [110] reported that deep learning can be used to predict neoadjuvant systemic therapy for triple negative breast cancer. DL was investigated to predict whether triple-negative breast cancer (TNBC) patients would show pathological complete response (pCR). In this procedure, dynamic contrast enhanced MRI (DCE) and diffusion weighted imaging are used at the early stage of neo-adjuvant systemic therapy (NAST). DL model was meticulously trained using 130 images from TNBC patients. Based on the areas under the receiver operating characteristic curves (AUCs) for both the training and validation phases, it showed commendable performance. The AUC of this model was consistently high despite rigorous testing on an independent set of 32 patients. In a prospective blinded test involving 48 patients, these results were further confirmed. These results suggest that DL models have promising capabilities when applied to multiparametric MRI data, especially. In the early stages of NAST, such models can differentiate TNBC patients with pCR from those without.

6.4 Histopathological and microscopic imaging

Deep learning is applied to analyzing histopathological and microscopic imaging for breast cancer detection in this subsection. The combination of advanced imaging techniques with deep learning offers detailed insights at the cellular level to help diagnose breast cancer more accurately and efficiently. Deep learning algorithms have successfully been applied to interpret and analyze these complex images in pivotal studies.

As Table 8 signifies Taheri and Golrizkhatami [111] proposed magnification-dependent and magnification-dependent classification of breast cancer histopathological images. Annually, breast cancer causes numerous deaths. Detecting malignant breast cancer can be simplified through automated computer methods that reduce pathologists' workload and diagnoses more accurately. Histopathological images of breast cancer are used in this study to introduce two methods for diagnosing the disease. Based on specific magnification factors, the first system uses pre-trained DenseNet201 CNN models, fine-tuned on the BreakHis dataset. In the second system, there are four subsystems, each customized for a specific magnification. Final diagnosis results are then combined from these subsystems. The proposed methods demonstrate superiority over existing techniques when tested on the BreakHis dataset. Histopathological breast images can be classified using CNN architectures originally designed for color objects. In comparison to all other state-of-the-art methods, the single magnification-based (MSB) system outperformed all others using DenseNet201 technology. To classify patients, a multi-magnification-based (MIB) system used predictions from four magnification-specific DenseNet201 models.

Table 8 Transforming breast cancer diagnosis in Histopathological and Microscopic imaging using deep learning approaches

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[111]	BreakHis dataset (7909 microscopic biopsy images and Four magnification factors of 40×, 100×, 200×, and 400×	Nan	DenseNet201 CNN and magnification-specific subsystems	Pre-processing (Data augmentation, Stain normalization), Transfer learning, Score-level fusion, Magnification-specific binary classification (MSB) using transfer learning technique, Magnification-independent binary classification (MIB) by score-level fusion of four MSB subsystems	The targeted problem is the automated detection and classification of benign and malignant breast cancer from histopathological images	Accuracy (ILA/PLA): 40×(93.76% ± 2.32 / 94.71% ± 0.88), 100×(92.04% ± 3.41 / 95.9% ± 4.2), 200×(91.31% ± 2.95 / 96.7% ± 1.09), 400×(89.81% ± 2.19 / 89.0% ± 0.12)
[112]	BreaKHis	752×582	MTRRE-Net (Multi-scale Two-fold Residual Recurrent Network)	Use of multi-scale two-fold residual recurrent groups (MTRRGs) for feature extraction, including dual residual blocks, max-pooling layers, and convolutions with different kernel sizes	Breast cancer identification from histopathological images	40×Magnification: Accuracy: 97.12%, Precision: 97%, Recall: 96%, F1 Score: 96%; 100×Magnification: Accuracy: 95.22%, Precision: 94%, Recall: 95%, F1 Score: 97%; 200×Magnification: Accuracy: 96.85%, Precision: 96%, Recall: 98%, F1 Score: 94%; 400×Magnification: Accuracy: 97.81%, Precision: 96%, Recall: 94%, F1 Score: 95%

Table 8 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[113]	BreaKHis and IDC (Invasive Ductal Carcinoma) and UCSB (Bio-Segmentation Benchmark dataset)	227 × 227	AlexNet-BC	Data Augmentation, Transfer Learning, New Loss Function	Breast Cancer Classification	Accuracy (BreaKHis: 40 × Magnification: 98.15 ± 0.9%, 100 × Magnification: 97.71 ± 1.9%, 200 × Magnification: 97.96 ± 0.7%, 400 × Magnification: 98.48 ± 1.1%, IDC: 86.31 ± 1.7%, and UCSB: 96.10 ± 0.8%)
[114]	BreaKHis	Varied (128 × 128, 64 × 64, 32 × 32, etc.)	CNN and DenseNet121	Use of CNN and DenseNet121 models with data augmentation (scaling, rotation, histogram normalization) and transfer learning. Experimentation with various image sizes and rotations to optimize performance	Automatic Classification of Breast Histopathological Images	CNN Test Accuracy: (128 × 128) 60°: 40X—71.25%, 100X—76.07%, 200X—70.79%, 400X—67.75%; (64 × 64) 90°: 40X—65%, 100X—86%, 200X—82.67%, 400X—80.87%; (32 × 32) 180°: 40X—68.12%, 100X—85.16%, 200X—84.65%, 400X—79.23%. DenseNet121 Test Accuracy: (224 × 224) 60°: 100X—77.11%, 200X—77.32%; (128 × 128) 90°: 100X—88.03%, 200X—71.03%; (64 × 64) 180°: 100X—62.9%, 200X—67.81%

Table 8 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[115]	collected	500×500	Multiclass CNN	THG Microscopy + CNN for Breast Tissue Cancer Grading	Automated Breast Tissue Classification by Grade	Mean Validation Accuracy: ~ 79.4%. Precision: 0.84, 0.78, 0.78, 0.81 (for each class respectively). Recall: 0.92, 0.85, 0.77, 0.68. F1-Score: 0.88, 0.81, 0.77, 0.74 (for each class respectively)
[116]	collected	The patches of size (50×50)	RANN-BCC Model based on ResNet	Neural network architecture (RANN-BCC) with attention mechanisms for breast cancer classification	Identification and Classification of IDC in Breast Cancer	Accuracy: 92.45%, Recall: 0.98, Precision: 0.91, F-score: 0.94
[117]	collected	224×224	DL model for grading histopathology slides CNN	Patient selection, Histopathological assessment, Data acquisition, Pre-processing, Multiple Instance Learning (MIL), Deep Learning Model Development, Multi-Task Learning (MTL), Evaluation, Survival analysis	Development of a DL-based solution for automated breast cancer grading, aiming to improve objectivity, reproducibility, reduce inter-observer variation, and explore impact on patient survival analysis and treatment decisions	80% accuracy (Cohen's Kappa score of 0.59)
[118]	collected	64×64	Convolutional Transfer Learning and Regional Attention	Pre-processing (Gaussian Filtering), ACCNN Based Segmentation, Classification (CTL-RAM), Attention Mechanism	Breast Cancer Detection (Localization and Classification)	Classification accuracy: 96%, Detection accuracy: 92%, MAP: 82%, Sensitivity: 92%, Specificity: 91%, RMSE: 70%

Table 8 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[119]	collected	224 × 224	MobileNetV2 and U-Net	Color Normalization, Nucleus Extraction, Combination of MobileNetV2 and U-Net	Recognition of Cancerous Regions in Pathology Slides	H&E Stained: Pinpointing Accuracy 89.6%, RGB Model Accuracy 97.6%, CD Model Accuracy 83.9%, CDACS Model Accuracy 89.6%, Fluorescent Stained: Pinpointing Accuracy 80.5%, RGB Model Accuracy 50.4%, CD Model Accuracy 72.5%, CDACS Model Accuracy 80.5%
[120]	BATCH (ICIAR 2018 Grand Challenge on Breast Cancer Histology)	512 × 512	Fusion of DCNNs (VGG16 and VGG19 and Xception and InceptionV3 and InceptionResnetV2)	Image Preprocessing, Feature Extraction using DCNN Architecture, Classification, Confidence Matrix, Fusion	Accuracy Improvement in Breast Cancer Histology Classification	Overall Accuracy: 96%, Precision: 96%, Recall: 96%
[121]	TCGA (The Cancer Genome Atlas)	512 × 512	CNN	Image Patching, Filtering, Color Normalization, Feature Extraction (ResNet50), Fusion with MCB	Prediction of Recurrence and Metastasis in Breast Cancer	AUC: 0.72
[122]	TCGA-BRCA (TCGA breast cancer subset)	224 × 224	Multimodal Deep Learning	Hybrid Model: DNN for gene modality, CNN for image modality, Fusion	Breast Cancer Subtype Prediction Using Multimodal Data	Prediction Accuracy: 88.07%, Average AUC: 0.94277

Table 8 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[123]	collected	512×512	CNN	HER2 Status Classification and Trastuzumab Response Prediction using transfer learning, data augmentation, color normalization, and deep generative models for stain-color normalization	Classifying HER2 Status and Predicting Trastuzumab Response	AUC of 0.90 for HER2 Status Prediction; AUC of 0.80 for Trastuzumab Response
[124]	collected	950×950	Multitask CNN (ResNet21 backbone)	ResNet21 CNN Backbone, Global Average Pooling, Fully Connected Layer for Risk Score Prediction, Guan-Rank Transformation, Multitask Learning	Breast Cancer Prognosis, BCSS Prediction, ER and ERBB2 Status Prediction	Accuracy for BCSS: 0.66
[125]	collected	512×512	CNN	MITNET-det for nuclei detection and MITNET-rec for mitosis recognition	Mitosis Classification in breast cancer tissue	Precision: 75.4%, Recall: 38.1%, F1-score: 49%
[126]	collected	224×224	CNN	Multimodal Imaging (BFMI, AFMI, OPMI) + Image Fusion (PLF, DLF) + Grad-CAM	Classification of Breast Tissue as Cancerous or Normal based on Imaging	Accuracy: 89.01%, Sensitivity: 90.59%, Specificity: 86.93%, AUC: 0.9366

Chattopadhyay et al. [112] described MTRRE-Net as a deep learning model for detecting breast cancer in histopathological images. In the MTRRE-Net framework, a multi-scale two-fold residual recurrent network (MTRRE-Net) is used to detect breast cancer using histopathology images. BreakHis is a dataset known for having intricate appearances that complicate straightforward classification. It emphasizes local features to discern intricate patterns to improve classification accuracy. It uses residual skip connections, multi-scale learning of image features, and residual recurrent connections. Feature delivery is enhanced by these elements, and residual skip connections integrate shallow and intricate features. As a result of its unique two-fold residual recurrent operation, the model addresses the degradation problem in deep architectures. With its multi-scaling operation, it excels at extracting significant features from small datasets, enhancing classification accuracy. Tests on the BreaKHis dataset showed impressive accuracy across a wide range of magnification levels. In comparison with other pre-trained models, such as GoogLeNet, DenseNet169, and VGG16, it outperformed them all. A comparison with state-of-the-art models further validated its effectiveness.

Liu et al. [113] reported that deep learning methods can be applied to pathology images to classify breast cancer. AlexNet-BC is a new breast cancer classification framework built on the foundational AlexNet model. Using the BreaKHis dataset and optimizing the original AlexNet network structure to compensate for the limitations of small-scale breast pathological images datasets. A novel low-entropy output penalty is introduced to counteract overfitting caused by softmax-cross-entropy learning. A penalty is applied when the predicted likelihood probability exceeds a threshold. Pre-trained and fine-tuned with a transfer learning approach, enhanced neural networks demonstrate superior performance. Additionally, IDC and UCSB datasets are used to validate the model's robustness. In spite of its high accuracy, the model does not segment the lesion area, which will be the subject of future research, integrating deep learning semantic segmentation.

Muntean and Chowkkar [114] proposed breast cancer detection from histopathological images using deep learning and transfer learning. Researchers compared CNN and DenseNet121 models for classifying breast histopathological images as malignant or benign. The impact of image magnification, scaling, and rotation on model accuracy was studied using 7,909 images. This study examines the performance of two deep learning models in comparison to one another: a Convolutional Neural Network (CNN) that was built from scratch with parameter tuning and a DenseNet121, which uses transfer learning to automatically categorize breast histopathological images. This study investigates how image magnification, scaling, and rotation affect the model's accuracy using 7909 histopathological images. It consists of convolutional neural networks, pooling neural networks, and fully connected neural networks, with a learning rate of 0.0001 and ReLU as the activation function. On the other hand, DenseNet121, with its 121 layers, utilizes ImageNet's pretrained weights and uses sigmoid activation functions. On Google Colab Notebook, Keras and TensorFlow libraries were used to implement both models. Data was sourced from Google Drive. The purpose of the research is to analyze the accuracy of these models in terms of test and training accuracy, emphasizing the importance of increased accuracy and true negative values in medical diagnosis.

As a result of scaling the images from 128×128 to 100X, DenseNet121 was able to achieve its highest accuracy (86.6% at 100X magnification). The CNN model, on the other hand, achieved a 90.9% accuracy rate under different conditions, namely 100X magnification, 64×64 image size scaling, and 180-degree rotation. Transfer learning led to a 16.4% increase in accuracy at 100X magnification during training. Study findings concluded that the DenseNet121 model accurately detected breast cancer in histopathological images, and future studies could explore other transfer learning models.

Tsafas et al. [115] employed a deep-learning technique to detect breast cancer using non-linear images from human tissue biopsies. This study explores the use of raw THG images to identify breast biopsy tissue types. There is no additional processing or staining required to highlight multilayered structures in membranes. By leveraging deep learning, the research attempted to differentiate benign from malignant breast tissues and assess disease severity using THG images from unlabeled samples. Breast cancer intratumoral heterogeneity and multiregional classification were examined in this study. It enables precise cancer grading (benign, grades I-III) using its multiclass CNN model. In order to determine the best treatment strategy for a patient, an accurate diagnosis and knowledge of his or her cancer grade is essential. The non-destructive and label-free properties of optical methods may be useful in future histopathology. As a result of this approach, fresh, unstained biopsy samples can be characterized quickly. The technique could increase cancer detection rates and shorten biopsy characterization time, which would improve histopathology. With this advance, patients might be able to undergo less unnecessary biopsies and enjoy a higher quality of life. THG imaging combined with recent advances in laser scanning microscopes paves the way for digital pathology to benefit both patients and healthcare systems.

Toa et al. [116] proposed a deep residual learning method using attention algorithms to classify breast cancers. This paper applies deep learning methods to both non-IDC and IDC classification. Several sample images are available for training deep learning models, which makes them ideal for processing medical images. Residual attention neural network breast cancer classification (RANN-BCC) is a model designed to assist physicians in rapidly and efficiently analyzing breast cancer images. Breast cancer diagnosis can be speeded up with RANN-BCC, which uses residual neural networks (ResNet) to support classification. In addition to comparing the results of the RANN-BCC model with other deep learning models, a classification test was conducted using a dataset of non-IDC and IDC images. Using residual attention neural networks breast cancer classification (RANN-BCC), they identify invasive ductal carcinoma (IDC) and non-invasive ductal carcinoma (non-IDC) in the given breast cancer dataset. Using the RANN-BCC model, they demonstrate that their model outperforms other deep learning models. As a result of integrating self-attention, cross-attention, collectors, and compressors into Residual Neural Network 34 (ResNet34), the accuracy increased from 79.49 to 92.45%. The authors claim this integrative approach will not only benefit medical practitioners but also contribute to computer-aided diagnosis.

Wetstein et al. [117] reported that breast cancers can be graded and rated with deep learning. A deep learning model was developed to grade histopathology slides using whole-slide images of breast cancer. 706 females between the ages of 20 and 40 were used to train the algorithm, which included tumor grade, tubule formation, and mitotic rate. Expert pathologist annotations were used as a benchmark to compare the model's performance with 686 patients. The Cohen's Kappa score of 0.59 indicates an 80% accuracy rate in distinguishing between low/intermediate and high grades of tumors. In terms of overall survival (OS), disease/recurrence-free survival (DRFS/RFS), and survival for both groups, the model's predictions revealed significant differences. Cox regression analysis of the univariate data further confirmed these findings. While molecular subtype stratification and clinicopathologic features were taken into account, a trend persisted. The deep learning model needs further refinement to better predict survival of breast cancer patients based on whole-slide images.

Sheeba et al. [118] proposed microscopic image analysis for breast cancer detection based on ensemble deep learning architectures. In this study, ensemble deep learning techniques are integrated with feature extraction and classification techniques to detect breast cancer. It encompasses both historical data and newly acquired microscopic images, gathered from the web of things (WoT). Initially, the input image is preprocessed. A Gaussian filter is used to eliminate noise, followed by an active contour convolutional neural network to segment the results. This process uses a combination of convoluted transfer learning and regional attention. Localization-centric cancer classification techniques are superior to existing methodologies in this field. Based on simulation results, the proposed method of cancer classification based on localization is superior. According to this study, its average classification accuracy stands at 96%, detection accuracy at 92%, mean average precision at 82%, sensitivity at 92%, specificity at 91%, and relative mean square error at 70%.

Huang et al. [119] found that deep learning can detect breast cancer through cross-staining histopathology images. The study used color normalization and nucleus extraction techniques. Furthermore, a workflow for analyzing fluorescent nucleic acid-stained images was introduced that uses an AI model designed to analyze H&E-stained segmentation. The accuracy of pinpointing H&E-stained and fluorescent-stained images was 89.6% and 80.5%, respectively. Using existing pathology AI models for cross-staining inference, the precision of the inference remained the same as the proposed workflow. This study provided a methodology for cross-staining recognition between H&E-stained bright-field and fluorescent-stained dark-field images, allowing for parallel analysis across staining methods. With this innovative approach, current H&E AI models can be extended and more opportunities for fluorescence-based clinical research can be created.

Bhowal et al. [120] proposed fuzzy ensembles of deep learning models for breast cancer histology using cognitive fuzzy integral theory, coalition games, and information theory. A new model for classifying breast cancer histology images outperformed most existing ones. To fuse deep learning models, they used Choquet Integral, which considers decisions from subsets of classifiers, a method not previously employed in this area. A particularly significant contribution was their method of simplifying a typically complex process by using Coalition Games and Information Theory. DCNN architectures like VGG16, VGG19, Xception, InceptionV3, and InceptionResnetV2 were fine-tuned to extract high-level bottleneck features. The features allow for two-class and four-class histology image classifications. Ensemble models, which combine confidence scores from DCNN models, are based on Choquet integrals, Coalition Game Theory, and Information Theory. A recent model failed to outperform their method when tested on the ICIAR 2018 BACH dataset. They achieved 96% accuracy with their fusion method, an improvement of 1% over their best individual models for 2-class problems. The precision and recall of the model also improved significantly after fusion. As a result of deep learning's overconfidence, their model greatly improved accuracy for the 4-class problem, but only marginally for the 2-class problem.

Yang et al. [121] proposed that histopathological images as well as clinical data could be used to predict the risk of HER2-positive breast cancer recurrence and metastasis. Based on histopathological images and clinical data, this study presents a novel application for deep learning. However, this research is still in its early stages. Clinical efficacy tests must be conducted before the model can be widely used for diagnosis and treatment. To evaluate the risk of recurrence in HER2-positive breast cancer patients, a predictive framework was developed based on pathological and clinical data. Images of whole slide sections stained with H&E were initially acquired from surgical specimens. After segmenting the H&E WSIs, they were enlarged to 512×512 pixels, followed by several image processing steps.

Using the CNN algorithm, image features were identified and integrated with the clinical data. Two-fold cross-validation (CV) was performed to confirm the validity of the novel multimodal prognostic prediction model. To measure the model's efficacy, all available HER2-positive breast cancer patients from The Cancer Genome Atlas (TCGA) were used. In order to assess HER2-positive breast cancer patient's risk of relapse and metastasis, imaging images and clinical data can be combined with advanced deep learning models.

An integrated deep learning model has been proposed by Liu et al. [122] for the prediction of molecular subtypes of human breast cancer. An algorithm is used to identify non-invasive breast cancer molecular subtypes. Multimodal deep learning model addresses the limitation of prior studies that relied on single-mode data and lacked adequate feature extraction. By integrating gene and image modal data, this model ensures a comprehensive extraction of deep features. By combining these features, breast cancer molecular subtypes can be predicted intelligently. PCA was used as part of the gene mode preprocessing to expedite the training of the network and to minimize the parameters of the network. The issue of large pixel sizes in pathological images was also addressed by slicing full-scale images. Upon submitting the model to ten iterations of tenfold cross-validation, the prediction accuracy reached 88.07 percent. For each of the four subtypes, an AUC test was conducted, yielding an average AUC of 0.94277.

Farahmand et al. [123] reported that deep learning could predict HER2 status and trastuzumab response in HER2 + breast cancer. In this study, a CNN method is introduced that is more accurate than previous techniques at predicting HER2 status. In cross-validation and independent testing, the classifier obtained an AUC of 0.90 with 188 manually annotated H&E WSIs. Additionally, the classifier aligned well with pathologist annotations. The classifier achieved a 0.80 AUC when trained on samples from 187 HER2 + patients treated with trastuzumab. It is possible to predict HER2 status and trastuzumab response using this H&E-based algorithm, which may support clinical decisions. With the developed method, patients could determine which HER2 drug would benefit them by combining AI with traditional approaches.

Bychkov et al. [124] used outcomes and biomarkers for the prediction of survival in two multinational breast cancer series. Specifically, the study aimed to predict breast cancer survival. These algorithms were trained using 354 TMA samples from the same series. For validation, 674 tumor slides from another multicenter study (FinHer) were used. In addition, a pathologist assessed TMA samples in the FinProg test set visually and then incorporated the results. In spite of factors such as histological grade, tumor size, and axillary lymph node status, the multitask CNN was statistically significant for predicting survival. A pathologist combined deep learning with tissue characteristics visually assessed to improve accuracy to 0.66.

Asyir et al. [125] presented MITNET as a novel dataset for mitosis recognition in breast cancer tissue whole slide images. MITNET is used in this paper to detect nuclei and classify mitoses in breast cancer whole slide images (WSI). Moreover, this paper introduces two new datasets. In the first dataset, 139,124 annotated nuclei are shown in 1749 patches extracted from 115 WSIs containing breast cancer tissue in the first dataset. The second dataset consists of 4908 mitotic and 4908 non-mitotic WSI images, which are used for mitosis classification. With the created datasets, the MITNET network is trained using two deep learning architectures, MITNET-det and MITNET-rec, to isolate nuclei cells and identify mitoses in WSIs. In MITNET-det, nucleus images are extracted from and fused using CSPDarknet and Path Aggregation Network (PANet). To detect nuclei on three different scales, scaled-YOLOv4 is used. Classification involves passing the WSI images through the MITNET-rec deep learning architecture to identify mitosis. MIDOG and ATYPIA datasets are used for training and validation of the created dataset. MIDOG and ATYPIA-based classifiers fail to recognize mitosis on our dataset, which indicates that the mitosis dataset has unique features and characteristics. The overall MITNET framework also detects the nucleus with high detection rates in WSIs, according to the experimental results. A high F1-score allows pathologists to recognize WSI, which improves their accuracy.

Wu et al. [126] reported that multimodal microscopic imaging combined with deep learning could be highly effective in diagnosing breast cancer. This study introduces a novel method of combining microscopical imaging with deep learning. An imaging approach based on bright-field imaging (BFMI), autofluorescence imaging (AFMI), and orthogonal polarization imaging (OPMI) captures comprehensive information about tissue morphology, collagen content, and structure. These details are extremely valuable because they reveal how collagen is distributed and arranged in tissues in relation to tumor progression. Multimodal images are combined at the pixel and decision level in the study to produce fusion classification models. A simulation approach outperforms a single-mode approach.

6.4.1 Datasets for histopathological and microscopic imaging

In this section, a comprehensive overview of several pivotal histopathological and microscopic imaging datasets is offered. Each dataset differs in its range, image quality, and annotations, making them invaluable resources for both developing diagnostic algorithms and furthering medical imaging research. The following is an exploration of prominent histopathological and microscopic datasets, including the highest accuracy levels achieved by algorithms utilizing these datasets.

- The **BreaKHis (Breast Cancer Histopathological Image)** dataset [127] provides microscopic images of benign and malignant breast tumors that can be used to develop and validate automated systems for breast cancer diagnosis. During January 2014 to December 2014, P&D Laboratory in Brazil conducted a clinical study on this dataset. In this study, 82 patients with breast cancer clinical indications were studied. To protect patient privacy, all participants provided informed consent. There are 7,909 images in the dataset, including 2,480 benign and 5,429 malignant samples. The images were captured using an Olympus BX-50 microscope coupled with a Samsung digital color camera, at four different magnifications: 40x, 100x, 200x, and 400x. Using RGB color space and 24-bit color depth, high-quality visual information was obtained. The effective pixel sizes for each magnification are 0.49 µm, 0.20 µm, 0.10 µm, and 0.05 µm, respectively, providing detailed views of the tissue samples. Images are saved as PNG files with dimensions of 700×460 pixels after they have been cropped to remove black borders and text annotations. Several steps were involved in preparing the tissue samples, including fixation, dehydration, clearing, infiltration, embedding, and trimming. The sections were cut to 3 µm thickness and stained with hematoxylin and eosin (HE). To capture images, pathologists selected regions of interest (ROIs) within each slide that had tumorous areas. In each slide, 40X magnification was used to cover the entire ROI, then 100X, 200X, and 400X magnifications were used to cover the remaining ROI. The multi-level imaging approach captures different aspects of the morphology and structure of the tissue. As part of the BreaKHis dataset, images are meticulously categorized into four types of benign tumors (adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma) as well as four types of malignant tumors (ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma). In the dataset, images are distributed across these categories and magnifications, making targeted research and analysis possible. In providing this extensive and well-annotated dataset, the creators hope to foster advances in computer-aided diagnosis (CAD) systems for breast cancer, thereby fostering the development of more accurate and reliable diagnostic tools. The highest accuracy achieved by algorithms in this review utilizing this dataset is 98.48%.

- The **ICIAR Breast Cancer** dataset [128], part of the Grand Challenge on BreAst Cancer Histology images (BACH) organized for the 15th International Conference on Image Analysis and Recognition (ICIAR 2018), serves as a crucial resource for advancing the field of breast cancer histology image analysis. In order to develop algorithms that can accurately identify different kinds of breast tissue, the dataset is designed to support the development of automatic classification algorithms. Breast cancer diagnosis relies on images stained with hematoxylin and eosin (H&E). To aid pathologists in detecting breast cancer, these images are classified into four categories: normal, benign, in situ carcinoma, and invasive carcinoma. There are two main parts to the BACH challenge dataset. The first step consists of analyzing histology microscopy images annotated image-by-image by experts. It contains 400 training and 100 test images representing the four aforementioned classes equally. The images were acquired with a Leica DM 2000 LED microscope and a Leica ICC50 HD camera, collected from patients in the Porto and Castelo Branco regions of Portugal. Three hospitals provided these samples, including Hospital CUF Porto, Centro Hospitalar do Tâmega e Sousa, and Centro Hospitalar Cova da Beira, and the annotation was performed by two medical experts. The second part of the dataset involves pixel-by-pixel labeling of whole-slide breast histology images. An expert reviewer ensures the quality and reliability of these annotations on whole-slide images (WSIs). In this part of the dataset, pixel-level labels are provided that can be used to train and evaluate segmentation algorithms for histopathological image analysis. In the BACH challenge, participants submitted their methods and results for evaluation. The Grand Challenge platform provided researchers with an organized and competitive environment. Convolutional neural networks (CNN) were the most successful approach for automatic classification of breast cancer histology images, achieving an accuracy of 87%. Digital pathology and the automation of breast cancer diagnosis are being encouraged with the availability of this dataset. The highest accuracy achieved by algorithms in this review utilizing this dataset is 96%.

The most commonly used dataset in histopathological and microscopic imaging research is the BreakHis (Breast Cancer Histopathological Image) database. Figure 16 provides a detailed overview of the process involved in creating this dataset. The following section describes various models and their performance results using this dataset.

6.4.2 Model accuracy comparison on the BreaKHis dataset

Several advanced models have demonstrated varying levels of performance across different magnifications (40x, 100x, 200x, and 400x) in the domain of histopathological and microscopic imaging based on the BreaKHis dataset, revealing the complexity and efficiency of different deep learning architectures tailored to different imaging tasks. Taheri and Golrizkhatami [111] proposed the DenseNet201 CNN model for lower magnifications, coupled with magnification-specific subsystems. With its dense connectivity patterns, this

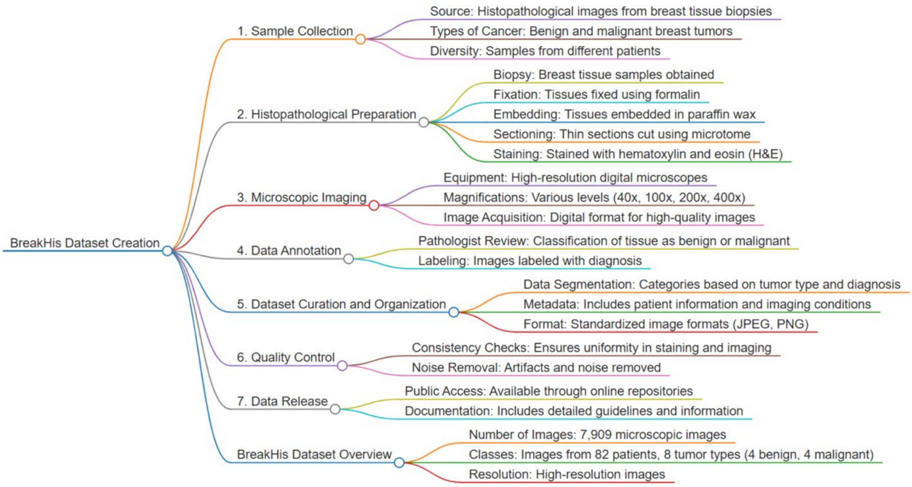


Fig. 16 Showing the process of generating the BreaKHis dataset

model is able to capture detailed patterns effectively, and enhances feature reuse and mitigates the vanishing gradient problem. As a result of the increased complexity and fine details of high-magnification images, performance drops to 89.81% at the highest magnification (400x). Across magnifications, Chattopadhyay et al. [112] reported 97.12% (40x), 95.22% (100x), 96.85% (200x), and 97.81% (400x) accuracy for the MTRRE-Net (Multi-scale Two-fold Residual Recurrent Network). Multiscale analysis and residual and recurrent networks are combined in this model, making it effective for capturing both spatial and sequential information at different magnification levels. Liu et al. [113] developed AlexNet-BC, which stands out with superior performance across all magnifications, achieving accuracies of 98.15% (40x), 97.71% (100x), 97.96% (200x), and 98.48% (400x). AlexNet still performs exceptionally well despite being an older model, possibly due to batch normalization and other techniques. As a result of its high performance, it is a robust and adaptable imaging solution for histopathology. The BreaKHis dataset illustrates the importance of advanced network architectures and specialized subsystems designed for specific magnifications. DenseNet201 performs well at lower magnifications, but less so at higher magnifications. In contrast, MTRRE-Net's consistent performance across magnifications and AlexNet-BC's superior accuracy reflect the effectiveness of multi-scale and residual learning approaches, as well as the potential to achieve high accuracy in histopathological imaging using well-optimized older architectures. Figure 17 shows the different accuracy levels for BreaKHis dataset.

6.5 Thermal imaging

This subsection examines the use of thermal imaging in breast cancer detection, highlighting the role of deep learning in the process. Thermal imaging is a non-invasive and cost-effective method for detecting breast abnormalities. Deep learning techniques combined with thermal imaging have the potential to significantly improve breast cancer detection and diagnosis. The purpose of this paper is to examine key studies that have analyzed thermal images for the detection of breast cancer using deep learning.

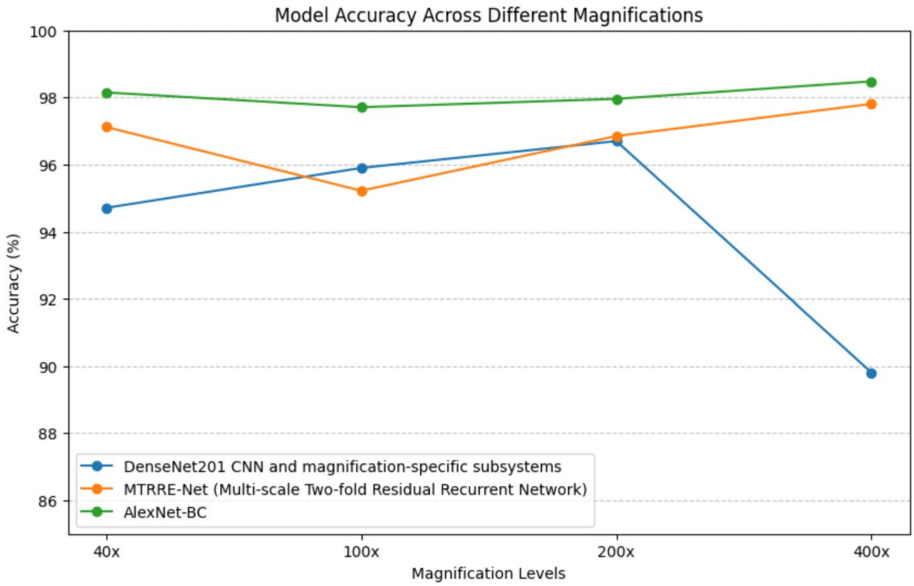


Fig. 17 Comparison of model accuracy on the BreaKHis dataset

As Table 9 represents Ensafi et al. [129] proposed combining multiple views of thermography images together to improve deep learning paradigms for breast cancer diagnosis. Breast cancer is one of the leading causes of death among women worldwide, and the study aims to enhance early detection of this disease through thermography. Thermography is an invasive, non-painless, and cost-effective method for detecting abnormalities of the breast surface. To optimize the performance of preexisting deep learning models, this study integrates multiple views of thermograms, including frontal-45, lateral-45, and lateral45. A comprehensive diagnosis is provided by the model by leveraging transfer learning. Infrared images are obtained from the Database for Mastology Research (DMR). The sensitivity, specificity, and efficiency of thermogram analysis can be improved by 2–15%, 2–30%, and 2–25%, respectively, utilizing transfer-based deep learning. This study presents multiple thermogram views combined with transfer learning in order to develop an innovative method of diagnosing breast cancer. The proposed model has the potential for interpreting breast thermography images in the future as compared to deep learning and handcrafted feature-based methods.

According to Tsietso et al. [130], thermal infrared imaging and clinical data can be combined to develop a multi-input deep learning approach for breast cancer screening. In this study, a multi-input network with transfer learning is used to classify breast thermograms. As part of the network, clinical data and three breast views are incorporated along with automatic Region of Interest (ROI) extractions for symmetry analysis. This model is simulated using a graphical user interface (GUI). This study incorporates transfer learning to reduce computation costs, develops a multi-input network for accurate classification of patients, and adds automatic ROI extraction. With clinical data integration, the study achieved a 90.48% accuracy rate. Besides detecting breast cancer, the proposed model has potential applications in medical radiology.

Table 9 Enhancing breast cancer detection through deep learning in Thermal imaging

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[129]	DMR (Database for Mastology Research)	640×480	Transfer learning with DenseNet121 and EfficientNetB0 and VGG19	Preprocessing, Feature extraction, Classification	Improved deep learning-based breast cancer diagnosis using thermography images	Accuracy: 0.93, Precision: 0.94, Sensitivity: 0.93, Specificity: 0.95
[130]	DMR	224×224 or 227×227	Multi-input network with transfer learning	Multi-input deep learning with automatic ROI extraction, using thermograms and clinical data, Involves data acquisition, preprocessing, feature extraction, and merging of features	Breast Cancer Screening with CADx System	Accuracy: 90.48%, Precision: 93.33, Sensitivity: 93.33, Specificity: 83.33, AUC: 0.94
[131]	DMR-RI (Database for Mastology Research with Infrared Thermal Imaging)	229×229	InceptionV3 and InceptionV4 and MV4	Application of Inception V3, Inception V4, and a modified version called Inception MV4 to classify thermal images of breast tissue. MV4 reduces layers to maintain computational cost while preserving accuracy	Early detection of breast cancer using thermal imaging and deep convolutional neural networks to improve detection accuracy and reduce computational costs	InceptionV4: Accuracy: 99.712%, Sensitivity: 1.000, Specificity: 1.000, Precision: 1.000, Negative Predictive Value: 1.000, AUC: 1.000, EER: 0.000, F1 Score: 1.000 MV4: Accuracy: 99.748%
[132]	DMR-IR	640×480	VGG16 and Dragonfly Grunwald–Letnikov Dragonfly algorithm (GLDA) and SVM classifier	A two-stage framework using VGG16 for feature extraction and GLDA for optimal feature selection	Detection of breast cancer from thermal images, addressing high dimensionality and redundant/irrelevant information	Accuracy: 100%, Precision, Recall, F1-Score: 1.0, 1.0, 1.0

Abdulla Salim Al Husaini et al. [131] found that combining thermal imaging with deep learning models, specifically Inception V3, Inception V4, and a modified Inception MV4, was effective in improving breast cancer detection. To identify breast disorders early using thermal images, this study evaluated these models' performance. The Inception V4 model demonstrated high accuracy when applied to color images, achieving a 100% accuracy rate with the SGDM optimization method and a learning rate of 1×10^{-4} in just 4 epochs. Inception MV4 also demonstrated high accuracy and 7% faster classification response time than Inception V4. These models gained little performance from adding more layers. Inception V3 performed better on grayscale images than its counterparts when more epochs were used when compared to Inception V4 and MV4.

Chatterjee et al. [132] presented a two-stage approach for breast cancer detection using thermographic images. Initially, features are extracted using a deep learning model, VGG16, and then an optimal subset of these features is selected using a memory-based Dragonfly Algorithm (DA) enhanced with the Grunwald–Letnikov (GL) method. Based on testing of the model on the DMR-IR dataset, 100% diagnostic accuracy was achieved with 82% less features than using the VGG16 model alone. Breast cancer detection can be made more accurate and efficient with this method since it filters non-essential features efficiently.

6.5.1 Datasets for thermal imaging

In this section, several pivotal datasets used in breast cancer detection and diagnosis research are analyzed in detail. These datasets are unique in terms of their range, image quality, and detailed annotations, making them invaluable tools for advancing the field of medical imaging and developing diagnostic algorithms. The following is an exploration of prominent thermal datasets, including the highest accuracy levels achieved by algorithms utilizing these datasets.

- The DMR-IR (Database for Mastology Research with Infrared Thermal Imaging) [133] is an open-access database provided by Federal Fluminense University, which aims to support early breast cancer detection. Data included in this database include infrared (IR) images, digitized mammograms, medical history, dietary habits, age, symptoms, and more. Breast cancer research can benefit from a multifaceted approach, including IR images, ultrasound images, and MRI images. Both static and dynamic protocols were employed to capture the IR images with the FLIR-SC 620 imaging camera. In the database, each infrared image measures 640×480 pixels. By providing detailed patient information alongside diverse imaging data, this resource facilitates the development and testing of diagnostic algorithms. The highest accuracy achieved by algorithms in this review utilizing this dataset is 100%.

The most commonly used dataset in thermal imaging research is the DMR-IR (Database for Mastology Research with Infrared Thermal Imaging) database. Figure 18 provides a detailed overview of the process involved in creating this dataset. The following section describes various models and their performance results using this dataset.

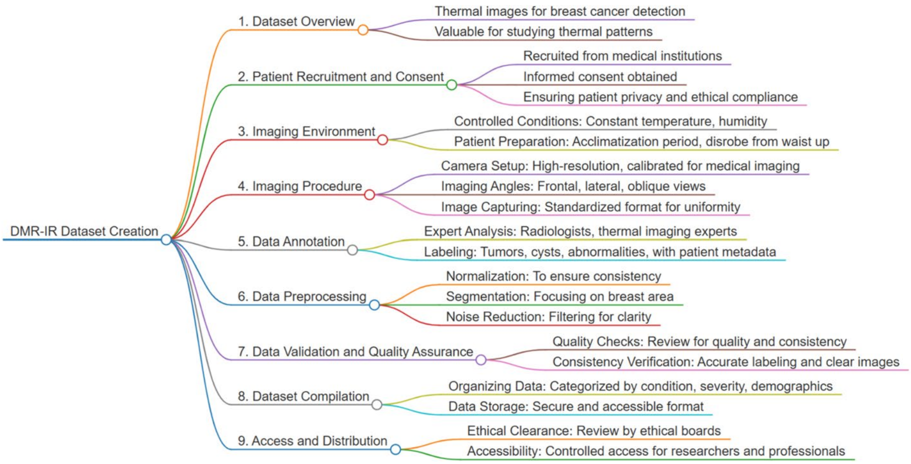


Fig. 18 Showing the process of generating the DMR-IR dataset

6.5.2 Model accuracy comparison on the DMR-IR dataset

Based on the DMR-IR dataset, two state-of-the-art models demonstrate outstanding performance, demonstrating the power of combining deep learning architectures with optimization and classification techniques. The combination of InceptionV3, InceptionV4, and MV4, as proposed by Abdulla Salim Al Husaini et al. [131], achieves an impressive accuracy of 99.748%. Inception architectures are renowned for their powerful feature extraction capabilities. Through their inception modules, InceptionV3 and InceptionV4 handle multi-scale features effectively, which is essential when capturing and processing thermal patterns. As a result of MV4, the model can distinguish subtle details with near-perfect accuracy. Surpassing even this impressive benchmark, the model incorporating VGG16, the Dragonfly Grunwald–Letnikov Dragonfly algorithm (GLDA), and an SVM classifier, as proposed by Chatterjee et al. [132], attains a flawless accuracy of 100%. The deep convolutional layers of VGG16 are effective at capturing hierarchical features, while the GLDA optimizes feature selection. In this way, only the most relevant features are retained, reducing noise and improving the performance of the classifier. With remarkable precision, the SVM classifier, known for its robustness in high-dimensional spaces, then classes these optimized features. This study highlights the importance of combining advanced convolutional neural networks with innovative optimization algorithms and robust classifiers based on the DMR-IR dataset. The near-perfect accuracy of the InceptionV3 + InceptionV4 + MV4 model and the flawless performance of the VGG16 + GLDA + SVM classifier illustrate the potential of these approaches in achieving state-of-the-art results in thermal imaging applications. Feature extraction, optimization, and precise classification techniques are key to the success of these models. Figure 19 shows the different model accuracies for this dataset.

6.6 Others

This subsection examines various innovative deep learning approaches for the detection and classification of breast cancer that do not strictly fall under conventional imaging

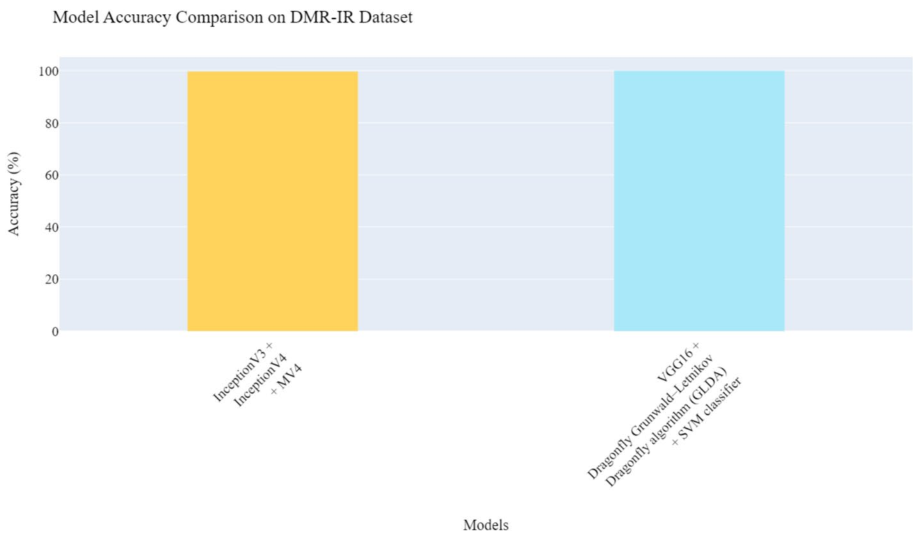


Fig. 19 Comparison of model accuracy on the DMR-IR dataset

techniques. To enhance the accuracy and efficiency of breast cancer diagnosis, these alternative methods utilize deep learning in unique ways, applying it to different types of data or combining it with novel algorithms. These diverse and creative applications of deep learning in breast cancer detection will be reviewed.

As Table 10 represents the classification of breast cancer, Sharma et al. [134] employed a snapshot ensemble deep learning model and a t-distributed stochastic neighbor embedding. The purpose of this study is to analyze historical breast cancer data to detect and predict the disease. To reduce dimension and assemble snapshot models for efficient diagnostics, the authors recommend t-distributed stochastic neighbor embedding (t-SNE). T-SNE enhances scatter plots and optimizes costs. The research uses a snapshot ensemble deep learning framework to combine predictions from multiple base models. From the UCI Machine Repository, Wisconsin Breast Cancer Dataset (WBCD) is used. This model outperformed state-of-the-art models such as averaging, weighted averaging, stacked ensembles, and Polyak Rupert. The proposed model seemed to have great potential for real-world application based on these promising results.

Kayikci and Khoshgoftaar [135] proposed a gated attentive multimodal deep learning approach for the prediction of breast cancer. This paper presents a multimodal deep learning model incorporating clinical, copy number alteration, and gene expression data. Mammography images can be used to detect subtle indications of cancer, and details specific to a patient, such as age and family history, can help refine their predictions. This study aims to bolster breast cancer prediction by using attention mechanisms. By combining multimodal data, breast cancer prognosis is enhanced. Two stages are involved in the process. In the first step, sigmoid gated attention convolutional neural networks are used to create stacked features. The second uses a bimodal attention process. This model may improve detection and outcomes of breast cancer patients. Deep learning and CNNs have made significant contributions to disease diagnosis with the advancements in artificial intelligence. It is possible to achieve classification results that surpass those of human experts without explicitly describing the domain of features. As a result of using deep learning

Table 10 The deep learning-based approaches to breast cancer detection from various studies

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[134]	Wisconsin Breast Cancer Dataset (WBCD)	Nan	Snapshot ensemble deep learning	Data Collection & Preprocessing, t-SNE Dimensionality Reduction, Features Division, Ensemble Deep Learning Model	the detection and prediction of breast cancer using historical data and snapshot ensemble deep learning model and t-distributed stochastic neighbor embedding	Training Accuracy: 86.6%. Performance Metrics: ROC Curve (0.816), Accuracy (0.812), Sensitivity (0.213), Specificity (99.0%), Precision (0.911)
[135]	METABRIC (Molecular Taxonomy of Breast Cancer International Consortium)	Nan	Gated attentive multi-modal deep learning with CNNs	Preprocessing, CNN, Attention, Max pooling, Flatten, Dense layers, Sigmoid activation, Dropout, Combines data from multiple sources	Deep learning model for breast cancer prediction using multimodal approach	AUC: 0.950, Accuracy: 0.912, Precision: 0.841, Sensitivity: 0.798
[136]	WBCD	Nan	Rule-based DL-based model (ANN)	Preprocessing, Feature estimation, Rule assessment, Feature selection + ANN	Breast cancer diagnosis via DL-based model with feature selection	Accuracy: 99.9%, F-Score: 99.0%, Recall: 99.0%, Precision: 98.0%, ROC Curve: 99.0
[137]	collected	Nan	Fine-tuned DL Model with Feature Selection	DL model based on densely-connected artificial neural networks, fine-tuned with feature selection (SULOV algorithm)	Differentiating patients with and without breast cancer based on demographic and anthropometric information, biomarkers, and relative risks related to age and BMI	AUC: 0.9220, TPR: 0.8021, TNR: 0.0780, FNR: 0.1979, FPR: 0.1425, FDR: 0.0983, FOR: 0.0655
[138]	TCGA (Cancer Genome Atlas project) and METABRIC	Nan	Autoencoder and VIPER algorithm	Data Preprocessing, Feature Transformation, Clustering, and Deep Learning	Breast Cancer Subtype Stratification and Prediction	Accuracy: 99.98%, AUC: 0.9663

Table 10 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[139]	collected	36	CNN-based	CNN with 6 hidden layers (128 nodes each), ReLU activation for hidden layers, Softmax for output layer. Data split: 70% training, 30% testing using Scikit-learn's 'train_test_split' with fixed 'random_state'. Loss function guides weight optimization. Training for 240 epochs, batch size of 16. Overfitting mitigation strategies employed. Model evaluation likely includes accuracy, precision, recall, and other metrics	Estimating complete pathological response to chemotherapy (Neoadjuvant Complete Response) in breast cancer patients	Accuracy: 91%, True Negative Ratio: 84%, Matthews Correlation Coefficient: 88%. Precision: 1.0, Recall: 0.84, F1 Score: 0.91
[140]	collected and TCGA (for validation)	256 × 256	Neural network framework for tissue decomposition and prediction CNN	Deep learning-based framework with two CNNs: 1) Tile-level tissue type classifier within WSIs, 2) Prediction of molecular features and TNBC subtypes based on identified tiles, Three-fold cross-validation for training and validation, external validation with TCGA cohort	Prediction of Various Molecular Features, TNBC Subtypes, and Relapse Risk from Pathological WSIs	AUC: 0.84, 0.85, 0.93, 0.73 (for Various Molecular Features), Classification Accuracy for Tissue Types: ~90%, F1 Score: 0.96

Table 10 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[141]	METABRIC	Nan	Heterogeneous Stacking of Multiple DL Models	Multi-modal ensemble classification using CNN and DNN for feature extraction from clinical, gene expression, and CNV data. The approach includes training individual networks, stacking features, and classification using Random Forest. Techniques like dropout, L2 regularization, Adam optimizer, and binary cross-entropy loss function are used to enhance model performance and prevent overfitting	Classification of Breast Cancer Patients based on Multi-modal Data	Accuracy: 97%, Sensitivity: 0.97, Precision: 0.98, F1 Score: 0.98, AUC: 0.92
[142]	METABRIC	Nan	Deep-learning-based stacked ensemble models	Stacked ensemble model using CNNs for feature extraction from multi-modal data (clinical, gene expression, CNA) with subsequent stacking using a machine learning model (e.g., Random Forest)	Breast Cancer Prognosis Prediction	AUC: 0.968, Accuracy: 0.881, Precision: 0.949, Sensitivity: 0.950

Table 10 (continued)

Ref	Dataset	Input size of model	Model	Proposed approach	Targeted problem	Output performance
[143]	ToxCast hit call dataset	Nan	Multi-layer Perceptron (MLP)	Deep learning with MLP neural networks to predict molecular toxicity of EDCs. Data preprocessing includes oversampling for imbalances in ToxCast dataset, Model features an input layer, dense layers, and sigmoidal activator	Prediction of Molecular Toxicity of EDCs Related to Breast Cancer	Accuracy: Between 56.5% to 99.9%

for mammogram-based breast cancer risk assessment, positive results have been obtained. Machine learning and big data can improve screening test accuracy and optimize imaging protocols.

Awotunde et al. [136] suggested combining rule-based feature selection algorithms with deep learning algorithms to detect breast cancer. Breast cancer is one of the leading causes of death among women, and this study focuses on it. Deep learning (DL) and expert systems are increasingly used in breast cancer diagnosis due to the importance of an early and accurate diagnosis. With the help of a hybrid feature selection mechanism based on rule-based DL-based models, the paper presents a model of breast cancer using DL-based features. In this approach, irrelevant features are filtered out to improve diagnostic accuracy. Testing and assessing the efficacy of the model was conducted on the Wisconsin Breast Cancer Dataset (WBCD). With feature selection, the DL model achieved a remarkable diagnostic accuracy of 99.5%. Additionally, it identifies five key diagnostic features. Through the use of diagnostic features that are relevant to breast cancer, the model is superior to current models in predicting the presence of breast cancer.

Martinez and Dongen [137] reported that deep learning algorithms can detect breast cancer at an early stage. Research suggests deep learning can enhance breast cancer screening. Mammograms and ultrasounds are traditional methods of screening for breast cancer, but deep learning models analyze images for abnormalities. Nevertheless, this study extends deep learning beyond image analysis. This article examines how deep learning can be used to detect breast cancer using a diverse set of data. In addition to demographic data, cancer risk information comes from international databases. The study involved 64 women with breast cancer and 52 healthy women. To identify effective prescreening predictors, this study was conducted. The performances of deep learning and traditional machine learning were compared using k-fold Monte Carlo cross-validation experiments. Results showed that a deep learning model fine-tuned with feature selection performed the best at distinguishing between cancerous and non-cancerous patients. Thus, deep learning is able to detect malignancies more accurately than traditional machine learning, reducing the risk of late cancer detection. The performance metrics of deep learning also showed lower prediction uncertainty. A deep learning algorithm may prove more cost-effective and non-invasive than traditional imaging-based cancer screening when used for prescreening. Furthermore, these algorithms can promote self-examination, mitigate the psychological effects associated with false positives, and identify individuals who may need more intensive testing. Additionally, it can reduce healthcare challenges and societal burdens associated with cancer treatment.

Xiong et al. [138] reported that deep learning-based transcription factor activity can be used to stratify breast cancer patients. In this study, transcription factors play an important role in gene expression and DNA epigenetic modification. It is crucial to understand how cancer cells undergo genomic changes. The VIPER algorithm was used in this study to identify the transcription factor activity profiles. The activity profile was compressed using deep learning, specifically an autoencoder, to extract valuable features that can be used to distinguish between two subtypes of breast cancer. In comparison with traditional methods based on transcription factor activity, deep learning displayed enhanced prognostic capabilities. Additionally, the study developed a ceRNA network for the subtypes identified and identified 11 master regulators for each cluster. Multiple breast cancer datasets were used to validate the model's efficacy, emphasizing its potential in prognosis prediction and hinting at its possible use in tailored breast cancer treatments. The accuracy and area under the ROC curve of the deep neural network developed using the derived features were 99.98% and 0.9663, respectively. Tumor immunogenicity and immune infiltration are closely related to these subtypes. Deep learning provides insight into prognosis prediction and potential therapeutic approaches for breast cancer subtypes, as well as improving our understanding of breast cancer subtypes.

Kirelli et al. [139] proposed CNN-based deep learning for predicting the outcome of NAC treatment in breast cancer. There are several convolutional layers that influence model training success, in addition to the quality of the dataset and the dependent variable. In order to evaluate CNN models' performance in classifying pathological data, researchers utilized standard pathological data. CNNs provided robust feature representations for deep learning, leading to accurate predictions. Models are developed that accurately predicted the Miller coefficient, tumor lymph node value, and axillary complete response, with accuracy rates of 87%, 77%, and 91%, respectively. Despite the large and diverse datasets in the study, deep learning methods work well for interpreting pathological results, and assisting in diagnosis, treatment, and prognosis. It suggests that deep learning and machine learning could enhance the interpretation and management of healthcare data.

Using deep learning, Zhao et al. [140] analyzed multiple molecular markers and prognostic factors for triple-negative breast cancer. The authors developed a deep learning model using a multi-omics dataset on TNBC ($N=425$), which N is number of cases, to predict molecular features, subtypes, and prognoses. To decompose tissues on WSIs, a neural network was used, and then a second network was trained on specific tissue types for diverse predictions. Molecular features that were investigated include somatic mutations, copy number changes, germline mutations, metabolic pathways, and immunotherapy biomarkers. All three molecular features predicted by the framework were successful: somatic mutation of PIK3CA, germline mutation of BRCA2, and expression of PD-L1. Moreover, the AUC values for the basal-like immune-suppressed, immunomodulatory, luminal androgen receptor, and mesenchymal-like subtypes of TNBC were 0.84, 0.85, 0.93, and 0.73, respectively. The morphological patterns revealed heterogeneity in TNBC. Moreover, a neural network stratified patients based on image features and clinical data (log-rank $P<0.001$). Models and prediction frameworks were externally validated on 143 cases of TNBC from TCGA ($N=143$), which N is number of cases, and were robust to changes in patient populations. The team created an online platform to deploy and modularize the framework, validate models, and enable real-time predictions for future cases.

Jadoon et al. [141] proposed a multi-modal ensemble classification approach for predicting survival in human breast cancer based on deep learning. This study aims to improve breast cancer prognosis and identifying novel prognostic factors by utilizing ensemble deep learning techniques. Models based on heterogeneous stacking are constructed by extracting features, stacking them, and then classifying them. In addition to using multiple deep learning models for extracting features from various data modalities, this model outperforms existing benchmarks for breast cancer prognosis. This study may contribute to new prognostic tools and treatment strategies for breast cancer patients due to its clinical significance. This model combines individual neural networks to extract features from a variety of data types, such as clinical data, gene expression data, and CNV data, with 97% accuracy. In order to use the model in clinical decision-making, it needs to be further validated and integrated with other sources of information. This flexible approach permits the use of different algorithms and incorporating more data modalities to predict other diseases.

Arya and Saha [142] proposed deep-learning-based stacked ensemble models for multimodal classification of human breast cancer prognosis. This study proposes deep learning-based predictive models to enhance breast cancer prognosis prediction. A convolutional neural network is used for feature extraction in the first stage, and this extracted feature is used as input to the stack-based ensemble model in the second stage. According to various metrics, the model performs better than existing approaches.

Guan et al. [143] identified potential targets of endocrine-disrupting chemicals causing breast cancer using ToxCast and deep learning. A deep learning model was used to predict the molecular toxicity of 47 endocrine-disrupting chemicals (EDCs) linked to breast cancer that had never been tested by the ToxCast program of the US-EPA before. With the help of SMOTE technology, 275 multi-layer perceptron (MLP) neural network models were created, ranging from 56.5% to 99.9% accurate on average. The toxicity of untested EDCs was determined by 142 models that were more than 90% accurate. Six potential targets of breast cancer-related EDCs have been identified, including MYC, PLAUR, and HIF1A.

6.6.1 Datasets

This section provides a detailed overview of several key datasets used in Sect. 6.6. Each dataset is unique in its scope, making them essential resources for developing and testing diagnostic algorithms and advancing the field of medical research. Below are the details of well-known datasets, along with the highest accuracy levels achieved by algorithms utilizing these datasets.

- There are 569 samples in the **Wisconsin Breast Cancer (WBC)** dataset [144], 357 samples representing benign cases and 212 samples representing malignant cases. This dataset contains 11 integer-valued attributes that describe various characteristics of breast cancer cases. It is widely used to determine whether a breast cancer case is benign or malignant by machine learning models, making it a valuable resource for medical data scientists. The highest accuracy achieved by algorithms in this review utilizing this dataset is 99.9%.
- The **METABRIC (Molecular Taxonomy of Breast Cancer International Consortium)** dataset [145] is a significant resource for the study of breast cancer genomics and transcriptomics. It consists of a comprehensive collection of over 2,000 clinically annotated primary fresh-frozen breast cancer specimens from tumor banks in the UK and Canada. The dataset is divided into two main cohorts: a discovery set of 997 primary tumors and a validation set of 995 tumors. These cohorts were carefully assembled to ensure a representative sample of the breast cancer population, facilitating robust genomic and transcriptomic analyses. The primary focus of the METABRIC dataset is to integrate genomic and transcriptomic data to identify novel subgroups of breast cancer with distinct clinical outcomes. By combining copy number and gene expression profiles, researchers aimed to uncover the molecular drivers of breast cancer and to stratify patients into clinically relevant subgroups. The study revealed that 40% of genes were associated with either inherited variants (such as single nucleotide polymorphisms and copy number variants) or acquired somatic copy number aberrations (CNAs), which played a significant role in influencing gene expression. One of the key findings from the METABRIC dataset is the identification of several novel subgroups with distinct genomic characteristics and clinical prognoses. For example, a high-risk subgroup characterized by cis-acting alterations on chromosomes 11q13/14 was identified, which was predominantly composed of estrogen receptor-positive tumors with poor prognosis. Additionally, a subgroup devoid of CNAs was found, which exhibited a favorable prognosis and was associated with an adaptive immune response driven by TCR deletions. These findings highlight the importance of integrating multiple genomic data types to

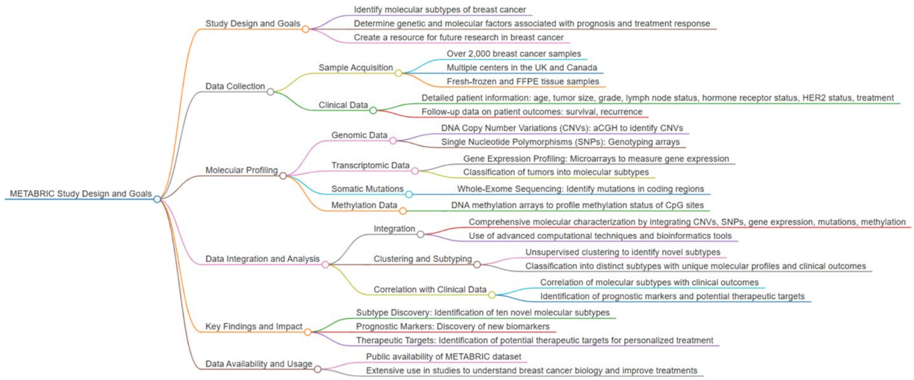


Fig. 20 Showing the process of generating the METABRIC dataset

achieve a more nuanced understanding of breast cancer heterogeneity. The METABRIC dataset also emphasizes the impact of specific genomic regions on breast cancer pathology. For instance, the study highlighted the roles of ZNF703, a driver gene specific to the luminal B subtype, and other significant alterations such as deletions in PPP2R2A, MTAP, and MAP2K4. These genomic insights provide a comprehensive framework for understanding the molecular underpinnings of breast cancer and have implications for developing targeted therapeutic strategies. Overall, the METABRIC dataset serves as a valuable resource for the breast cancer research community, offering detailed molecular portraits that can inform both clinical practice and future research directions. The highest accuracy achieved by algorithms in this review utilizing this dataset is 97%.

The METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset is one of the most frequently used datasets in this category. Figure 20 provides a comprehensive overview of the process involved in creating this dataset. The subsequent section discusses various models and their performance outcomes using this dataset.

6.6.2 Model accuracy comparison on the METABRIC dataset

In the classification and prognosis of breast cancer using the METABRIC dataset, state-of-the-art (SOTA) models excel by integrating multimodal data via sophisticated deep learning architectures. With 91.2% accuracy, Kayikci and Khoshgoftaar [135] proposed a gated attention multimodal deep learning model. It utilizes gated attention mechanisms and convolutional neural networks (CNNs) to dynamically focus on the most relevant features and capture complex spatial patterns in the data, resulting in a high level of effectiveness when processing diverse genomic and clinical inputs. Using multiple deep learning models, Jadoon et al. [141] achieved a remarkable accuracy rate of 97%. In this method, more than one DL model is trained on different aspects of the dataset, and their outputs are combined to leverage the strengths of each, resulting in highly generalized and accurate predictions. Arya and Saha [142] proposed Deep-Learning-Based Stacked Ensemble Models that reduce variance and bias by stacking multiple deep learning models in a hierarchical manner to achieve 88.1% accuracy, but their lower accuracy indicates a need to improve model integration or feature selection in order to match the higher performance of other SOTA methods. Based on the complexity of the data and the sophistication of the models used,

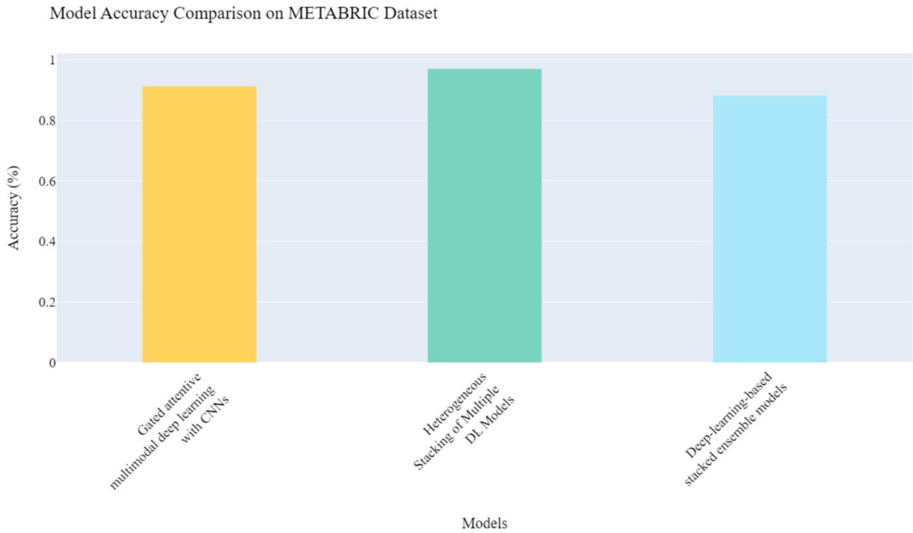


Fig. 21 Comparison of model accuracy on the METABRIC dataset

the performance of machine learning models differs greatly on the METABRIC dataset. With an accuracy of 91.2%, gated attentive multimodal deep learning effectively integrates diverse data types, emphasizing relevant features through gated attention. Gene expression data with complex spatial relationships can be captured with its use of CNNs. Although the model is highly effective, there may still be some challenges in fully capturing the nuances of multimodal data given its accuracy. Furthermore, heterogeneous stacking achieves a remarkable 97% accuracy, illustrating its superior ability to generalize across different data modalities. Combining different deep learning models that excel at different aspects of a dataset has the advantage of high performance. Stacking of models provides a rich, complementary set of information, resulting in a highly accurate and robust prediction system. However, compared to the other SOTA models, the deep-learning-based stacked ensemble model, with 88.1% accuracy, fails to meet expectations. Many factors could contribute to this, such as the choice of base models, the method of combining their outputs, or even the specifics of feature engineering. As a result of the slightly lower accuracy, the model may not use the full potential of the multimodal data or integrate the model outputs as effectively as heterogeneous stacking. Overall, the performance disparities between these models illustrate the importance of model architecture, data integration strategies, and capturing complex relationships within datasets. Due to its richness, the METABRIC dataset lends itself well to advanced machine learning models to explore and improve breast cancer prognosis. Figure 21 shows the different model accuracies for this dataset.

6.6.3 Model accuracy comparison on the Wisconsin dataset

Several state-of-the-art (SOTA) models have been developed using machine learning in breast cancer diagnostics to improve prediction accuracy and reliability. Deep learning approaches, such as Convolutional Neural Networks (CNNs) and ensemble methods, have demonstrated significant success in machine learning applications. In order to ensure early detection of breast cancer and to plan treatment accordingly, high accuracy,

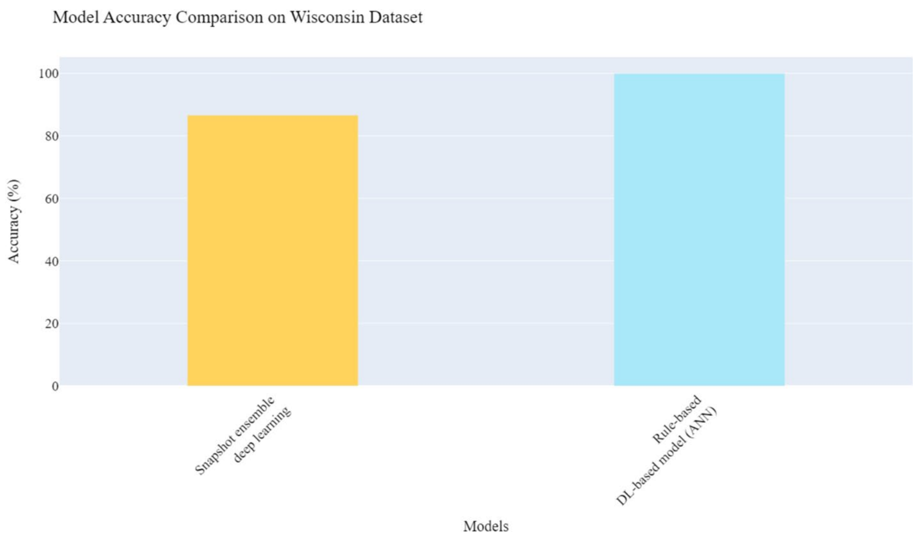


Fig. 22 Comparison of model accuracy on the Wisconsin dataset

sensitivity, and specificity must be achieved. In processing complex patterns within a dataset, deep learning models, especially those utilizing CNNs and ensemble techniques, have demonstrated remarkable performance. For instance, Sharma et al. [134] employed the snapshot ensemble deep learning approach, which achieved an accuracy of 86.6%, which takes multiple snapshots of the model at different stages of training to form an ensemble, thus capturing diverse representations of the data and improving robustness. Awotunde et al. [136] have proposed rule-based AI models that have enhanced the accuracy of deep learning to 99.9%. The high accuracy indicates the models' ability to learn intricate relationships within data. Model interpretability and reliability are enhanced by the rule-based component, which adheres to predefined rules. As a result of the inherent complexity and design of each model, the performance of machine learning models on the Wisconsin Breast Cancer dataset varies. Though snapshot ensemble deep learning achieves 86.6% accuracy, its ability to capture global data distributions and variability may limit its ability to accurately capture all nuances of the dataset. Typically, ensembles combine weak learners to create a stronger one, but snapshot methods might not exploit model diversity to its full potential, leading to lower accuracy. However, the rule-based ANN model achieves a near-perfect accuracy of 99.9%, demonstrating the effectiveness of deep learning architectures in identifying and learning from complex patterns in medical data. Its high performance can be attributed to its ability to capture nonlinear relationships and its versatility in adapting to various dataset features. For critical applications like cancer diagnosis, rule-based logic enhances its robustness and interpretability. Nonetheless, the model may not generalize as well to unseen data when it performs exceptionally well on the training data, such as 99.9% accuracy. Predictive systems that are highly accurate and generalizable are essential in a medical context to avoid misleading diagnoses. As a result of the performance disparities among models, selecting the right algorithm and validation strategy is crucial to achieving optimal results in the detection of breast cancer using Wisconsin data. Figure 22 shows the different model accuracies for this dataset.

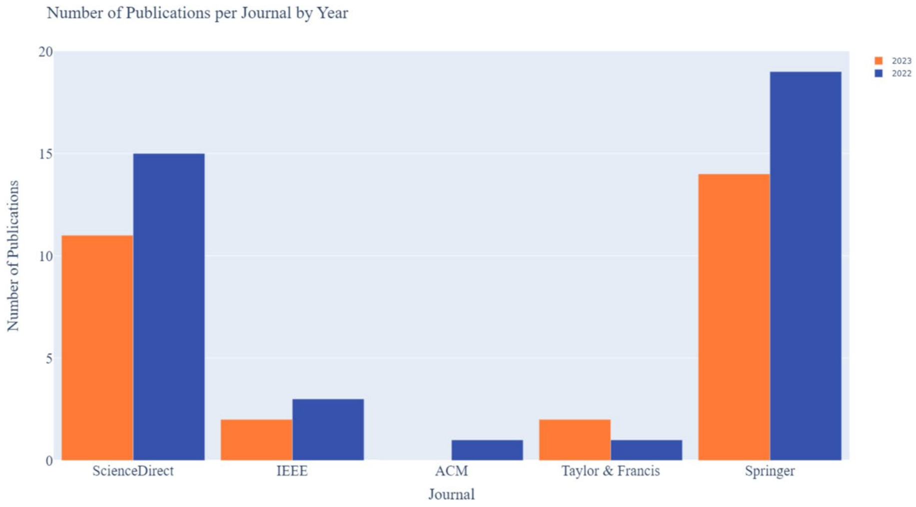


Fig. 23 Evolution of deep learning publications in breast cancer detection (2022–2023) across leading academic publishers

7 Analytical discussion

Deep learning approaches are explored in this comprehensive review of breast cancer detection with deep learning. In this review, the literature is systematically examined, providing an analytical perspective. Additionally, the review examines various deep learning applications in the context of breast cancer detection and develops a taxonomy and parameter analysis of them. There has been a critical analysis of research in the field of deep learning and how it can be applied to early detection and diagnosis of breast cancer.

7.1 Analyze based on the year and publisher

Figure 23 illustrates a notable trend in the publication of articles related to breast cancer detection using deep learning across various academic platforms between 2022 and 2023. It is identified that Springer is the leading publisher, with its articles decreasing from 19 in 2022 to 14 in 2023. ScienceDirect is closely following, with 15 articles produced in 2022 and 11 in 2023. However, IEEE and Taylor & Francis maintain lower publication rates in this research area. Over the past two years, the ACM has published only one article. Springer and ScienceDirect are emerging as key disseminators of research in breast cancer detection using deep learning techniques. Developing computer science to address critical health problems is a growing trend in interdisciplinary research.

In Fig. 24, the most prominent keywords pertain to breast cancer and deep learning, indicating strong and broad interest in this topic. Deep learning research focuses specifically on convolutional neural networks, emphasizing their importance. Furthermore, there is noticeable attention paid to broader machine learning concepts such as transfer learning and general machine learning techniques, suggesting an exploration of varied computational approaches. In addition to infrared thermography, feature extraction, prognosis prediction, segmentation, and ultrasound imaging, there are a variety of imaging methodologies and

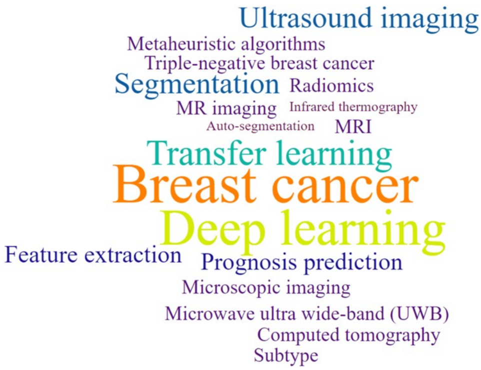


Fig. 24 The frequency analysis of keywords in the reviewed articles

techniques being explored. There are also terms such as microwave ultra-wideband, MR imaging, and radiomics, as well as specific breast cancer types, including triple-negative breast cancer, and imaging modalities such as MRI and computed tomography. Within the broader topic, these are niche but vital areas of research. Moreover, current studies include metaheuristic algorithms, auto-segmentation, and subtype analysis. The keyword distribution illustrates a research landscape that combines diverse methodological approaches and utilizes advanced computational techniques to address breast cancer detection challenges. Imaging and diagnosis applications of deep learning are a testament to the field's evolution and its potential for breakthroughs in the future.

7.2 Analyze based on the type of imagining

The realm of breast cancer detection research is inherently intertwined with the selection of imaging modalities. In this section, an examination is conducted on the distribution of articles among various imaging techniques to elucidate dominant patterns. Figure 25 offers a visual depiction of this analysis, presenting the number of articles associated with each imaging modality. This graphical presentation serves as a pivotal reference point for delving into the nuanced insights derived from the survey.

Based on Fig. 25 It turns out that mammographic imaging has been extensively explored, with the highest number of articles. In this way, it demonstrates its effectiveness and established role in breast cancer screening and diagnosis. Ultrasound Imaging is also a significant focus, but is less prominent than mammography, indicating its relevance as a non-invasive and accessible diagnostic tool. MRI and CT imaging, with fewer articles, are used for more detailed and complex diagnostic scenarios. Microscopic and histopathological imaging are also receiving considerable attention. As a result, a microscopic approach to detection is more important when understanding and diagnosing breast cancer than other imaging methods. In the field of thermographic imaging, there are fewer articles, which may indicate its niche or emerging status. The category labeled "Others" indicates a diverse range of alternative or less common techniques being explored, showing the field's openness to innovative approaches and methodologies.

Figure 26 provides a comprehensive overview of the diagnostic performances associated with different imaging modalities in the context of breast cancer detection using deep learning techniques. The analysis indicates distinct success rates and efficiencies among modalities.

Fig. 25 The categorization of articles based on imaging techniques

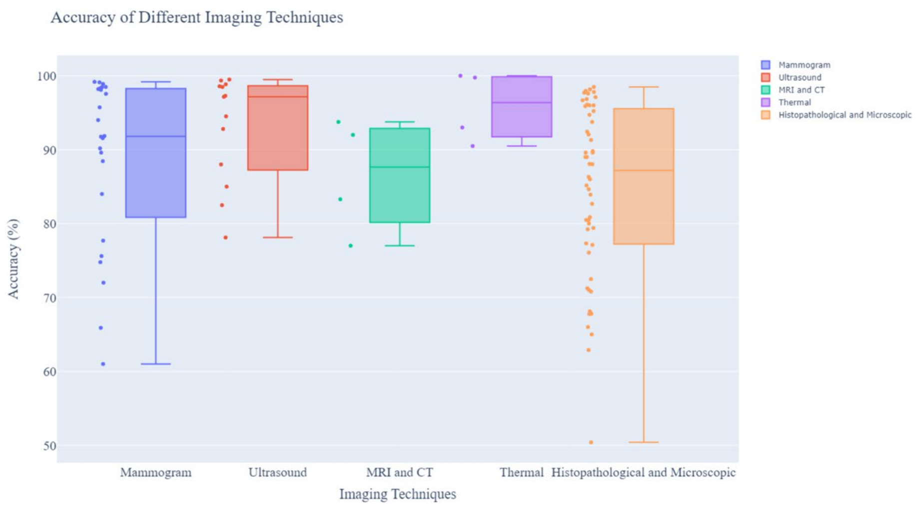


Fig. 26 Accuracy boxplot for different imaging

Figure 26 shows the accuracy percentages of various imaging techniques for breast cancer detection using deep learning. The techniques compared include Mammogram, Ultrasound, MRI and CT, Thermal, and Histopathological and Microscopic. For each method, there is a clear range of accuracy, with medians, quartiles, and outliers indicating central tendency and variability. The accuracy of mammograms and histopathological and microscopic techniques varies based on conditions and implementation. It seems that although mammograms are often reliable, some cases or datasets result in significantly lower accuracy than the median of 95%, with some outliers as low as 60%. Microscopic

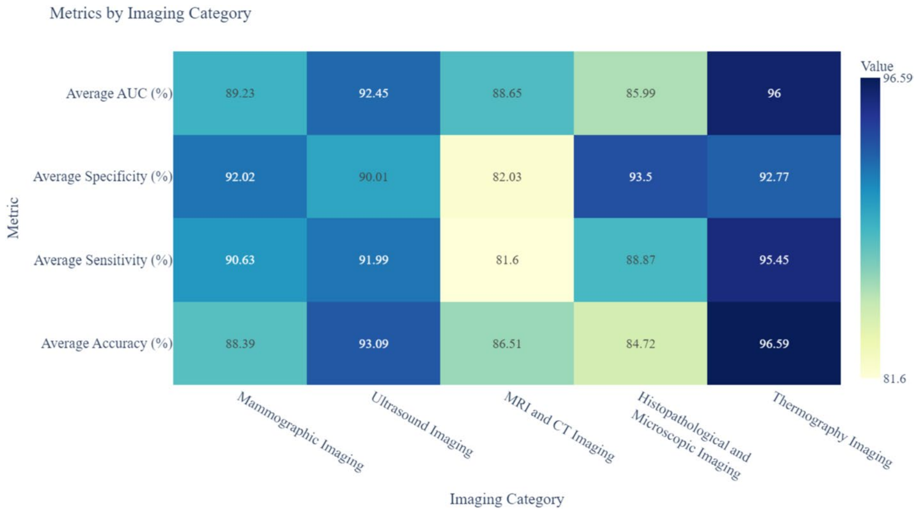


Fig. 27 Comparative diagnostic performance of imaging modalities in deep learning-based breast cancer detection

and histopathological methods, on the other hand, exhibit high median accuracy and a wide distribution, with numerous outliers on both ends, indicating an inconsistent interpretation pattern or differences in sample quality. There are differences between other imaging techniques. There is a significant difference in median accuracy between ultrasound and thermal imaging, particularly with ultrasound, which clusters tightly around high accuracy values. Even though thermal imaging shows high median accuracy, it also shows a broader range without extreme outliers, suggesting reliability with occasional variances. Even though MRI and CT are advanced imaging modalities, they show a lower median accuracy and a wide interquartile range, suggesting resolution limitations or interpretation challenges specific to these modalities. It is possible to attribute the distinct performance levels across these techniques to the inherent capabilities of each imaging method to capture specific aspects of breast tissue, the quality of the image data, and the efficiency with which deep learning models are used to interpret the images.

Figure 27 shows the performance metrics of different imaging categories used for breast cancer detection using deep learning. These metrics include Average Accuracy, Average Sensitivity, Average Specificity, and Average AUC. Its superior performance in detecting breast cancer is evident from the high average accuracy (96.59%) and average AUC (96%) among imaging techniques. Due to the ability of thermal imaging to detect temperature variations associated with cancerous tissues, deep learning models are able to leverage this unique diagnostic feature. Both mammography and ultrasound imaging perform well, especially in terms of average sensitivity (90.63% for mammography and 91.99% for ultrasound) and average specificity (92.02% for mammography). In addition to its high specificity, mammography is effective in identifying non-cancerous cases, reducing the possibility of false positive results. Due to its high sensitivity (93.09%) and accuracy (93.09%), ultrasound imaging is a reliable diagnostic tool that can diagnose dense breast tissues as well as abnormalities. The average sensitivity (61.60%) and specificity (81.60%) of MRI and CT imaging are generally low, suggesting that certain types of breast cancer features cannot be captured accurately or that the quality of the images may be variable. Histopathological

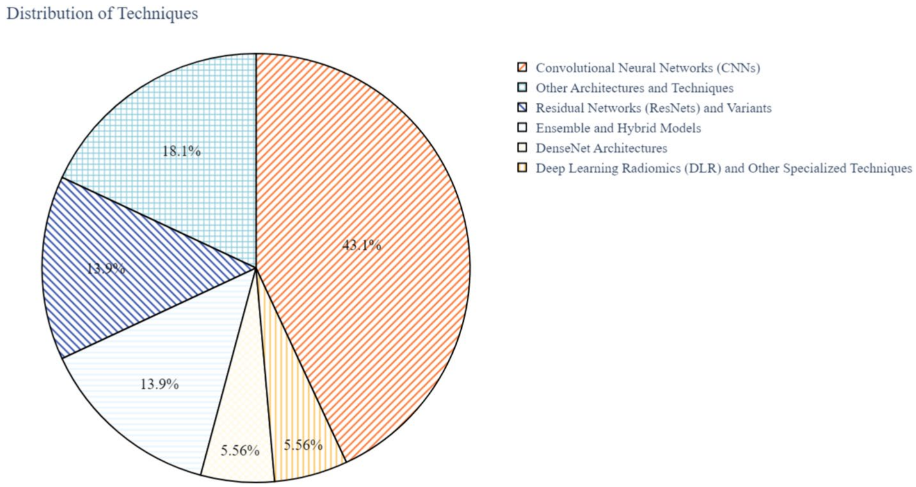


Fig. 28 Distribution of articles based on utilized deep learning algorithms in breast cancer detection research

and microscopic imaging exhibit strong specificity (93.5%) but lower accuracy (84.72%), possibly due to challenges with sample preparation and interpretation. The differences between these imaging techniques emphasize the importance of choosing the appropriate technique for the diagnostic requirement based on each modality's strengths.

7.3 Analyze based on the deep learning algorithms

Within the scope of this section, an in-depth analysis is conducted on the application of deep learning algorithms in breast cancer detection. The objective is to discern the prevalence and effectiveness of well-known deep learning methods and architectures. In Fig. 28, the distribution of articles across these notable algorithms is visualized, offering insights into research trends and preferences.

As shown in Fig. 28, Convolutional Neural Networks (CNNs) stand out as the most commonly used architecture. Medical imaging data can be analyzed effectively with their help because of their effectiveness in image recognition and classification tasks. Even though ResNets and their variants are less prominent than CNNs, they remain significant. Their ability to address issues such as vanishing gradients makes them a valuable tool in complex image analysis. Despite being less frequently mentioned, DenseNet Architectures are becoming more popular. As a result of their unique connectivity pattern, which promotes feature reuse, they can be particularly useful in medical imaging when subtle features are essential. In addition to ensemble and hybrid models, the research community is interested in combining different models and techniques for greater accuracy and robustness. Performance is often enhanced by leveraging the strengths of various architectures. It may not be extensively discussed, but Deep Learning Radiomics (DLR) focuses on extracting quantitative features from medical images utilizing deep learning. The presence of other architectures and techniques also suggests continuous innovation and experimentation in the field, indicating that researchers are exploring a wide range of methods to find breast cancer detection solutions.

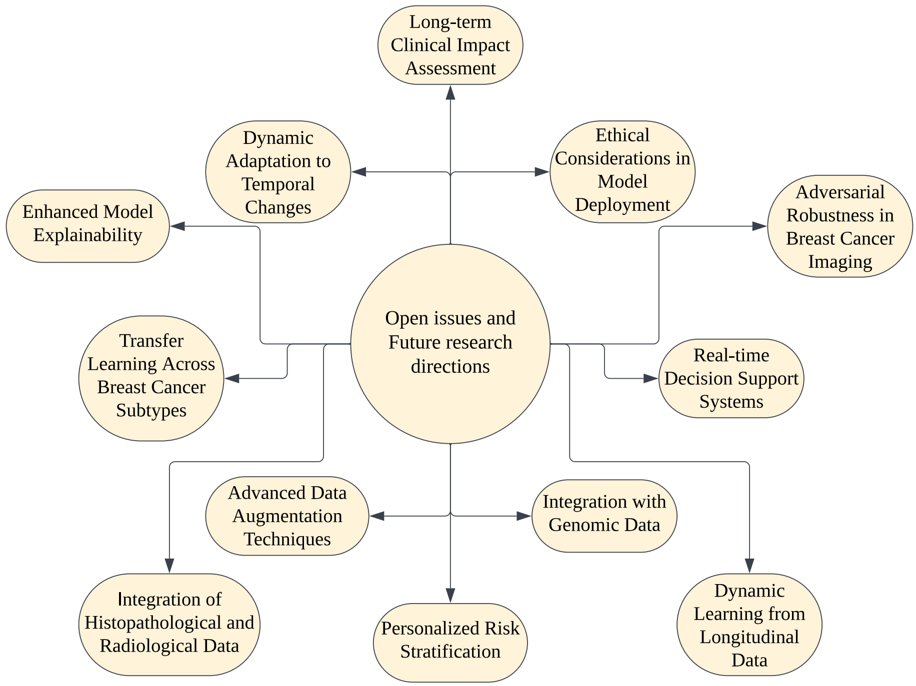


Fig. 29 An overview of open issues and future research directions

8 Future research directions

The field of breast cancer diagnosis using deep learning remains subject to several open issues and possible directions for future research. It is imperative to address these issues in order to advance the technology to a point where it can be used safely and effectively in clinical practice. Figure 29 provides an overview of these future research directions.

- **Integration of histopathological and radiological data**
  Deep learning models can enhance diagnostic accuracy by incorporating histopathological and radiological data to provide a more comprehensive understanding of breast cancer characteristics. A combination of tissue pathology and imaging modalities may lead to better diagnostic outcomes and improved patient care.

- **Enhanced model explainability**
  It is crucial to develop methodologies for enhancing the explainability of deep learning models for breast cancer diagnosis. Interpretability of models is essential for gaining clinicians' trust and integrating them into clinical workflows. The use of explainable AI (XAI) methods can help clinicians gain confidence and acceptance of models.

- **Dynamic learning from longitudinal data**

In order to better understand breast cancer's evolving nature, dynamically learning models will be developed using longitudinal patient data, which will identify biomarkers and treatment response patterns. Long-term patient outcomes can be improved by such models, which enhance personalized treatment plans.

- **Adversarial robustness in breast cancer imaging**
  Maintaining model reliability in clinical scenarios requires analyzing adversarial attacks on breast cancer imaging models. In order to deploy these models safely in clinical settings, robustness against maliciously crafted inputs is crucial.

- **Personalized risk stratification**
  Breast cancer management can be improved by developing personalized risk stratification strategies based on a patient's unique clinical, genetic, and imaging data. A more effective monitoring and treatment plan can be achieved by customizing risk assessment models according to individual profiles.

- **Integration with genomic data**
  Deep learning models and genomic data can be integrated to understand the molecular mechanisms underlying breast cancer. Developing targeted treatments and improving diagnostic accuracy can be achieved through this integration.

- **Advanced data augmentation techniques**
  Data augmentation techniques can be developed and evaluated in order to enhance the generalization of breast cancer imaging datasets. Large, diverse datasets are especially important in scenarios where data augmentation is difficult.

- **Real-time decision support systems**
  Real-time decision support systems can streamline clinical workflows and expedite patient care in breast cancer diagnosis. Clinical practice can benefit significantly from algorithms that provide immediate feedback during diagnostic procedures.

- **Transfer learning across breast cancer subtypes**
  Model performance can be improved by exploring transfer learning techniques across breast cancer subtypes. In order to improve the accuracy and robustness of models for all breast cancer subtypes, it is critical to utilize knowledge from well-represented subtypes.

- **Dynamic adaptation to temporal changes**
  Models must be dynamically adapted to changes in breast cancer characteristics over time for accurate prognostic assessments and treatment plans. Long-term patient management can be enhanced by considering tumor morphology and behavior over time.

- **Ethical considerations in model deployment**
  A fair and equitable breast cancer diagnosis requires an assessment of the ethical implications of deep learning models. It is vital for ethical deployment of these models to address biases caused by data imbalances.

- **Long-term clinical impact assessment**
  The integration of deep learning models into breast cancer diagnostic workflows is essential for long-term evaluation of their clinical impact. These technologies are more likely to be adopted into clinical practice if their effectiveness, resource utilization, and patient outcomes are evaluated.

9 Conclusion

In order to enhance early detection and treatment outcomes of breast cancer, advances in diagnostic methodologies are necessary. As part of our survey, we review a diverse range of articles in order to investigate imaging modalities, deep learning algorithms, and publication trends in this critical field of medical research. In this paper, we have categorized and analyzed various imaging techniques, with a particular focus on ultrasound imaging's role in improving diagnostic accuracy. The prevalence and effectiveness of convolutional neural networks (CNNs) and their variants in breast cancer detection have been highlighted. We found that ultrasound imaging greatly enhances sensitivity and specificity, leading to a high level of diagnostic accuracy. Additionally, CNNs and their variants are widely used for image analysis in deep learning algorithms. Increasing numbers of publications in this area reflect technology and medicine's rapidly evolving nature. Although these findings are promising, the review acknowledges inherent limitations. There are theoretical limitations in model robustness, generalizability, evaluation metrics, algorithmic complexity, transparency, scalability, and reproducibility. Practical limitations include data quality and availability, computational resources, clinical validation, regulatory approval, and integration into clinical workflows. To improve breast cancer diagnosis and patient outcomes, it is crucial to address these challenges.

By identifying key challenges and examining the current landscape, this survey prepares the ground for future advancements in breast cancer detection. The integration of computer science and healthcare is essential to overcome theoretical and practical limitations, enabling deep learning to improve diagnostic accuracy and patient care to its full potential.

9.1 Limitations

Despite the fact that this survey provides a broad understanding of the current state of breast cancer detection using deep learning, it is also important to acknowledge that the reviewed studies have inherent limitations. It is essential to understand these limitations in order to guide future research and to improve the practical applicability of deep learning models in clinical settings. A broad classification of limitations can be made based on theoretical and practical aspects, with each presenting its own challenges.

9.1.1 Theoretical limitations

Deep learning approaches have theoretical limitations concerning model design, algorithm performance, and conceptual frameworks. Current technologies and methodologies often have inherent limitations that make it impossible to develop robust, adaptable, and generalizable models.

- **Model robustness and generalizability**
  The datasets used for training deep learning models heavily affect their performance. The current models often rely on specific datasets that do not reflect the full spectrum of clinical variability, potentially leading to biased results and reduced generalizability. The generalization of models trained on limited datasets may fail, especially in diverse clinical settings. This specificity can lead to biases that limit the applicability of these models to larger populations of patients. As well as overfitting, deep learning models perform poorly on new data if they're trained on small or homogeneous datasets. Clearly, more diverse and comprehensive datasets are needed.

- **Evaluation metrics**
  There are no standardized evaluation metrics, which makes it difficult to compare results across studies. The development of universally applicable models and the comparison of different approaches require consistent metrics. In the absence of standardized metrics, it is difficult to compare and validate different approaches, making it difficult to develop universally applicable and reliable models.

- **Algorithmic complexity and transparency**
  Despite the complexity of deep learning models, particularly CNNs, clinical acceptance is limited due to transparency and interpretability challenges. Deep learning models can be very complex, making their decision-making processes difficult to understand. Clinicians need to understand how a model comes up with its conclusions in order to trust and accept it. For increasing clinical trust and acceptance, Explainable AI methods (XAI) are crucial for providing insights into model decisions.

- **Scalability and reproducibility**
  It is critical to ensure that deep learning models can be applied across a variety of settings and populations in order for them to be widely adopted. It is common to develop and test models in controlled environments, however, their scalability to a variety of clinical settings remains uncertain. Models must be scalable in order to be effective. The ability to reproduce results across different studies and datasets is also crucial to validating models. Clinical applications relying on deep learning models may suffer from reproducibility issues.

9.1.2 Practical limitations

In clinical environments, deep learning models face real-world challenges when implementing and deploying them. For models to be effective and reliable, high-quality data, standardized protocols, and seamless integration into clinical workflows are essential.

- **Data quality and availability**
  It is essential to have access to high-quality, annotated datasets when training deep learning models. The lack of such datasets, however, creates significant barriers to progress. Training and validating models effectively require high-quality annotated datasets. Large-scale annotated datasets are scarce, making collaborative efforts to create and share them essential for the development of robust models. In diverse clinical settings, imaging techniques and equipment can vary widely, making it challenging to develop models that will perform consistently. Imaging protocols can be standardized to mitigate this problem.

- **Computational resources and infrastructure**
  In resource-constrained settings, deep learning models can be prohibitively expensive to develop and deploy. There are some research and clinical environments where deep learning models cannot be trained due to a lack of computing power. In order to develop and refine these models, high-performance computing infrastructure is essential. To ensure deep learning models can be deployed efficiently and scalable in real-world settings, substantial computational resources and infrastructure are required.

- **Clinical validation and regulatory approval**
  It is challenging to move from research to clinical practice without rigorous validation, regulatory approval, and seamless integration into existing healthcare workflows. The adoption of AI models in clinical settings requires rigorous validation and regulatory approval. As a result, models are safe for use in patient care and meet clinical standards. It can be challenging to navigate the regulatory landscape for AI in healthcare, posing a significant barrier to clinical adoption.

- **Integration into clinical workflows**
  Successful adoption of AI models requires seamless integration into existing healthcare workflows. It is important for models to complement and enhance current practices rather than complicate them. Integrating AI models seamlessly into existing clinical workflows is imperative for their practical implementation, as disrupting established practices can hinder adoption and reduce their perceived value. The use of AI tools by clinicians requires adequate training. It is crucial for a successful implementation to build trust, ensure user acceptance, and demonstrate model reliability.

In spite of the challenges outlined, this survey provides valuable insight into the changing landscape of breast cancer detection, paving the way for future developments in this crucial intersection between technology and healthcare. Deep learning can be fully harnessed to improve breast cancer diagnosis and outcomes with continued research to address these limitations. Computer science and healthcare are convergent fields that hold great promise for improving diagnostic accuracy and patient care, but overcoming theoretical and practical limitations will be key.

9.2 Advances and obstacles in deep learning-based breast *cancer* detection

In recent years, deep learning has gained popularity in the area of breast cancer detection and diagnosis, with numerous papers proposing methods, models, and algorithms aimed at improving detection, classification, and prognosis accuracy. The advantages and disadvantages of cutting-edge technologies will be discussed.

Cutting-edge advancements:

- Improved Detection Accuracy: Medical images such as mammograms, ultrasounds, and MRIs can be detected using deep learning techniques, particularly convolutional neural networks (CNNs). In this way, these models can identify subtle patterns and abnormalities that might go unnoticed by radiologists, leading to a more accurate and earlier diagnosis.
- Personalized Medicine: Personalized treatment plans can be developed by analyzing large datasets of patient records, genetic information, and medical images. Different tumor characteristics, genetics, and responses to treatment can be accounted for in these models, which can aid in tailoring treatments to improve outcomes.
- Automated Image Analysis: Healthcare professionals can save time and potentially reduce human error by automating the process of analyzing medical images with deep learning algorithms. Automation can result in faster diagnosis and treatment planning, especially when resources are limited and radiologists are scarce.
- Integration with Multi-Modal Data: A comprehensive understanding of breast cancer can be gained through deep learning, which integrates diverse data sources including imaging, genomic, and clinical records. Research can uncover hidden relationships and biomarkers by combining data from different modalities.
- Transfer Learning and Pretrained Models: The transfer learning method allows researchers to fine-tune deep learning models trained on large datasets (e.g., ImageNet) for specific tasks in breast cancer detection and diagnosis. As a result, model development can be accelerated and performance improved, particularly when limited labeled medical imaging datasets are available.

Dark sides and challenges:

- **Data bias and generalization issues:** Data sets that are biased or unrepresentative may result in discrepancies in diagnosis and treatment recommendations across demographic groups. In order to address these biases and improve model generalization, diverse and representative training data must be provided.
- **Interpretability and transparency:** The process of understanding how deep learning models reach specific predictions is often described as a "black box." Interpretability is an issue that can undermine trust and acceptability among healthcare professionals and patients, especially in critical medical decision-making situations. The goal of deep learning systems is to increase transparency and explain model predictions.
- **Overfitting and robustness:** Overfitting can occur with deep learning models, especially when data is limited or noisy. When overfitting models are applied to unseen data, they won't perform well, resulting in inaccurate predictions in real-life situations. In clinical practice, robustness and generalization capability are essential.

- **Ethical and legal concerns:** In healthcare, deep learning raises a variety of ethical and legal considerations, such as patient privacy, consent, and liability. To protect patient privacy and rights, HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation) compliance is essential.
- **Integration with clinical workflow:** Deep learning models are challenging to integrate into existing clinical workflows due to factors such as compatibility with medical imaging systems, regulatory approval processes, and healthcare provider acceptance. Deep learning tools must be seamlessly integrated and have user-friendly interfaces in order to be adopted and used in clinical practice.

To maximize the potential of deep learning for improving patient outcomes and health care delivery, addressing the associated challenges and ethical considerations is paramount to achieving its full potential. It is essential that researchers, clinicians, policymakers, and industry stakeholders collaborate to navigate the complexities and ensure that deep learning is used responsibly and effectively in breast cancer treatment.

9.3 Leveraging deep learning for enhanced breast *cancer* management

Breast cancer remains one of the most prevalent and deadly forms of cancer affecting women worldwide. Recent advancements in deep learning, a subset of artificial intelligence (AI), have shown significant promise in transforming breast cancer diagnosis, treatment, and overall patient management. This section explores how deep learning can further enhance the performance and outcomes in breast cancer management.

- **Fine-grained subtype classification:** The use of deep learning techniques has the potential to enhance the classification of breast cancer subtypes based on molecular profiles, histopathological characteristics, and clinical characteristics. Research groups can identify subtle molecular signatures associated with breast cancer subtypes by analyzing multi-omics data (e.g., genomics, transcriptomics, proteomics). Finer-grained classification can improve patient stratification and prognosis prediction for targeted therapies.
- **Early detection and risk assessment:** Incorporating clinical records, imaging data, and genomic data into deep learning algorithms can enhance risk assessment models. Deep learning models can detect early signs of breast cancer development and predict individual risk trajectories using longitudinal data and temporal patterns. Detection and assessment of breast cancer early can enable timely interventions, such as screenings, lifestyle adjustments, and preventive measures.
- **Drug discovery and therapeutic response prediction:** Drug discovery and development can be facilitated through deep learning approaches in breast cancer research. With deep learning models, drug-target interactions, drug efficacy, and adverse effects can be predicted with unprecedented accuracy based on large-scale biomedical datasets and chemical libraries. As well, deep learning techniques like generative adversarial networks (GANs) can generate synthetic data to augment limited experimental datasets and accelerate preclinical drug screening. Additionally, deep learning facilitates personalized treatment strategies based on molecular profiles, clinical characteristics, and treatment histories.
- **Radiogenomics and radiomics integration:** By integrating radiomic features extracted from medical images with genomic data, deep learning can uncover imaging-

genomic associations in breast cancer and identify biomarkers. Deep learning models can be used to reveal underlying biological mechanisms, tumor heterogeneity, and treatment response by analyzing imaging phenotypes and genetic signatures. It is also possible to use radiogenomics to guide image-guided biopsies, treatment planning, and monitoring of therapeutic response in breast cancer patients.
- **Cross-domain knowledge transfer:** Breast cancer research can be advanced through the use of deep learning models trained on diverse datasets from related domains (e.g., pathology, oncology, bioinformatics). Transfer learning reduces the need for large annotated datasets and accelerates model development by allowing pretrained models and feature representations to be reused across a variety of tasks and datasets. In order to address complex challenges in breast cancer detection, diagnosis, and treatment, interdisciplinary research communities can collaborate.

Overall, deep learning has immense potential to advance our understanding of breast cancer, improve diagnostic accuracy, guide treatment decisions, and ultimately, improve patient outcomes. Deep learning can transform breast cancer prevention, diagnosis, and treatment by embracing interdisciplinary approaches, integrating heterogeneous data, and fostering collaborative partnerships.

Acknowledgements Not applicable

Funding No funds, grants, or other support was received.

Data availability The datasets generated and/or analyzed during the current study are not publicly available due but are available from the corresponding author on reasonable request.

Declarations

Consent for publication Not applicable.

Ethics approval and consent to participate Not applicable.

Competing interests The authors declare that they have no competing interests.

References

1. Siddiqui S, Akram T, Ashraf I, Raza M, Khan MA, Damaševičius R (2024) CG-Net: A novel CNN framework for gastrointestinal tract diseases classification. Int J Imaging Syst Technol 34:e23081
2. Anand A, Bedi J, Aggarwal A, Khan MA, Rida I (2024) Authenticating and securing healthcare records: a deep learning-based zero watermarking approach. Image Vis Comput 145:104975
3. Bibi S, Khan MA, Shah JH, Damaševičius R, Alasiry R, Marzougui A, Alhaisoni M, Masood M, Anum A (2023) MSRNet: Multiclass skin lesion recognition using additional residual block based fine-tuned deep models information fusion and best feature selection. Diagnostics 13:3063
4. Boyle P, Levin B (eds) (2008) World cancer report 2008. International Agency for Research on Cancer (IARC). https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2008
5. R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal, Cancer statistics, 2023, CA: A Cancer Journal for Clinicians, 73, 17–48, (2023).
6. Dar RA, Rasool M, Assad A (2022) Breast cancer detection using deep learning: datasets, methods, and challenges ahead. Comput Biol Med 149:106073

7. Iqbal S, Qureshi AN, Aurangzeb K, Alhussein M, Haider SI, Rida I (2023) AMIAC: adaptive medical image analysis and classification, a robust self-learning framework. Neural Comput Appl. https://doi.org/10.1007/s00521-023-09209-1
8. Rehman SU, Khan MA, Masood A, Almujally NA, Baili J, Alhaisoni M, Tariq U, Zhang Y-D (2023) BRMI-Net: Deep learning features and flower pollination-controlled regula falsi-based feature selection framework for breast cancer recognition in mammography images. Diagnostics 13:1618
9. Jabeen K, Khan MA, Hameed MA, Alqahtani O, Alouane MT-H, Masood A (2024) A novel fusion framework of deep bottleneck residual convolutional neural network for breast cancer classification from mammogram images. Front Oncol 14:1347856
10. Parashar A, Rishi R, Parashar A, Rida I (2023) Medical imaging in rheumatoid arthritis: a review on deep learning approach. Open Life Sci 18:20220611
11. Fatima M, Khan MA, Shaheen S, Almujally NA, Wang S-H (2023) B2C3NetF2: breast cancer classification using an end-to-end deep learning feature fusion and satin bowerbird optimization controlled Newton Raphson feature selection. CAAI Trans Intell Technol 8:1374–1390
12. Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, Brenner RJ, Bassett L, Berg W, Feig S, Hendrick E, Mendelson E, D'Orsi C, Sickles E, Burhenne LW (2010) Breast cancer screening with imaging: recommendations from the society of breast imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. J Am Coll Radiol 7:18–27
13. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y (2009) Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE Trans Inf Technol Biomed 13:236–251
14. Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA (2001) Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 91:1724–1731
15. Dromain C, Balleyguier C, Adler G, Garbay JR, Delaloge S (2009) Contrast-enhanced digital mammography. Eur J Radiol 69:34–42
16. Kelly KM, Dean J, Comulada WS, Lee SJ (2010) Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur Radiol 20:734–742
17. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C (2019) Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. Radiology 290:290–297
18. Hizukuri A, Nakayama R, Nara M, Suzuki M, Namba K (2021) Computer-aided diagnosis scheme for distinguishing between benign and malignant masses on breast DCE-MRI images using deep convolutional neural network with Bayesian optimization. J Digit Imaging 34:116–123
19. Yang X, Wu L, Ye W, Zhao K, Wang Y, Liu W, Li J, Li H, Liu Z, Liang C (2020) Deep learning signature based on staging CT for preoperative prediction of sentinel lymph node metastasis in breast cancer. Acad Radiol 27:1226–1233
20. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S (2017) Breast cancer multi-classification from histopathological images with structured deep learning model. Sci Rep 7:1–10
21. Frize M, Herry C, Roberge R (2002) Processing of thermal images to detect breast cancer: comparison with previous work. In Proceedings of the IEEE conference, pp 23–26. https://doi.org/10.1109/IEMBS.2002.1106325
22. Ng EYK, Ung LN, Ng FC, Sim LSJ (2001) Statistical analysis of healthy and malignant breast thermography. J Med Eng 25:253–263
23. Farooq MA, Corcoran P (2020) Infrared imaging for human thermography and breast tumor classification using thermal images. 31st Irish Signals and Systems Conference (ISSC), pp 1–6. https://doi.org/10.1109/ISSC49989.2020.9180164
24. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL (2015) Breast cancer surveillance consortium, diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 175:1828–1837
25. Schaller RR (1997) Moore's law: past, present and future. IEEE Spectr 34:52–59
26. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56
27. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J (2019) A guide to deep learning in healthcare. Nat Med 25:24–29
28. Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 49:1225–1231
29. Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ (2021) Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. Neurocomputing 444:92–110

30. Fang W, Love PE, Luo H, Ding L (2020) Computer vision for behavior-based safety in construction: a review and future directions. Adv Eng Inform 43:100980
31. Li H-C, Deng Z-Y, Chiang H-H (2020) Lightweight and resource-constrained learning network for face recognition with performance optimization. Sensors 20:6114
32. Wang W, Gang J (2018) Application of convolutional neural network in natural language processing. International conference on information systems and computer aided education (ICISCAE). https://doi.org/10.1109/ICISCAE.2018.8666928
33. Palaz D, Magimai-Doss M, Collobert R (2019) End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. Speech Commun 108:15–32
34. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press. https://doi.org/10.1007/s10710-017-9314-z
35. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324
36. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inform Process Syst 25
37. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556
38. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. Proc IEEE Conf Comput Vis Pattern Recognit: 2818–2826. https://doi.org/10.1109/CVPR.2016.308
39. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9. https://doi.org/10.1109/CVPR.2015.7298594
40. Deng J, Dong W, Socher R, Li LJ, Kai L, Li F-F (2009) ImageNet: a large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, pp 248–255. https://doi.org/10.1109/CVPR.2009.5206848
41. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 223–226. https://doi.org/10.1109/CVPR.2016.90
42. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, p 31. https://doi.org/10.1609/aaai.v31i1.11231
43. Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T et al (2017) MobileNets: efficient convolutional neural networks for mobile vision applications. https://doi.org/10.48550/arXiv.1704.04861
44. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) MobileNetV2: inverted residuals and linear bottlenecks. IEEE/CVF conference on computer vision and pattern recognition, pp 4510–4520. https://doi.org/10.48550/arXiv.1801.04381
45. Howard A, Sandler M, Chu G, Chen L, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H (2019) Searching for MobileNetV3. ArXiv. https://doi.org/10.48550/arXiv.1905.02244
46. Tan M, Le Q (2019) EfficientNet: rethinking model scaling for convolutional neural networks. Proceedings of the 36th international conference on machine learning, proceedings of machine learning research (PMLR). https://doi.org/10.48550/arXiv.1905.11946
47. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6:60
48. Kumar S, Bhupati Bhambu P, Pachar S, Cotrina-Aliaga JC, Arias-Gonzáles JL (2023) Deep learning-based computer-aided diagnosis model for the identification and classification of mammography images. SN Comput Sci 4:502
49. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc. https://dl.acm.org/doi/10.5555/534133
50. Bacanin N, Stoean R, Zivkovic M, Petrovic A, Rashid TA, Bezdan T (2021) Performance of a novel chaotic firefly algorithm with enhanced exploration for tackling global optimization problems: application for dropout regularization. Mathematics 9:2705
51. Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm intelligence: from natural to artificial systems. Oxford University Press. https://academic.oup.com/book/40811
52. Aleksendrić D, Carlone P (2015) Composite materials – modelling, prediction and optimization. In: Soft computing in the design and manufacturing of composite materials. Woodhead Publishing, pp 61-289, https://doi.org/10.1533/9781782421801.61
53. Malakar S, Ghosh M, Bhowmik S, Sarkar R, Nasipuri M (2020) A GA-based hierarchical feature selection approach for handwritten word recognition. Neural Comput Appl 32:2533–2552

54. Gajic L, Cvetnic D, Zivkovic M, Bezdan T, Bacanin N, Milosevic S (2021) Multi-layer perceptron training using hybridized bat algorithm. Springer, Singapore, pp 689–705
55. Rida I, Al-Maadeed N, Al-Maadeed S, Bakshi S (2020) A comprehensive overview of feature representation for biometric recognition. Multimedia Tools and Applications 79:4867–4890
56. Strumberger I, Bacanin N, Tuba M (2017) Enhanced firefly algorithm for constrained numerical optimization. IEEE congress on evolutionary computation (CEC), pp 2120–2127. https://doi.org/10.1109/CEC.2017.7969561
57. Wani NA, Kumar R, Mamta Bedi J, Rida I (2024) Explainable AI-driven IoMT fusion: unravelling techniques, opportunities, and challenges with explainable AI in healthcare. Inf Fusion 110:102472
58. Saraswat D, Bhattacharya P, Verma A, Prasad VK, Tanwar S, Sharma G, Bokoro PN, Sharma R (2022) Explainable AI for healthcare 5.0: opportunities and challenges. IEEE Access 10:84486–84517
59. Rajpal S, Rajpal A, Saggar A, Vaid AK, Kumar V, Agarwal M (2023) XAI-MethylMarker: explainable AI approach for biomarker discovery for breast cancer subtype classification using methylation data. Expert Syst Appl 225:120130
60. Loizidou K, Elia R, Pitris C (2023) Computer-aided breast cancer detection and classification in mammography: a comprehensive review. Comput Biol Med 153:106554
61. Sahu A, Das PK, Meher S (2023) Recent advancements in machine learning and deep learning-based breast cancer detection using mammograms. Phys Med 114:103138
62. Radak M, Lafta HY, Fallahi H (2023) Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies. J Cancer Res Clin Oncol 149:10473–10491
63. Thakur N, Kumar P, Kumar A (2024) A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities. Multimed Tools Appl 83:35849–35942
64. Chugh G, Kumar S, Singh N (2021) Survey on machine learning and deep learning applications in breast cancer diagnosis. Cogn Comput 13:1451–1470
65. Yadavendra Chand S (2020) A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method. Mach Vis Appl 31:46
66. Li Y, Wu X, Yang P, Jiang G, Luo Y (2022) Machine learning for lung cancer diagnosis, treatment, and prognosis. Genom Proteomics Bioinformatics 20:850–866
67. Wen X, Guo X, Wang S, Lu Z, Zhang Y (2024) Breast cancer diagnosis: a systematic review. Biocybernetics and Biomedical Engineering 44:119–148
68. Kerschke L, Weigel S, Rodriguez-Ruiz A, Karssemeijer N, Heindel W (2022) Using deep learning to assist readers during the arbitration process: a lesion-based retrospective evaluation of breast cancer screening performance. Eur Radiol 32:842–852
69. Houssein EH, Emam MM, Ali AA (2022) An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators' algorithm. Neural Comput Appl 34:18015–18033
70. Qu J, Zhao X, Chen P, Wang Z, Liu Z, Yang B, Li H (2022) Deep learning on digital mammography for expert-level diagnosis accuracy in breast cancer detection. Multimed Syst 28:1263–1274
71. Nemade V, Pathak S, Dubey AK (2024) Deep learning-based ensemble model for classification of breast cancer. Microsyst Technol 30:513–527
72. Majji R, G OPP, R C (2023) Smart IoT in breast cancer detection using optimal deep learning. J Digit Imaging 36(4):1489–1506
73. Kavitha T, Mathai PP, Karthikeyan C, Ashok M, Kohar R, Avanija J, Neelakandan S (2022) Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images. Interdiscip Sci: Comput Life Sci 14:113–129
74. Jafarzadeh Ghoushchi S, Ranjbarzadeh R, Najafabadi SA, Osgooei E, Tirkolaee EB (2023) An extended approach to the diagnosis of tumour location in breast cancer using deep learning. J Amb Intell Human Comput 14:8487–8497
75. Alloqmani A, Abushark YB, Khan AI (2023) Anomaly detection of breast cancer using deep learning. Arab J Sci Eng 48:10977–11002
76. Sannasi Chakravarthy SR, Bharanidharan N, Rajaguru H (2023) Deep learning-based metaheuristic weighted k-nearest neighbor algorithm for the severity classification of breast cancer. IRBM 44:100749
77. Zhang M, Wang C, Cai L, Zhao J, Xu Y, Xing J, Sun J, Zhang Y (2023) Developing a weakly supervised deep learning framework for breast cancer diagnosis with HR status based on mammography images. Comput Struct Biotechnol J 22:17–26
78. Dadsetan S, Arefan D, Berg WA, Zuley ML, Sumkin JH, Wu S (2022) Deep learning of longitudinal mammogram examinations for breast cancer risk prediction. Pattern Recogn 132:108919

79. Jiang J, Peng J, Hu C, Jian W, Wang X, Liu W (2022) Breast cancer detection and classification in mammogram using a three-stage deep learning framework based on PAA algorithm. Artif Intell Med 134:102419
80. Raaj RS (2023) Breast cancer detection and diagnosis using hybrid deep learning architecture. Biomed Signal Process Control 82:104558
81. Kumbhare S, Kathole AB, Shinde S (2023) Federated learning aided breast cancer detection with intelligent heuristic-based deep learning framework. Biomed Signal Process Control 86:105080
82. Ramesh S, Sasikala S, Gomathi S, Geetha V, Anbumani V (2022) Segmentation and classification of breast cancer using novel deep learning architecture. Neural Comput Appl 34:16533–16545
83. Maria HH, Kayalvizhi R, Malarvizhi S, Venkatraman R, Patil S, Kumar AS (2023) Real-time deployment of BI-RADS breast cancer classifier using deep-learning and FPGA techniques. J Real-Time Image Proc 20:80
84. Hekal AA, Moustafa HE-D, Elnakib A (2023) Ensemble deep learning system for early breast cancer detection. Evol Intell 16:1045–1054
85. Ortega-Martorell S, Riley P, Olier I, Raidou RG, Casana-Eslava R, Rea M, Shen L, Lisboa PJG, Palmieri C (2022) Breast cancer patient characterisation and visualisation using deep learning and fisher information networks. Sci Rep 12:14004
86. Shimokawa D, Takahashi K, Oba K, Takaya E, Usuzaki T, Kadowaki M, Kawaguchi K, Adachi M, Kaneno T, Fukuda T, Yagishita K, Tsunoda H, Ueda T (2023) Deep learning model for predicting the presence of stromal invasion of breast cancer on digital breast tomosynthesis. Radiol Phys Technol 16:406–413
87. Mohapatra S, Muduly S, Mohanty S, Ravindra JVR, Mohanty SN (2022) Evaluation of deep learning models for detecting breast cancer using histopathological mammograms images. Sustain Oper Comput 3:296–302
88. Shimokawa D, Takahashi K, Kurosawa D, Takaya E, Oba K, Yagishita K, Fukuda T, Tsunoda H, Ueda T (2023) Deep learning model for breast cancer diagnosis based on bilateral asymmetrical detection (BilAD) in digital breast tomosynthesis images. Radiol Phys Technol 16:20–27
89. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP (2001) The digital database for screening mammography. In: Yaffe MJ (ed) Proceedings of the fifth international workshop on digital mammography. Medical Physics Publishing, pp 212–218
90. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL (2017) A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 4:170177
91. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I, Stamatakis E, Cerneaz N, Kok S, Taylor P, Betal D, Savage J (2015) Mammographic Image Analysis Society (MIAS) database v1.21. Apollo - University of Cambridge Repository. https://doi.org/10.17863/CAM.105113
92. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) INbreast: Toward a full-field digital mammographic database. Acad Radiol 19:236–248
93. Wang Q, Chen H, Luo G, Li B, Shang H, Shao H, Sun S, Wang Z, Wang K, Cheng W (2022) Performance of novel deep learning network with the incorporation of the automatic segmentation network for diagnosis of breast cancer in automated breast ultrasound. Eur Radiol 32:7163–7172
94. Yu FH, Miao SM, Li CY, Hang J, Deng J, Ye XH, Liu Y (2023) Pretreatment ultrasound-based deep learning radiomics model for the early prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Eur Radiol 33:5634–5644
95. Balaha HM, Saif M, Tamer A, Abdelhay EH (2022) Hybrid deep learning and genetic algorithms approach (HMB-DLGAHA) for the early ultrasound diagnoses of breast cancer. Neural Comput Appl 34:8671–8695
96. Fujioka T, Kubota K, Hsu JF, Chang RF, Sawada T, Ide Y, Taruno K, Hankyo M, Kurita T, Nakamura S, Tateishi U, Takei H (2023) Examining the effectiveness of a deep learning-based computer-aided breast cancer detection system for breast ultrasound. J Med Ultrason 50:511–520
97. Boulenger A, Luo Y, Zhang C, Zhao C, Gao Y, Xiao M, Zhu Q, Tang J (2023) Deep learning-based system for automatic prediction of triple-negative breast cancer from ultrasound images. Med Biol Eng Compu 61:567–578
98. Taleghamar H, Jalalifar SA, Czarnota GJ, Sadeghi-Naini A (2022) Deep learning of quantitative ultrasound multi-parametric images at pre-treatment to predict breast cancer response to chemotherapy. Sci Rep 12:2244
99. Atrey K, Singh BK, Bodhey NK, Bilas Pachori R (2023) Mammography and ultrasound based dual modality classification of breast cancer using a hybrid deep learning approach. Biomed Sig Process Control 86:104919

100. Cho SW, Baek NR, Park KR (2022) Deep learning-based multi-stage segmentation method using ultrasound images for breast cancer diagnosis. J King Saud Univ - Comput Inform Sci 34:10273–10292
101. Qi X, Yi F, Zhang L, Chen Y, Pi Y, Chen Y, Guo J, Wang J, Guo Q, Li J, Chen Y, Lv Q, Yi Z (2022) Computer-aided diagnosis of breast cancer in ultrasonography images by deep learning. Neurocomputing 472:152–165
102. Welhenge A (2022) Deep learning-based breast cancer detection system using fog computing. Journal of Discrete Mathematical Sciences and Cryptography 25:661–669
103. Huang Y, Yao Z, Li L, Mao R, Huang W, Hu Z, Hu Y, Wang Y, Guo R, Tang X, Yang L, Wang Y, Luo R, Yu J, Zhou J (2023) Deep learning radiopathomics based on preoperative US images and biopsy whole slide images can distinguish between luminal and non-luminal tumors in early-stage breast cancers. eBioMedicine 94:104706
104. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2020) Dataset of breast ultrasound images. Data Brief 28:104863
105. Guo Y, Xie X, Tang W, Chen S, Wang M, Fan Y, Lin C, Hu W, Yang J, Xiang J, Jiang K, Wei X, Huang B, Jiang X (2024) Noninvasive identification of HER2-low-positive status by MRI-based deep learning radiomics predicts the disease-free survival of patients with breast cancer. Eur Radiol 34:899–913
106. Li L, Zhou X, Cui W, Li Y, Liu T, Yuan G, Peng Y, Zheng J (2023) Combining radiomics and deep learning features of intra-tumoral and peri-tumoral regions for the classification of breast cancer lung metastasis and primary lung cancer with low-dose CT. J Cancer Res Clin Oncol 149:15469–15478
107. Zhang J, Yin W, Yang L, Yao X (2024) Deep learning radiomics nomogram based on multiphase computed tomography for predicting axillary lymph node metastasis in breast cancer. Mol Imag Biol 26:90–100
108. Koh J, Yoon Y, Kim S, Han K, Kim E-K (2022) Deep learning for the detection of breast cancers on chest computed tomography. Clin Breast Cancer 22:26–31
109. Vivancos Bargalló H, Stick LB, Korreman SS, Kronborg C, Nielsen MM, Borgen AC, Offersen BV, Nørrevang O, Kallehauge JF (2023) Classification of laterality and mastectomy/lumpectomy for breast cancer patients for improved performance of deep learning auto segmentation. Acta Oncol 62:1546–1550
110. Zhou Z, Adrada BE, Candelaria RP, Elshafeey NA, Boge M, Mohamed RM, Pashapoor S, Sun J, Xu Z, Panthi B, Son JB, Guirguis MS, Patel MM, Whitman GJ, Moseley TW, Scoggins ME, White JB, Litton JK, Valero V, Hunt KK, Tripathy D, Yang W, Wei P, Yam C, Pagel MD, Rauch GM, Ma J (2023) Prediction of pathologic complete response to neoadjuvant systemic therapy in triple negative breast cancer using deep learning on multiparametric MRI. Sci Rep 13:1171
111. Taheri S, Golrizkhatami Z (2023) Magnification-specific and magnification-independent classification of breast cancer histopathological image using deep learning approaches. SIViP 17:583–591
112. Chattopadhyay S, Dey A, Singh PK, Oliva D, Cuevas E, Sarkar R (2022) MTRRE-Net: a deep learning model for detection of breast cancer from histopathological images. Comput Biol Med 150:106155
113. Liu M, Hu L, Tang Y, Wang C, He Y, Zeng C, Lin K, He Z, Huo W (2022) A deep learning method for breast cancer classification in the pathology images. IEEE J Biomed Health Inform 26:5025
114. Muntean CH, Chowkkar M (2022) Breast cancer detection from histopathological images using deep learning and transfer learning. Proceedings of the 2022 7th international conference on machine learning technologies (ICMLT '22), pp 164–169. https://doi.org/10.1145/3529399.3529426
115. Tsafas V, Oikonomidis I, Gavgiotaki E, Tzamali E, Tzedakis G, Fotakis C, Athanassakis I, Filippidis G (2022) Application of a deep-learning technique to non-linear images from human tissue biopsies for shedding new light on breast cancer diagnosis. IEEE J Biomed Health Inform 26:1188
116. Toa CK, Elsayed M, Sim KS (2023) Deep residual learning with attention mechanism for breast cancer classification. Soft Comput. https://doi.org/10.1007/s00500-023-09152-2
117. Wetstein SC, de Jong VMT, Stathonikos N, Opdam M, Dackus GMHE, Pluim JPW, van Diest PJ, Veta M (2022) Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. Sci Rep 12:15102
118. Sheeba A, Kumar PS, Ramamoorthy M, Sasikala S (2023) Microscopic image analysis in breast cancer detection using ensemble deep learning architectures integrated with web of things. Biomed Sig Process Control 79:104048
119. Huang PW, Ouyang H, Hsu BY, Dah-Tsyr Chang YR, Lin YC, Chen YA, Hsieh YH, Fu CC, Li CF, Lin CH, Lin YY, Chang M, Pai TW (2023) Deep-learning based breast cancer detection for cross-staining histopathology images. Heliyon 9(2):e13171

120. Bhowal P, Sen S, Velasquez JD, Sarkar R (2022) Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification. Expert Syst Appl 190:116167
121. Yang J, Ju J, Guo L, Ji B, Shi S, Yang Z, Gao S, Yuan X, Tian G, Liang Y, Yuan P (2022) Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. Comput Struct Biotechnol J 20:333–342
122. Liu T, Huang J, Liao T, Pu R, Liu S, Peng Y (2022) A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. IRBM 43:62–74
123. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, Zarringhalam K (2022) Deep learning of longitudinal mammogram examinations for breast cancer risk prediction. Pattern Recogn 132:108919
124. Bychkov D, Joensuu H, Nordling S, Tiulpin A, Kücükel H, Lundin M, Sihto H, Isola J, Lehtimäki T, Kellokumpu-Lehtinen P-L, von Smitten K, Lundin J, Linder N (2022) Outcome and biomarker supervised deep learning for survival prediction in two multicenter breast cancer series. J Pathol Inform 13:100171
125. Çayır S, Solmaz G, Kusetogullari H, Tokat F, Bozaba E, Karakaya S, Iheme LO, Tekin E, Yazıcı Ç, Özsoy G, Ayaltı S, Kayhan CK, İnce Ü, Uzel B, Kılıç O (2022) MITNET: A novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue. Neural Comput Appl 34:17837–17851
126. Wu J, Xu Z, Shang L, Wang Z, Zhou S, Shang H, Wang H, Yin J (2023) Multimodal microscopic imaging with deep learning for highly effective diagnosis of breast cancer. Opt Lasers Eng 168:107667
127. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) A dataset for breast cancer histopathological image classification. IEEE Trans Biomed Eng 63:1455–1462
128. Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Marami B, Prastawa M, Chan M, Donovan M, Fernandez G, Zeineh J, Kohl M, Walz C, Ludwig F, Braunewell S, Baust M, Vu QD, To MNN, Kim E, Kwak JT, Galal S, Sanchez-Freire V, Brancati N, Frucci M, Riccio D, Wang Y, Sun L, Ma K, Fang J, Kone I, Boulmane L, Campilho A, Eloy C, Polónia A, Aguiar P (2019) BACH: grand challenge on breast cancer histology images. Med Image Anal 56:122–139
129. Ensafi M, Keyvanpour MR, Shojaedini SV (2022) A new method for promoting the performance of deep learning paradigm in diagnosing breast cancer: improving role of fusing multiple views of thermography images. Health Technol 12:1097–1107
130. Tsietso D, Yahya A, Samikannu R, Tariq MU, Babar M, Qureshi B, Koubaa A (2023) Multi-input deep learning approach for breast cancer screening using thermal infrared imaging and clinical data. IEEE Access 11:52101–52116
131. Al Husaini MAS, Habaebi MH, Gunawan TS, Islam MR, Elsheikh EA, Suliman F (2022) Thermal-based early breast cancer detection using Inception V3, Inception V4 and modified Inception MV4. Neural Comput Appl 34:333–348
132. Chatterjee S, Biswas S, Majee S, Sen A, Oliva S, Sarkar D, Ram R (2022) Breast cancer detection from thermal images using a Grunwald-Letnikov-Aided dragonfly algorithm-based deep feature selection method. Comput Biol Med 141:105027
133. da Silva LF, Saade DCM, Sequeiros GO, Silva AC, de Paiva AC, R. de S. Bravo, A. Conci, (2014) A new database for breast research with infrared image. J Med Imaging Health Infor 4(1):92–100
134. Sharma N, Sharma KP, Mangla M, Rani R (2023) Breast cancer classification using snapshot ensemble deep learning model and t-distributed stochastic neighbor embedding. Multimedia Tools and Applications 82:4011–4029
135. Kayikci S, Khoshgoftaar TM (2023) Breast cancer prediction using gated attentive multimodal deep learning. Journal of Big Data 10:62
136. Awotunde JB, Panigrahi R, Khandelwal B, Garg A, Bhoi AK (2023) Breast cancer diagnosis based on hybrid rule-based feature selection with deep learning algorithm. Research on Biomedical Engineering 39:115–127
137. Gonzales Martinez R, van Dongen DM (2023) Deep learning algorithms for the early detection of breast cancer: a comparative study with traditional machine learning. Inform Med Unlocked 41:101317
138. Xiong Y, Wang S, Wei H, Li H, Lv Y, Chi M, Su D, Lu Q, Yu Y, Zuo Y, Yang L (2022) Deep learning-based transcription factor activity for stratification of breast cancer patients. Biochim Biophys Acta Gene Regul Mech 1865:194838
139. Kirelli Y, Arslankaya S, Koçer HB, Harmantepe T (2023) CNN-based deep learning method for predicting the disease response to the neoadjuvant chemotherapy (NAC) treatment in breast cancer. Heliyon 9(6):e16812

140. Zhao S, Yan C-Y, Lv H, Yang J-C, You C, Li Z-A, Ma D, Xiao Y, Hu J, Yang W-T, Jiang Y-Z, Xu J, Shao Z-M (2022) Deep learning framework for comprehensive molecular and prognostic stratifications of triple-negative breast cancer. Fundam Res 4(3):678
141. Jadoon EK, Khan FG, Shah S, Khan A, ElAffendi M (2023) Deep learning-based multi-modal ensemble classification approach for human breast cancer prognosis. IEEE Access 11:85760–85769
142. Arya N, Saha S (2022) Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. IEEE/ACM Trans Comput Biol Bioinf 19:1032–1041
143. Guan M, Qi G, Li Z, Hou X (2023) Potential targets of endocrine-disrupting chemicals related to breast cancer identified by ToxCast and deep learning models. Toxicol Environ Chem 105:128–143
144. Wolberg WH, Street WN, Heisey DM, Mangasarian OL (1995) Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. Arch Surg 130:511–516
145. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Caldas C, Aparicio S, Brenton JD, Ellis I, Huntsman D, Pinder S, Purushotham A, Murphy L, Bardwell H, Ding Z, Jones L, Liu B, Papatheodorou I, Sammut SJ, Wishart G, Chia S, Gelmon K, Speers C, Watson P, Blamey R, Green A, Macmillan D, Rakha E, Gillett C, Grigoriadis A, de Rinaldis E, Tutt A, Parisien M, Troup S, Chan D, Fielding C, Maia A-T, McGuire S, Osborne M, Sayalero SM, Spiteri I, Hadfield J, Bell L, Chow K, Gale N, Kovalik M, Ng Y, Prentice L, Tavaré S, Markowetz F, Børresen-Dale AL (2012) METABRIC Group, Writing Committee, Steering Committee, Tissue and Clinical Data Source Sites, University of Cambridge/Cancer Research UK Cambridge Research Institute, British Columbia Cancer Agency, University of Nottingham, King's College London, Manitoba Institute of Cell Biology, Cancer Genome/Transcriptome Characterization Centres, Data Analysis Subgroup, The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352