Artificial Intelligence Review (2024) 57:327
https://doi.org/10.1007/s10462-024-10984-z

A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification

Fatima-Zahrae Nakach[1] · Ali Idri[2] · Evgin Goceri[3]

Accepted: 26 September 2024 / Published online: 12 October 2024

Abstract

In breast cancer research, diverse data types and formats, such as radiological images, clinical records, histological data, and expression analysis, are employed. Given the intricate nature of natural phenomena, relying on the features of a single modality is seldom sufficient for comprehensive analysis. Therefore, it is possible to guarantee medical relevance and achieve improved clinical outcomes by combining several modalities. The presen study carefully maps and reviews 47 primary articles from six well-known digital libraries that were published between 2018 and 2023 for breast cancer classification based on multimodal deep learning fusion (MDLF) techniques. This systematic literature review encompasses various aspects, including the medical modalities combined, the datasets utilized in these studies, the techniques, models, and architectures used in MDLF and it also discusses the advantages and limitations of each approach. The analysis of selected papers has revealed a compelling trend: the emergence of new modalities and combinations that were previously unexplored in the context of breast cancer classification. This exploration has not only expanded the scope of predictive models but also introduced fresh perspectives for addressing diverse targets, ranging from screening to diagnosis and prognosis. The practical advantages of MDLF are evident in its ability to enhance the predictive capabilities of machine learning models, resulting in improved accuracy across diverse applications. The prevalence of deep learning models underscores their success in autonomously discerning complex patterns, offering a substantial departure from traditional machine learning approaches. Furthermore, the paper explores the challenges and future directions in this field, including the need for larger datasets, the use of ensemble learning methods, and the interpretation of multimodal models.

Keywords Breast cancer · Classification · Fusion · Multimodality · Deep learning · Review

Extended author information available on the last page of the article

1 Introduction

Breast cancer (BC) is characterized by the uncontrolled growth of malignant cells in breast tissue; this disease is widespread and life-altering (Arnold et al. 2022). It not only affects millions of women worldwide but also poses complex challenges in terms of diagnosis and treatment (García-Aranda et al. 2019). One of the main challenges in addressing BC is the observation that patients with similar disease characteristics often exhibit varying responses to treatment and overall prognoses (Zhang et al. 2018). It is essential for healthcare providers to make informed decisions and enable personalized, effective, and patient-centered care to improve treatment success rates and enhance the overall quality of life for individuals affected by BC (Lu et al. 2009). Artificial intelligence (AI) has emerged as a powerful tool for predicting BC (Bahl and Bahl 2022; Sugimoto 2023). By leveraging advanced machine learning (ML) algorithms and analyzing a vast array of patient data, AI systems can significantly improve cancer diagnosis, treatment, and prevention (Hanahan et al. 2011). The mortality rates associated with BC could decrease by offering physicians accurate risk prediction models to assist in directing patient treatment and management approaches (Yassin et al. 2018). The advent of deep learning (DL) techniques and the availability of multi-dimensional data have presented promising prospects for conducting a more thorough analysis of the molecular attributes associated with BC (Romeo et al. 2021). Consequently, these advancements have the potential to improve the accuracy and effectiveness of BC diagnosis, treatment, and prevention using different modalities, including gene expression, clinical records, and medical images (James et al. 2014). Although one of those modalities alone is not accurate enough, their combination can significantly improve diagnostic accuracy and treatment decisions due to their complementary nature (Yuan et al. 2010).

Recently, multimodal data fusion based on DL has undergone rapid growth in the development of healthcare AI systems (Lahat et al. 2015). Unimodal systems, in reality, have limitations in dealing with nonuniversal, distinct, and noisy data. On the other hand, multimodal architectures address these limitations by integrating pertinent features from various sources (Jain and Ross 2004; Wang et al. 2009). Multimodal DL has shown potential across various tasks and domains, with applications extending to the medical field for the diagnostic and prognostic objectives of cancer. However, the intricate nature of medical modalities, each characterized by its own distinct set of features, shapes, and dimensions, leads to the generation of high-dimensional data that represent a complex interplay of information and a significant hurdle in the implementation of multimodal DL (Metzger-Filho et al. 2013). The main challenge lies in identifying the optimal combination of medical modalities, feature processing methods, feature extraction techniques, and decision-fusion algorithms tailored to address a particular clinical issue. Moreover, themes such as interpretability and explainability are emerging fields of research and hold significant relevance within the realm of DL (Ito 2018). For clinical applications, interpretability is fundamental since physicians need to confirm the predictions of AI models by verifying whether the rationale behind them aligns with established medical knowledge (Brito-Sarracino et al. 2019).

Recognizing the pivotal role of developing and evaluating AI systems that seamlessly integrate different medical modalities, researchers have acknowledged the paramount significance of these systems, leading to a series of comprehensive reviews (Huang et al. 2020; Lipkova et al. 2022; Stahlschmidt et al. 2022; Salvi et al. 2024; Steyaert et al. 2023; Pei et al. 2023). These reviews, conducted by esteemed scholars in the field, aim to elucidate the current state of knowledge, identify gaps, and provide insights into the evolving landscape of multimodal ML applied to medicine, either in a general context or specifically in oncology. Consequently, the primary objective of the present review is to refine the focus by concentrating on BC, where multimodal fusion techniques have been applied for clinical decision support based on classification. This undertaking involves conducting a Systematic Literature Review (SLR) with a more specific and focused scope. This approach allows for a more in-depth analysis of the applications, challenges, and implications of BC within its context, resulting in a review that is targeted, valuable, and contextually relevant.

This paper explores the primary studies published in six digital libraries—ScienceDirect, SpringerLink, Wiley, IEEE Xplore, ACM Digital Library, and Google Scholar—until December 2023 and presents a comprehensive review in the field of multimodal DL fusion applied for BC classification, emphasizing its pivotal role as a ML objective. The classification task serves as a versatile tool capable of addressing multiple facets within BC research and clinical practice. It enables the refinement of screening methods, enhancement of diagnostic accuracy across diverse subtypes, prediction of prognosis, and customization of treatment strategies. This review provides a synthesis of the available research with the following specific aims:

- Identifying papers investigating multimodal DL fusion for BC classification.
- Specifying the types of data and sources commonly used.
- Enumerating the medical tasks, classification target and the combinations of modalities that are most frequently addressed.
- Describing the fusion strategies and concatenation techniques employed.
- Exploring the application of DL and ML models.
- Identifying he most commonly used validation methods and metrics for evaluation.
- Assessing the performance of multimodal DL models.
- Comparing the performance of models based on multimodal data and on an individual modality.
- Identifying the strengths and weaknesses of each fusion approach.
- Investigating the use of Explainable AI (XAI) in the selected papers.

The present paper is structured as follows: Section 2 presents a brief background on multimodal DL fusion and covers the related reviews. Section 3 explains the research strategy used in this SLR. Section 4 reports the statistical trends of the selected papers. Section 5 is dedicated to the datasets and modalities fused in the selected studies. In Section 6, an extensive examination of approaches to multimodal fusion is provided, along with a detailed evaluation of the concatenation methods and ML/DL models employed. Section 7 outlines the patterns discovered in the fused features learned from the studies that were selected. In Section 8, a thorough analysis is presented, discussing the strengths and weaknesses of each multimodal fusion approach and providing a comparative assessment of the selected works. Section 9 elaborates on the performance metrics utilized for evaluating multimodal fusion techniques. Section 10 details the performance of the multimodal fusion models and compares them with the best model based on a single modality. Section 11 offers practical guidance and implications for researchers, emphasizing key takeaways and providing actionable recommendations for future investigations. Finally, Section 12 concludes the paper and outlines potential directions for future research.

2 Background and related work

This section provides an overview of multimodal DL fusion and its relevance in BC research. It also summarizes related reviews to highlight key insights and identify gaps in current knowledge.

2.1 Multimodal fusion taxonomy

The accuracy of BC classification using a single modality falls short of meeting therapeutic requirements; due to the complexity of natural factors, a single modality struggles to provide comprehensive information for analysis (Abhisheka et al. 2023). Consequently, leveraging multimodal data offers distinct advantages for intricate analyses (Jain and Ross 2004). At the heart of multimodal ML is multimodal data fusion, a method aimed at amalgamating data from diverse distributions, sources, and types into a unified space capable of enhancing clinical accuracy (Stahlschmidt et al. 2022). The taxonomy of multimodal data fusion is subject to considerable variation from one scholarly work to another, often leading to confusion and ambiguity in establishing a uniform nomenclature for its diverse types. The absence of standardized terminology across the literature makes it challenging to consistently categorize and define different fusion approaches, particularly when specific details are not explicitly articulated in the papers. In our investigation of multimodal ML fusion techniques, we underscore a paper (Holste et al. 2021) that criticizes an existing survey's taxonomy (Baltrušaitis et al. 2019), deeming it limited in expressive power. This paper (Huang et al. 2020) proposes an alternative naming scheme based on a more recent review of techniques, specifically focused on fusing medical imaging with tabular clinical data. Figure1summarizes the different taxonomies of multimodal data fusion we identified in the literature. By adhering to the nomenclature introduced in Huang et al. (2020) and Holste et al. (2021), we aim to provide a more nuanced and expressive classification of multimodal fusion methods, emphasizing the nature of the features involved and the fusion methodology employed. This process encompasses three distinct fusion approaches:

- **Decision fusion**, also known as late fusion or probability fusion, refers to the process of utilizing predictions from multiple models to generate a final prediction. This procedure employs distinct models that are trained using various modalities, and a consolidation function merges their decisions (D). The available functions include averaging, majority voting, weighted voting, or utilizing a meta-classifier that relies on model predictions. The selection of the aggregation function typically relies on empirical evaluation and may differ based on the application and input modalities (Huang et al. 2020).
- **Feature fusion** refers to the process of combining various input modalities into a single feature vector, which is subsequently fed into a single model for training. The process of integrating these input modalities may involve various techniques, such as concatenation, pooling, or the utilization of a gated unit (Huang et al. 2020). Within feature fusion, subtypes emerge based on the nature of the features:
  - Semantic features (S): Represents the raw, original data.
  - Learned Features (L): Extracted features based on a DL model.

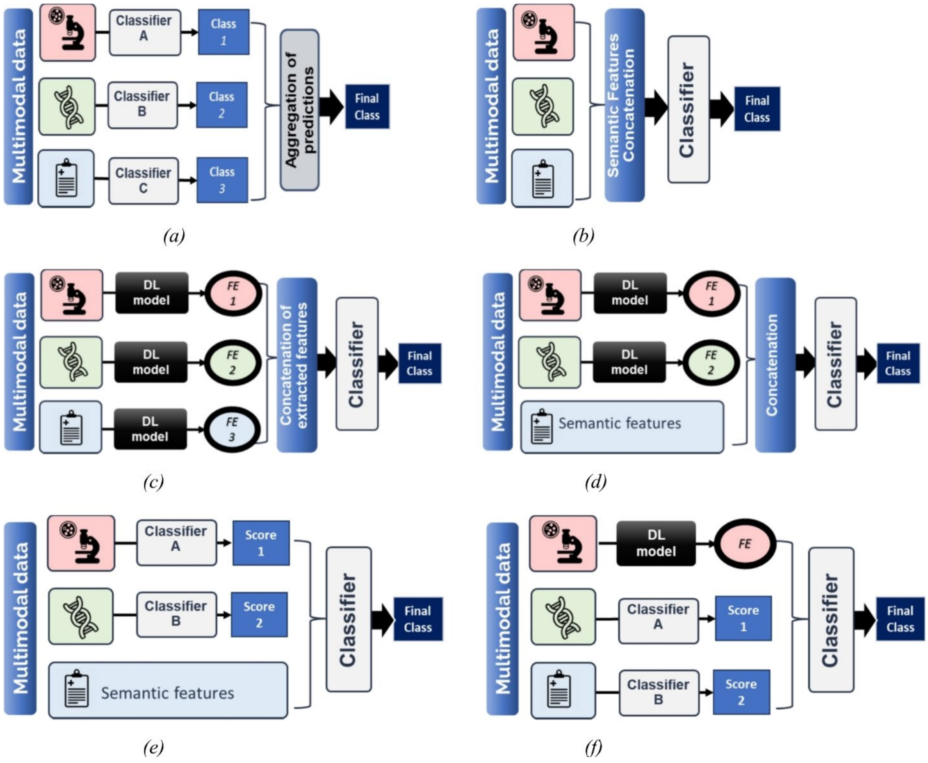

Fig. 1 Multimodal fusion approaches: **a** decision-fusion (D+D), **b** semantic feature-fusion (S+S), (**c**earned feature-fusion (L+L), **d** semantic and learned feature-fusion (L+S), **e** hybrid-fusion (D+S) and **f** hybrid-fusion (D+L)

When semantic features are fused, the fusion approach is referred to as "early fusion." When the features are learned, the fusion is known as "joint fusion."

– **Hybrid fusion** combines decision fusion and feature fusion approaches.

This process results in six possible fusion scenarios: S+S, L+L, D+D, D+L, D+S, and L+S, as illustrated in Fig. 1.

2.2 Multimodal deep learning

Multimodal representation learning challenges have been effectively addressed by DL, which uses a common approach to learn joint representations shared between modalities on top of layers of modality-specific networks. In the DL literature, researchers often focus on innovating by introducing new layers, optimization techniques, or novel methods tailored to specific data types (Akkus et al. 2023). However, in the field of biomedical AI, the emphasis is primarily on determining the most suitable architecture for the task at hand. Researchers are attempting to experiment with the number and configuration of layers and explore different loss functions. This challenge intensifies when addressing BC classification using multiple modalities, necessitating the fusion of various data types. Consequently, researchers find themselves employing diverse networks and integrating them at different levels. This subsection reviews the most impactful models utilized in the reviewed works, facilitating a quicker understanding of new research and, ideally, aiding them in developing models tailored to their specific tasks. New developments in DL are highly important for making diagnoses more accurate and faster:

– Deep Neural Networks

A Deep Neural Networks (DNN) is a form of artificial neural network that consists of multiple layers, such as an input layer, one or more hidden layers, and an output layer. The term "deep" denotes the existence of numerous concealed layers, enabling the network to acquire hierarchical representations from the input data. DNNs can automatically extract complex features and patterns. This makes them highly suitable for tasks such as image recognition, natural language processing, and regression analysis (Abdou and Abdou 2022). A concrete instance of a distinct DNN structure is the multilayer perceptron (MLP), which comprises three or more tiers of nodes, encompassing an input tier, one or more concealed tiers, and an output tier. The network establishes connections between each node in one layer and every node in the following layer while incorporating activation functions to introduce nonlinear behavior. MLPs are commonly used for supervised learning tasks such as classification and regression, and they often incorporate fully connected layers (FCLs) to facilitate information flow between layers (Taud et al. 2018). Notably, FCL and MLP emerged as the most commonly used models for multimodal feature classification across the selected studies. This indicates the strategic utilization of DNN classifiers for feature fusion when combining information from various modalities, and the prevalence of FCL in multimodal feature classification underscores its effectiveness in capturing intricate relationships within diverse datasets.

– Convolutional Neural Network

Convolutional Neural Networks (CNNs) are the most commonly used DL networks for BC detection and use multiple modalities. They have greatly advanced the field of computer vision by greatly enhancing the precision of image recognition tasks. CNNs are highly proficient at capturing hierarchical representations of visual features, which makes them particularly suitable for tasks such as object detection, segmentation, and classification (Battleday et al. 2021). CNNs have demonstrated significant efficacy not only in image classification tasks but also as reliable learners of representations. They have found utility in various domains, including image enhancement and medical diagnosis (Zerouaoui et al. 2021). The essential elements of a CNN consist of convolutional layers, pooling layers, and fully connected layers. Convolutional layers employ filters or kernels to execute convolutions, extracting localized patterns and features from the input data. Pooling layers decrease the spatial dimensions of the data by down-sampling, aiding in preserving crucial information while decreasing computational complexity. Fully connected layers establish connections between every neuron in one layer and every neuron in the subsequent layer, allowing the network to generate predictions by utilizing the acquired features (Alzubaidi et al. 2021). Beyond their ability to extract features from image data, CNNs have emerged as popular choices for information fusion in prediction tasks, especially at the feature level (Guo et al. 2018). This popularity can be attributed to the quality of the features they generate. However, CNNs must confront significant performance challenges when confronted with distributional shifts in datasets. In the realm of medical prediction, addressing limited dataset challenges is vital, and many studies advocate for pretraining networks as a crucial strategy to overcome these limitations.

- Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a class of neural networks that possess a form of memory, allowing them to access previous information during processing. Unlike typical neural networks where inputs and outputs are independent, RNNs can leverage their memory to consider past data, albeit within a limited temporal scope. However, RNNs face challenges, particularly short-term memory issues, as they struggle to retain information over extended sequences (Yin et al. 2017). To address this limitation, long short-term memory (LSTM) networks were introduced as a subclass of RNNs, and LSTMs exhibit a distinct advantage in their ability to overcome short-term memory constraints and facilitate long-term dependency learning (Wu et al. 2019). This feature makes LSTMs particularly effective in tasks requiring an understanding of intricate relationships across extended sequences. Gated recurrent units (GRUs), like LSTMs, belong to the category of recurrent neural networks and were specifically designed to address the vanishing gradient problem (Dey 2017). GRUs employ three primary gates and an internal cell state, making them more efficient than LSTMs in certain contexts. GRUs outperform LSTMs due to parameter reduction, resulting in a higher convergence rate and requiring less computational time. The simplicity of the GRU structure, compared to that of LSTMs, contributes to reduced matrix multiplication, saving time without compromising performance (Dey 2017).

- Autoencoders

An autoencoder is an unsupervised neural network composed of an encoder and a decoder. It learns to create a low-dimensional representation of data and reconstruct the original input data. In medical imaging, AEs are employed for representation learning and tasks such as denoising and compression (Li et al. 2022). Across the selected studies, three types of autoencoders were employed:

- *Denoising autoencoder (DAE)* A denoising autoencoder is a variant of the basic autoencoder. It shares the same structure but is trained on noisy versions of the raw data. This approach enhances the robustness of feature representation by preventing the model from simply learning identity function mapping (Gondara 2016).
- *Conditional autoencoder* A conditional autoencoder is designed to extract related features from different modalities. This approach typically involves the introduction of specific encoders, such as a gene encoder for dimension reduction, and aims to perform correlated feature extraction. This type of autoencoder is commonly used for multisource data fusion and feature extraction (Choi 2019).

- *Variational autoencoder (VAE)* Variational autoencoders are employed as a dimensionality reduction technique. They introduce explicit regularization during training, ensuring the regularity of the latent space. VAEs encode raw features as a normal distribution over latent space and use the Kullback–Leibler divergence to enforce the encoder to return a distribution resembling a standard normal distribution (Cheng et al. 2021).

- Generative Adversarial Networks.

Generative Adversarial Networks (GANs) hold a distinctive position among researchers owing to their ability to produce high-quality generated data (Goodfellow et al. 2020). The fundamental concept behind GANs involves the simultaneous training of two models—a discriminator and a generator. The discriminator aims to distinguish between real and generated images, while the generator endeavors to deceive the discriminator by producing realistic synthetic data. Along with their variants, such as CycleGAN (Chu 2017)and WGAN (Weng and From 2019), GANs have found widespread applications in medical diagnosis, contributing to tasks such as data augmentation, representation learning, and image enhancement. Notably, GANs have achieved significant advancements by utilizing adversarial training, allowing them to map one distribution to another (Logan et al. 2021). Their utilization has significantly benefited various research disciplines by introducing innovative methods for data generation and augmentation. However, as the dimensionality of the data increases, the quality of the generated synthetic data may also decrease (Goodfellow et al. 2020).

- Graph Neural Networks.

Graphs are fundamental data structures representing sets of interconnected objects (nodes) and their relationships (edges). In the realm of ML, the analysis of graphs has gained substantial attention due to their exceptional expressive power, which has made them versatile representations of various systems across disciplines such as social science, natural science, and knowledge graphs. In particular, graph neural networks (GNNs) are neural models explicitly designed to comprehend dependencies within graph-structured data through messages passing among nodes (Zhou et al. 2020). As an influential methodology in recent years, GNNs, including variants such as graph convolutional networks (GCNs), graph attention networks (GATs), and graph recurrent networks (GRNs), have demonstrated breakthrough performances in diverse learning tasks. These tasks span from modeling physical systems to predicting molecular fingerprints and classifying diseases, demonstrating the adaptability of GNNs.

2.3 Related work

Several review papers have been published on the application of ML and DL in BC research (Yassin et al. 2018; Abhisheka et al. 2023; Thakur et al. 2024). These reviews provide valuable insights into the advancements and trends in the field. However, they lack a comprehensive analysis of recent developments of multimodal DL fusion, since multimodal data fusion is not the primary concern of these reviews. Additionally, the existing SLRs of BC have explored various modalities but often without integrating them through fusion techniques (Murtaza et al. 2020; Nassif et al. 2022; Mahmood et al. 2020; Madani et al. 2022). Our review aims to address this gap by specifically examining the fusion of multiple modalities in the context of BC, providing a more comprehensive understanding of the advancements and challenges in this specific domain. Table 1 summarizes the contributions and limitations of these existing review papers, underscoring the necessity for this SLR to provide a more holistic and updated overview.

3 Research methodology

The present mapping and review process follows the guidelines set forth by Kitchenham and Charters (2007) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al. 2021); they consist of six steps, as shown in Fig. 2. We began the process by creating a series of mapping and review questions. Subsequently, we established the search query and discerned pertinent sources. The selection phase entailed the application of specific criteria to include or exclude papers, ensuring that only relevant papers were chosen for the current review. Subsequently, we evaluated the caliber of the chosen papers by employing a questionnaire and scoring mechanism. Ultimately, information was extracted from these papers and combined to generate answers to the mapping and review inquiries. Each of these steps is further explained in the following sections.

3.1 Mapping and review questions

The mapping questions (MQs) aim to identify, describe, and classify the primary studies published in the field of multimodal fusion for BC classification. The review questions (RQs) further analyze and synthesize these studies to gain insights into the field. The ten questions (six RQs and four MQs) are listed in Table 2, along with their respective motivations. By addressing these MQs and RQs, this review seeks to provide a thorough understanding of the current state of multimodal fusion based on DL for BC classification.

3.2 Search strategy

To respond to the MQs and RQs, we implemented a four-stage search procedure. (1) We developed a search query from the earliest available studies up to December 2023; (2) This query was used in an automated search across six chosen digital libraries to find primary papers; (3) As part of a second search, the bibliographies of relevant papers (those meeting certain inclusion and exclusion criteria) were examined to ensure that all relevant literature was included; and (4) Finally, the same query was used in a third search to find newly published articles while the search string was constructed by employing key terms and their synonyms derived from the review questions. Boolean AND was used to connect essential components, while Boolean OR was employed to link alternative words. The final search string was formulated as follows:

(Breast OR "Mammary gland") AND (cancer* OR tumor OR malignancy OR masses) AND ("machine learning" OR "deep Learning" OR "artificial intelligence" OR prediction

Table 1 Summary of review papers published on multimodal DL for BC classification

Paper	Scope	Contributions	Limitations
(Luo et al. 2024)	The paper provided an extensive review of DL-based BC imaging research over the past decade.	-Introduction of widely applied DL paradigms, including supervised learning, semi-supervised learning, weakly-supervised learning, unsupervised learning, transfer learning, and multimodal learning.	-Not a systematic review. -Not all the selected studies have fully utilized the comprehensive context provided by multimodal data fusion. -Fused modalities, predominantly focused on imaging modalities, without encompassing clinical data or genomics.
(Mathur et al. 2024)	The review presented a comprehensive survey of the current state-of-the-art in BC detection and prognosis, emphasizing the shift from unimodal to multimodality information.	-Various topics are discussed including available databases, feature selection, dimensionality reduction, variations in survival prediction windows, handling of minority classes, and the utility of multimodal approaches over single modalities like genomics.	-Not a systematic review. -The review primarily focused on prognosis, particularly molecular subtype and survival prognosis predictions, which may limit the scope of its applicability to other aspects of BC research.
(Abdullakutty et al. 2024)	The review explored the integration of histopathology images with non-image data and the application of XAI to elucidate decision-making processes in BC diagnosis.	- A detailed investigation of multi-modal datasets incorporating histopathology and non-image data, often overlooked in existing literature. - Discussion on multi-modal techniques using these datasets, providing insights into their application and effectiveness in BC diagnosis. - An investigation of explainable multimodal methods within histopathology-based BC diagnosis, addressing a critical research gap. - Identification of research gaps in multi-modality and explainability, guiding future studies and contributing to the strategic development of the field.	-Not a systematic review. -The review focused solely on whole slide imaging (WSI) histopathology as the imaging modality fused with clinical or genomic data.
(Thakur et al. 2024)	The authors analyzed various ML and DL techniques in the context of BC. The study explored the classification of BC using different image modalities, alongside discussions on diagnosis methodologies utilizing publicly and privately available datasets, pre-processing techniques, and feature extraction methods.	-Discussion of the future potential of BC classification through multi-modal integration. -Recognition of leveraging complementary information from different modalities as a key area of ongoing research in the field of BC detection.	-The SLR neglected papers that integrate multiple modalities, and only considered papers that use single imaging modalities separately.

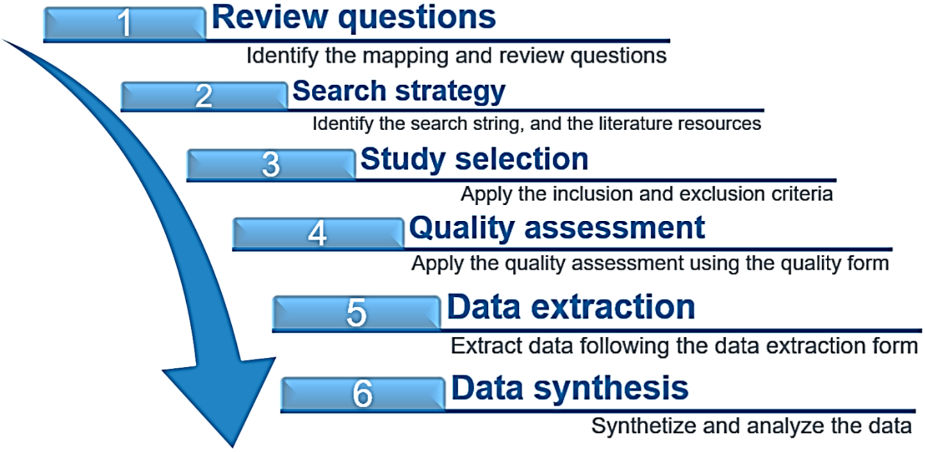
1
Review questions
Identify the mapping and review questions
2
Search strategy
Identify the search string, and the literature resources
3
Study selection
Apply the inclusion and exclusion criteria
4
Quality assessment
Apply the quality assessment using the quality form
5
Data extraction
Extract data following the data extraction form
6
Data synthesis
Synthetize and analyze the data

Fig. 2 Mapping and review process

Table 2 The mapping/review questions and their motivations

Mapping/Review Questions	Motivations
MQ1: What were the publication years, publication channels, and sources of the selected papers in the field of multimodal ML for BC?	To ascertain the presence of a particular publication channel and determine the frequency of studies on multimodal fusion for BC over the years.
MQ2: What are the types of papers that use multimodal ML for BC classification?	To categorize the various studies that have focused on multimodal fusion and analyze the specific evidence that has been generated in these selected studies.
MQ3: What specific medical tasks and challenges were addressed in the selected papers?	To determine the specific medical tasks for BC in which researchers utilized multimodal DL techniques and expressed interest.
MQ4: Which classification target were most frequently investigated in the context of multimodal ML for BC?	To discover the most investigated classification target in the literature of multimodal fusion for BC.
RQ1: What were the attributes and structures of the multimodal datasets employed for assessing and validating the explored multimodal fusion techniques in BC research?	-To enumerate the different multimodal datasets used for BC, and identify the most commonly utilized modalities and combinations.
RQ2: What were the key multimodal fusion approaches for integrating diverse modalities in the context of BC classification?	To determine the multimodal fusion approaches that were predominantly employed for BC classification.
RQ3: What techniques were predominantly employed for fusing the information from diverse modalities in multimodal ML studies focused on BC?	To determine the aggregation methods, ML and DL models used for the fusion of diverse BC modalities and identify the ones that had the highest interest.
RQ4: What are the validation methods used to measure the performance for the classification of BC using multimodal fusion?	To define the measures/metrics and validation methods used to evaluate the multimodal fusion techniques.
RQ5: Did the reviewed studies propose any interpretability methods to enhance understanding of multimodal ML models in the context of BC?	To discover if interpretability was considered when fusing different medical modalities for the classification of BC.
RQ6: What is the overall performance of the multimodal fusion techniques for BC classification?	To report the performance of multimodal fusion techniques and compare it with the performance on models relying on a single modality

OR classification) AND (modalities OR multimodality OR multimodal OR multimodalities OR fusion).

We underscore that our primary goal was to survey the literature related to the utilization of various modalities in the context of BC, aiming for a broad and comprehensive selection of candidate papers. This is why the search string incorporates terms such as "modalities" and "fusion." In the selection process, we applied inclusion and exclusion criteria to filter studies, specifically focusing on those where diverse modalities were integrated to formulate the ultimate ML prediction rather than being individually evaluated within the same study. We conducted an automated search using the specified search string across various digital libraries (ScienceDirect, ACM Digital Library, Wiley, IEEE Xplore, SpringerLink, and Google Scholar). These libraries encompass millions of articles across various publication channels within the field of computer science. The selection of search terms was tailored to the characteristics of each electronic database's search engine. Furthermore, we had the choice to incorporate papers based on the authors' personal awareness, under the condition that these papers had not yet been recognized among the candidate studies obtained from digital libraries.

3.3 Study selection procedure

In this phase, relevant studies that addressed the RQs were selected from a group of potential studies based on their titles, abstracts, and keywords. To achieve this goal, every candidate study identified in the initial search was subjected to evaluation by the authors. The authors employed inclusion and exclusion criteria to ascertain whether a study should be retained or discarded. If the title and/or abstract did not provide a definitive answer, a thorough analysis of the entire paper was conducted. The inclusion and exclusion criteria were connected using the OR Boolean operator. Inclusion was granted if a paper met at least one inclusion criterion, while exclusion resulted from satisfying any exclusion criterion. If both researchers classified a paper as retained, it was considered relevant.

Inclusion criteria (ICs):

1. This paper provides a summary of current multimodal fusion techniques employed for BC classification.
2. Papers presenting an improvement of existing multimodal fusion techniques for BC classification.
3. Papers introducing novel multimodal fusion techniques applied to the classification of BC.
4. Papers assessing or contrasting existing multimodal fusion techniques applied to BC classification.

Exclusion criteria (ECs):

1. Papers dealing with other types of cancer (not focusing on BC).
2. Different modalities are represented using a single type (as an example, the features of different imaging modalities are extracted manually and given in a tabular format).
3. Papers that evaluate each modality separately, using a separate DL model for each modality.

4. Studies that address multi-omics data only.
5. Duplicated papers.
6. Written in a language other than English.
7. Short deal abstract paper.
8. Preprints or posters.

3.4 Study quality assessment

To enhance the selection criteria and ensure the pertinence of the papers, we devised a questionnaire to evaluate the quality of the 53 relevant papers. The quality assessment process was applied to each paper, following the checklist specified in Table 3. The questionnaire consisted of four questions designed to evaluate the pertinent papers. The evaluation of QA1 is contingent upon the pertinence of the empirical findings presented in the paper to address the research inquiries. The evaluation of QA2 is contingent upon the level of transparency exhibited in the empirical design and methodologies utilized during the experiment. The evaluation of the study's outcomes in QA3 relies on the utilization of suitable performance metrics. QA4 focuses on evaluating the ranking of the paper. Conferences were assessed using the Computing Research and Education Association of Australasia (CORE Conference Ranking Exercise 2023), while journals were evaluated using Journal Citation Reports (JCR 2022). The authors autonomously performed the quality assurance process, and any inconsistencies were resolved through a collaborative meeting to arrive at a definitive conclusion. The quality score for each pertinent study was calculated by summing the scores of the quality assurance questions. A study was selected if its quality score exceeded 3.5, which is 50% of the ideal quality score of 7. This criterion was used to ensure the reliability and strength of the study's findings. Comparable checklists were utilized in ElOuassif et al. (2021).

Table 3 Quality assessment checklist

ID	Questions	Answers
QA1	Are the empirical findings presented in the study clearly articulated?	Yes (+1)/No (0)
QA2	Does the study exhibit a well-justified empirical design?	Yes (+1)/No (0)
QA3	Has the performance of the developed solution been thoroughly assessed in the study?	Yes (+1)/No (0)
QA4	Is the study published in a reputable and recognized source?	For conference: (1.5) Rank CORE A or A* (1) Rank Core B (0.5) Rank Core C (0) if not Core ranking For journals: (2) Rank JCR 2023 Q1 (1.5) Rank JCR 2023 Q2 (1) Rank JCR 2023 Q3 or Q4 (0) if not in JCR 2023 ranking

3.5 Data extraction strategy and synthesis

To handle the mapping and review inquiries, a data extraction form was created and completed for each of the chosen papers. The data obtained from these studies are presented in Table 4. The data extraction process consisted of two stages: first, the primary author diligently extracted pertinent data from each selected study by thoroughly examining the complete text; second, the remaining two authors verified the extracted data. Disagreements were resolved by engaging in collaborative discussions among the researchers. Following the extraction of the data, a synthesis was conducted, and the data were organized in a way that corresponded to the research questions to gather evidence for their resolution. Three methodologies were utilized for data synthesis: (1) vote counting, which entailed calculating the frequency of different outcomes across the chosen studies; (2) narrative synthesis, which involved creating a descriptive summary of the findings from the selected papers; and (3) visualization tools such as bar charts, funnel plots, pie charts, and scatter plots. Crucially, the reciprocal translation technique was found to be useful in examining and combining qualitative data obtained from specific papers. This approach involves the translation of the primary concepts or themes documented in various studies to ascertain any similarities or discrepancies among them.

Table 4 Extracted data

MQ/RQ	Data extracted
–	Authors, title, digital library, abstract
MQ1	Publication year Publication Channel: Journal, Conference, Book. Source name
MQ2	- HBE: Incorporating historical existing data in the evaluation. - Case study: An empirical assessment based on real-world datasets from hospitals/clinics.
MQ3	Medical tasks encompass various stages to ensure the well-being of patients: -Screening: Involves examining individuals who outwardly show no symptoms of an underlying illness, yet may harbor a concealed condition, presenting as overall good health. -Diagnosis: Focuses on the identification of a disease by analyzing its observable signs and symptoms. The choice of an appropriate treatment method is subsequently determined based on the outcomes of diagnostic procedures. -Prognosis: Entails forecasting the likelihood of recovery, guided by the nature of the disease and the exhibited symptoms. -Treatment: Aims to facilitate patient recovery and impede the progression of the disease through the application of suitable therapeutic interventions. -Monitoring: Encompasses the ongoing observation of both the disease's evolution and the patient's condition, providing valuable insights over time. -Management: Involves activities related to health promotion and the provision of necessary medical services to support overall well-being.
MQ4	The targets to predict with classification.
RQ1	Datasets used, modalities employed and the combinations that were exploited.
RQ2	The multimodal fusion approaches that were predominantly employed for BC classification.
RQ3	Aggregation and concatenation methods, with the ML and DL models used.
RQ4	The performance measures used (accuracy, sensitivity, specificity and others), and the evaluation methods employed.
RQ5	The interpretability technique employed to explain the predictions given by ML/DL models.
RQ6	The performance of multimodal fusion techniques and the performance on models relying on a single modality.

3.6 Threats to validity

To ensure the credibility of our study, it is essential to acknowledge the limitations inherent in the present review. In line with our study selection criteria, we specifically focused on fusion techniques applied for classification. However, we acknowledge the potential applicability of fusion in other compelling medical tasks, such as regression, segmentation and registration. Differences in the goals, input methods, and performance metrics were reported in the studies that were included, but some did not provide confidence limits. This makes it difficult to combine or statistically compare performance gains through meta-analysis. Furthermore, the validity of the reported metrics was not consistent, as certain studies lacked an independent test set, hindering the ability to provide an unbiased performance estimate. The scarcity of studies within each medical field, coupled with the intrinsic heterogeneity of each study, introduces further challenges in making qualitative comparisons. Additionally, a few studies have adopted unconventional approaches to fusion, potentially introducing subjectivity when classifying each study into categories such as decision-level, feature-level, or hybrid fusion. The primary threats to validity are outlined below:

- A comprehensive search string was formulated with appropriate keywords, and multiple iterations were conducted to cover a maximum of primary studies from the chosen digital libraries. Duplicate papers were removed.
- To prevent the exclusion of relevant papers, the authors rigorously applied selection criteria to titles, abstracts, and keywords. In cases of uncertainty, a thorough analysis of the full article was undertaken, with disagreements resolved through consensus meetings.
- To ensure high-quality paper selection, the authors established a minimum level of quality assessment.
- The reliability of the extracted data from the selected studies for addressing the research questions was a concern. Two researchers independently performed this task, minimizing the risk of incorrect data extraction. Disagreements were resolved through discussions.

4 Statistical trends of the map questions

The chosen studies and their findings concerning the MQs indicated in Table 2 are summarized in this section, along with a discussion of the findings. Figure 3 displays the flowchart of articles acquired during each phase of the selection procedure. The exploration of six electronic databases yielded a total of 838 potential papers from both searches. The inclusion and exclusion criteria were meticulously applied to identify pertinent studies, resulting in the selection of 53 articles. The selection process involved a thorough evaluation of titles, abstracts, and keywords, with a comprehensive review of full articles in cases of uncertainty. Subsequently, a rigorous quality assessment was employed, leading to the final inclusion of 47 qualified articles published from January 2018 to December 2023. The selected papers, along with some extracted results, are presented in Tables 8 and 9 for feature-level and decision-level fusion, respectively. For additional details and results related to the data

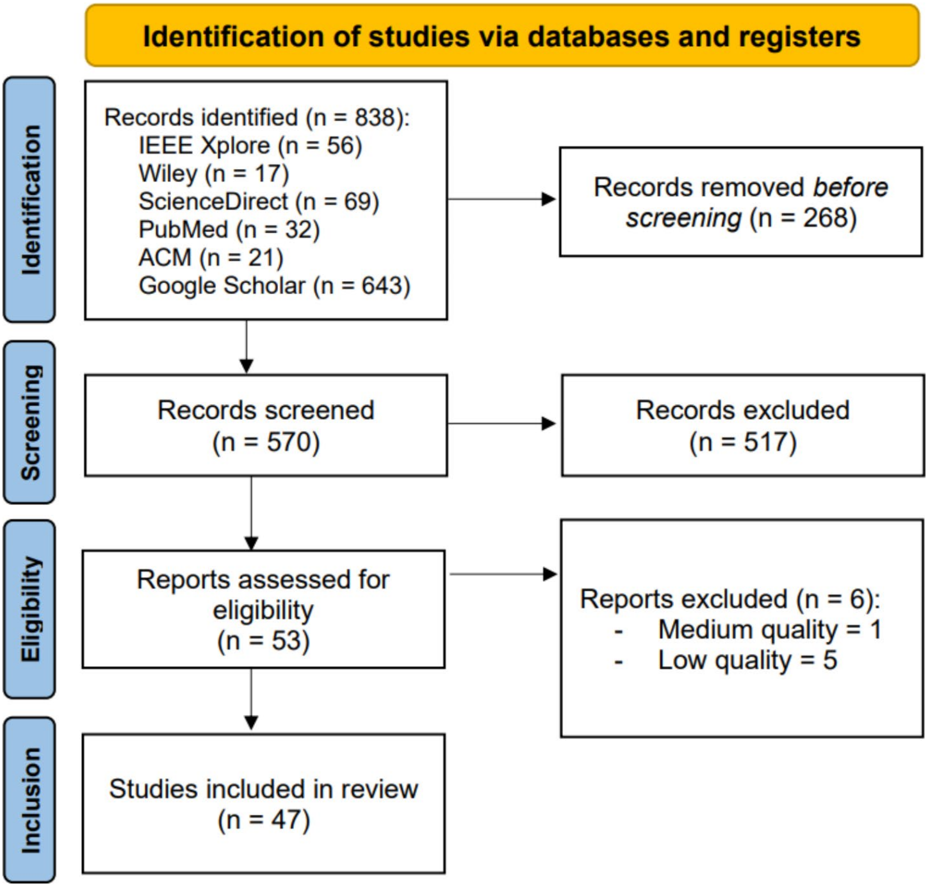
Identification of studies via databases and registers
Identification
Screening
Eligibility
Inclusion
Records identified (n = 838):
IEEE Xplore (n = 56)
Wiley (n = 17)
ScienceDirect (n = 69)
PubMed (n = 32)
ACM (n = 21)
Google Scholar (n = 643)
Records removed *before screening* (n = 268)
Records screened (n = 570)
Records excluded (n = 517)
Reports assessed for eligibility (n = 53)
Reports excluded (n = 6):
- Medium quality = 1
- Low quality = 5
Studies included in review (n = 47)

Fig. 3 PRISMA (preferred reporting items for systematic reviews and meta-analyses) flow diagram of the literature selection scheme

extraction, answers to MQs, and overall review, interested readers can request additional information from the corresponding author via email.

4.1 Publications' trends

By examining different sources and using mapping and review questions, it seems that journals have published more in-depth studies on multimodal data fusion for BC classification than did conferences. There was a total of 38 journal articles and 9 conference articles on this topic. Figure 4 illustrates the distribution of selected studies retrieved from six digital libraries. Google Scholar takes the lead, with the majority at 42.6%, while IEEE and ScienceDirect closely follow as the second-highest contributors, each comprising 20.3%. The next tier includes Springer and Scholar at 6.4%, and ultimately, Wiley represents a minimal share, with only one paper found.

Figure 5 displays the yearly distribution of articles spanning from 2018 to 2023, with no studies published prior to 2018. A substantial increase in publication rate was evident

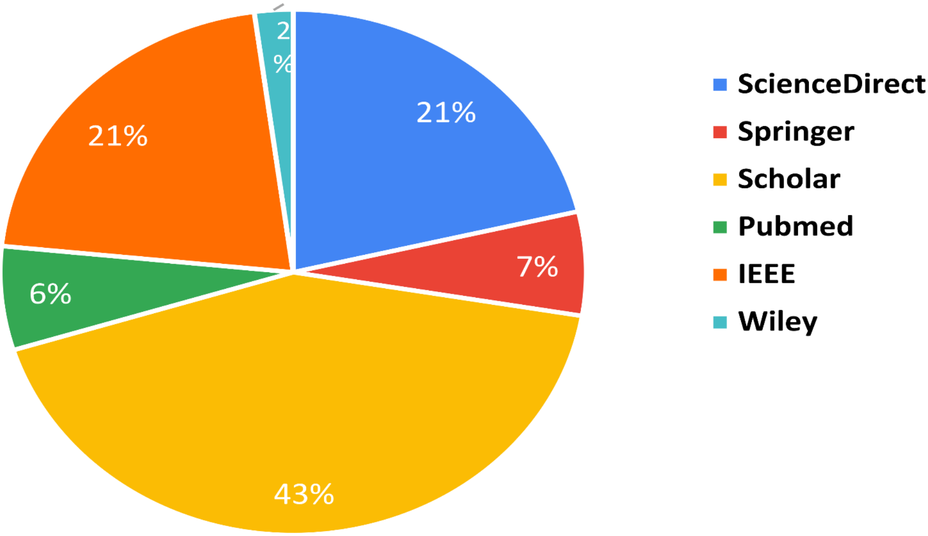

Fig. 4 Distribution of the digital libraries of the selected papers

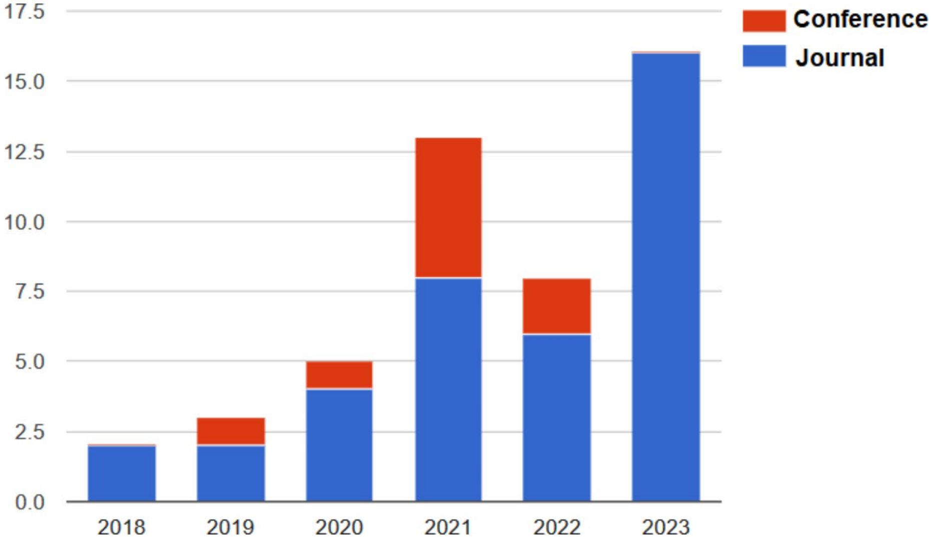

Fig. 5 Number of papers published per year and publication channel

from 2018 onward, with 2021 emerging as a standout year, encompassing 27.7% of the total selected studies. The large increase in publications in 2021 occurred because people suddenly realized how important it is to combine different types of data to improve DL models when making predictions, especially in the medical field. However, in 2022, there was a decline in the number of publications, only to witness a surprising resurgence in the subsequent year (2023). This resurgence may be attributed to the imperative need for developing new solutions and models, a process that typically requires time and refinement. The majority of the selected papers were indeed published in 2023, indicating that this acknowledgment was not an abrupt occurrence; rather, it was facilitated by the accessibility of ample computational power and extensive datasets. These enabling factors paved the way for the development of potent yet intricate models, such as DNNs.

As illustrated in Table 5, a total of 47 chosen studies were disseminated across various sources, predominantly journals and conferences. Journals accounted for 80.8% of these studies, while conferences constituted 19.2%. Table 5 exclusively highlights publication venues with a minimum of two primary studies. Among the journals, Biomedical Signal Processing and Control and Nature Scientific Reports emerged as the most frequently targeted publications. Regarding conferences, the International Workshop on Breast Imaging and the International Conference on Medical Image Computing and Computer-Assisted Intervention were the most common.

4.2 Contribution type

The examination identified two distinct types of empirical studies: human-based evaluation (HBE) and case studies. Figure 6 illustrates that 55.3% and 38.3% of the articles were categorized as HBE and case studies, respectively. Notably, three papers employed both HBE and case study methodologies, resulting in 26 papers exclusively utilizing HBE and 18 papers solely employing case study approaches. A noticeable trend is the widespread use of public datasets for evaluating solutions. This is probably because the results can be compared with those of other techniques tested on the same datasets, and public datasets are easy to find in the medical domain.

4.3 Medical tasks

In Fig. 7, the distribution of the 47 selected papers is presented based on medical tasks. Prognosis received the most research attention, accounting for 70.2% (33 papers), and diagnosis received 25.5% (12 papers). Screening and treatment were the least explored, each comprising 4.3% (2 papers), while none of the selected papers investigated the monitoring or management task.

Table 5 Publication channels

Publication source	#Of papers
Conference	
International Workshop on Breast Imaging (IWBI)	2
International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)	2
Journal	
Scientific reports	3
IEEE/ACM Transactions on Computational Biology and Bioinformatics	2
Biomedical Signal Processing and Control	3
Computers in Biology and Medicine	2
IEEE/ACM Transactions on Computational Biology and Bioinformatics	2

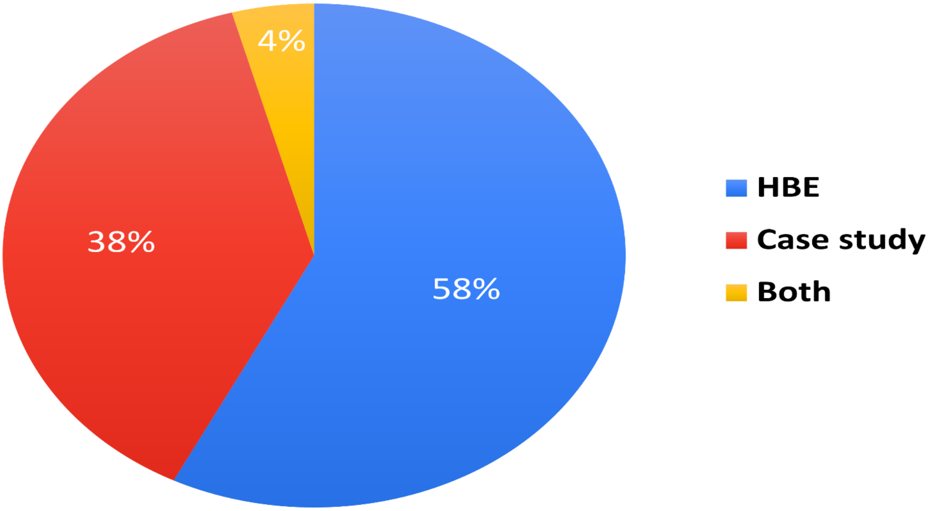

Fig. 6 Distribution of the empirical study types of the selected papers

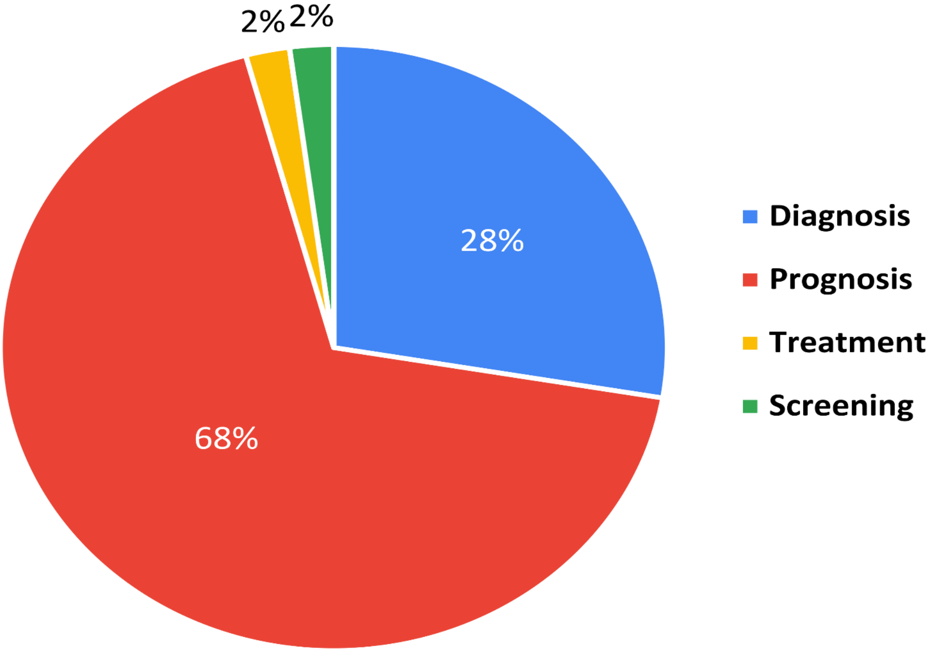

Fig. 7 Distribution of medical tasks

4.4 Classification targets

The primary objective of the MQ is to pinpoint the most extensively explored target within BC research. Within the realm of classification, various objectives were explored, as shown in Fig. 8, with the long-term and short-term survival of patients emerging as the predominant ones, accounting for 44.2% of the selected papers. Similarly, binary classification distinguishing between malignant and benign tumors was the second most common (21.2%), while multiclassification of BC molecular subtypes was the third most common. The recurrence of BC was another area of investigation, representing 7.7% of the selected papers. Additionally, 11.5% of the selected papers delved into predicting diverse targets, such as normal or abnormal breast conditions, pathological complete response to neoadjuvant chemotherapy, and lymph node metastasis. Notably, four papers successfully predicted more than one target simultaneously.

5 Datasets and modalities

In our SLR, the primary focus of the first RQ was to identify diverse datasets and modalities across the selected studies. For the datasets, a predominant trend emerged, with the majority being private (20 papers); only four public datasets were identified, namely, the METABRIC, TCGA-BRCA, the PathoEMR and BCNB cohorts, as detailed in Table 6. Seven studies evaluated their models with two datasets (TCGA-BRCA and METABRIC).

Distinguishing between three distinct types of modalities is a key aspect of this SLR, as it focuses on their application in BC research. The first modality, clinical data, encompasses various forms of tabular data derived from electronic health records, including patient medical history, diagnostic information, and treatment plans. The second modality, genomic data, involves the analysis of genetic information, including gene expression, copy number

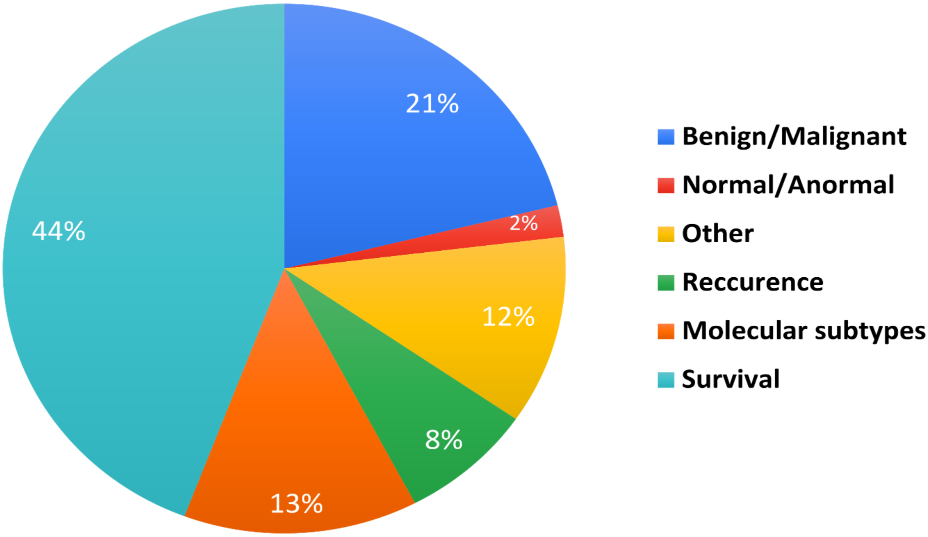

Fig. 8 Distribution of the targets to predict

Table 6 Public datasets used across the selected studies

Dataset	*N*# of Papers	Overview	Modalities
TCGA-BRCA (The 2023)	19	Subset of The Cancer Genome Atlas (TCGA) project focusing on BC. TCGA was developed between the NCI and National Human Genome project starting in 2006. The data for BC is standardized to 1031 patients.	Clinical Data, Gene Expression, CNV, Histopathological Imaging
METABRIC (cBioPortal 2023)	13	The METABRIC (Molecular Taxonomy of BC International Consortium) dataset was created through the collaborative efforts of researchers involved in the METABRIC consortium. It includes data from over 2,000 BC cases.	Clinical Data, Gene Expression, CNA, Histological Imaging
PathoEMR (Yan et al. 2021)	1	Peking University International Hospital released a dataset containing structured data from 185 patients for benign and malignant BC classification.	Pathological images with their Clinical EMR attributes
BCNB (Xu et al. 2021)	1	the Institutional Ethical Committees of Beijing Chaoyang Hospital affiliated to Capital Medical University released the Early BC Core-Needle Biopsy WSI dataset for micro-metastatic ALN preoperatively of 1058 patients.	Clinical Data, H&E Stained

variation (CNV), copy number alteration (CNA), and alternative genetic alterations. These molecular-level insights provide a deeper understanding of the genetic factors influencing BC. The third modality, medical images, encompasses visual representations obtained through techniques such as mammography, magnetic resonance imaging, and computed tomography. These images play a crucial role in diagnosing and monitoring BC, providing detailed insights into anatomical structures and aiding in treatment decision-making. Figure 9 illustrates the distribution of studies for each modality combination. Notably, combinations involving clinical and genomic data or clinical and imaging data were the most prevalent, each comprising 14 papers. Modalities with an imaging component were present in 12 papers, while the combination of medical images and genomics modalities occurred in 5 studies. Intriguingly, only four papers incorporated all three types of modalities simultaneously.

Upon thorough analysis of the selected studies, six modalities were identified in addition to clinical data for the following genomics types: gene expression, CNV/CNA, and four medical imaging modalities: histopathology, mammography, magnetic resonance imaging (MRI) and ultrasound (US). Figure 10 depicts the distribution of various modalities within the chosen studies. Clinical data emerged as the most predominant, constituting 28.6%. Similarly, the gene expression was 20.5%. histopathology and CNV/CNA shared the third position at 14.3%. US followed up at 8.9%, followed by mammography at 7.1%, and finally MRI at 6.3%.

In Table 7, a comprehensive breakdown of the combinations explored by the studies is presented, highlighting the intended classification targets. Notably, 20 multimodal combinations were documented: clinical data and gene expression, clinical data and gene expression combined with CNV/CNA or histopathology or both, as well as the amalgamation of clinical data, CNV/CNA, and histopathology with MRI. Additionally, the combination of histopathology and gene expression with or without CNV/CNA was explored. In terms of imaging combinations, variations such as histopathology with US, the combination of MRI with mammography or US, or the combination of mammography with US were observed.

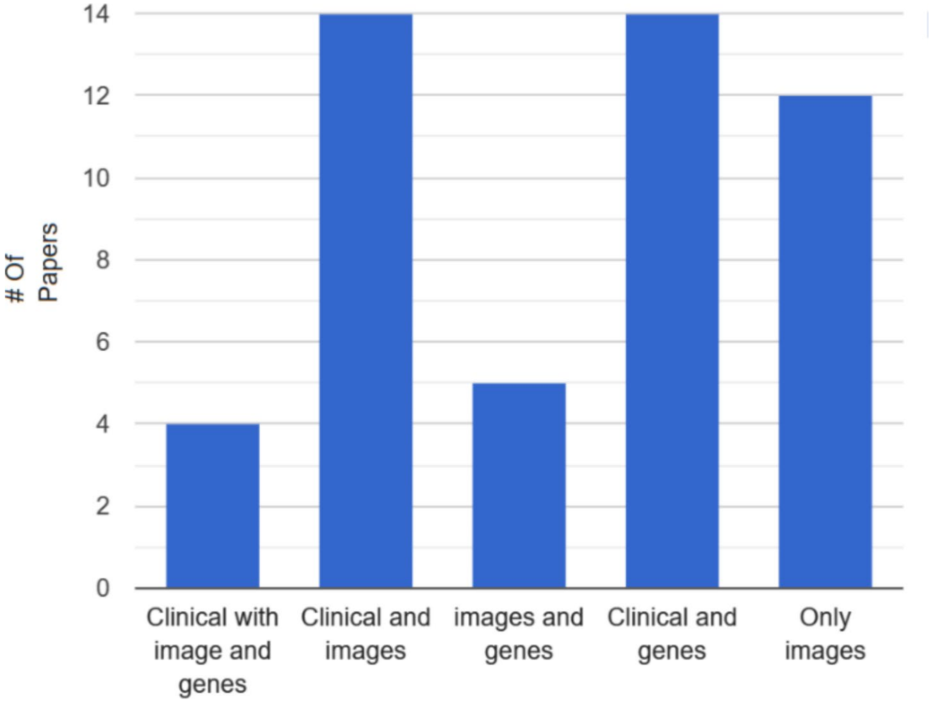

Fig. 9 Distribution of the combinations of the different modalities

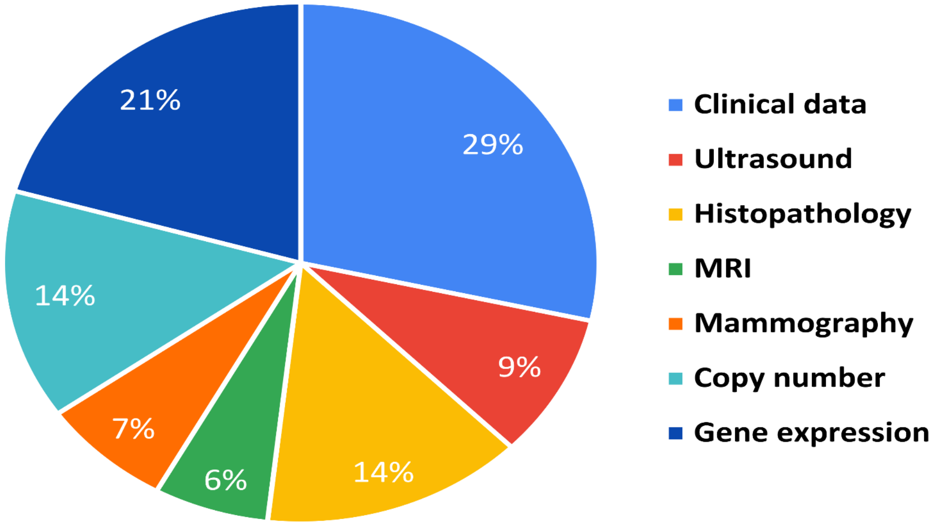

Fig. 10 Distribution of the different modalities across the selected studies

Table 7 Combinations of modalities

	Modalities combined								Targets to predict						
Combs	CD	GE	CN	HI	MG	MRI	US	DBT	IMC	DOT	B/M	MS	SURV	REC	Total
C1	X	X											O		2
C2	X			X							O		O	O	6
C3	X				X						O		O		2
C4	X					X					O			O	3
C5	X						X				O				2
C6	X	X	X												12
C7	X	X		X										O	1
C8	X	X	X	X									O		2
C9	X	X		X		X						O			1
C10		X	X	X								O			2
C11		X		X									O		3
C12				X			X				O				1
C13					X	X									1
C14						X	X				O	O			2
C15					X		X					O			4
C16							X	X			O				1
C17						X		X			O				1
C18	X								X				O		1
C19					X			X			O				1
C20							X			X	O				1

Clinical Data (CD), Gene expression (GE), Copy Number Variation (CN), histopathology (HI), Mammography (MG), magnetic resonance imaging (MRI), Ultrasound (US), digital breast tomosynthesis (DBT), Imaging Mass Cytometry (IMC), and diffuse Optical Tomography (DOT) and classification targets: benign or malignant tumor (B/M), multiple BC subtypes (MS), long-term or short-term survival (SUR), and recurrence (REC)

When examining the frequency of combinations utilized across studies, the most prevalent combination was clinical data with gene expression or CNV/CNA, appearing in 12 studies. The following closely followed: histopathology and clinical data (6 studies); MG and US (4 studies); and combinations of histopathology and gene expression, as well as MRI and clinical data, each featured in 3 studies. All the other combinations were represented in one or two studies. Specifically, in one study, the clinical data was presented in the form of BIRADS scoring data. Furthermore, an innovative approach in two separate studies included the incorporation of digital breast tomosynthesis (DBT) with US and MRI. Another study explored a comprehensive combination of clinical data, CNV/CNA, gene expression, and histopathology, as well as DNA and RNA. Importantly, two studies utilized modalities that were not employed elsewhere. Notably, US was combined with diffuse optical tomography (DOT), and clinical data was combined with imaging mass cytometry (IMC). Notably, for the predicted targets, the binary classification of benign and malignant tumors (sometimes with normal types) exhibited the highest diversity, followed by the multiclassification of molecular subtypes with four combinations, the long-term and short-term survival of patients with five combinations, and the recurrence of BC with only three combinations. The references of the papers that used combinations of modalities can be found in Tables 8 and 9.

6 Multimodal fusion

In the context of multimodal learning for BC classification, a comprehensive analysis of the selected papers reveals diverse strategies employed for fusion. Figure 11 illustrates the distribution of fusion types across the selected papers. It is evident that feature-level fusion is the predominant choice, with the majority of papers adopting this approach. In contrast, only a limited number of papers, specifically six, utilized decision fusion. This observation could account for the relatively low prevalence of decision-fusion implementations. However, it is crucial to acknowledge that only two papers directly compared these approaches (decision-fusion and feature-fusion), limiting the robustness of this comparison. Interestingly, as of 2023, no studies have reported using the decision-fusion approach. This absence in recent literature may suggest a growing recognition among researchers regarding the efficacy of feature fusion over decision fusion. The indication here is that the field is evolving, and decision fusion may be diminishing in popularity as researchers increasingly appreciate the advantages offered by feature-level fusion. Notably, two papers innovatively fused features at the semantic level: Arya et al. (2023) fused clinical data with gene expression and CNV, and Muramatsu et al. (2022) fused mammography data with US images. Additionally, four papers adopted a unique approach by combining the semantic features of one modality with the extracted features of another modality, such as histopathological images with clinical data in Yang et al. (2022), Yan et al. (2019), Wang et al. (2023), and histopathological images with US images (Qu et al. 2023). The remaining papers focused on fusing extracted features from both modalities. Moreover, one paper used hybrid fusion and combined both decision fusion and feature fusion approaches.

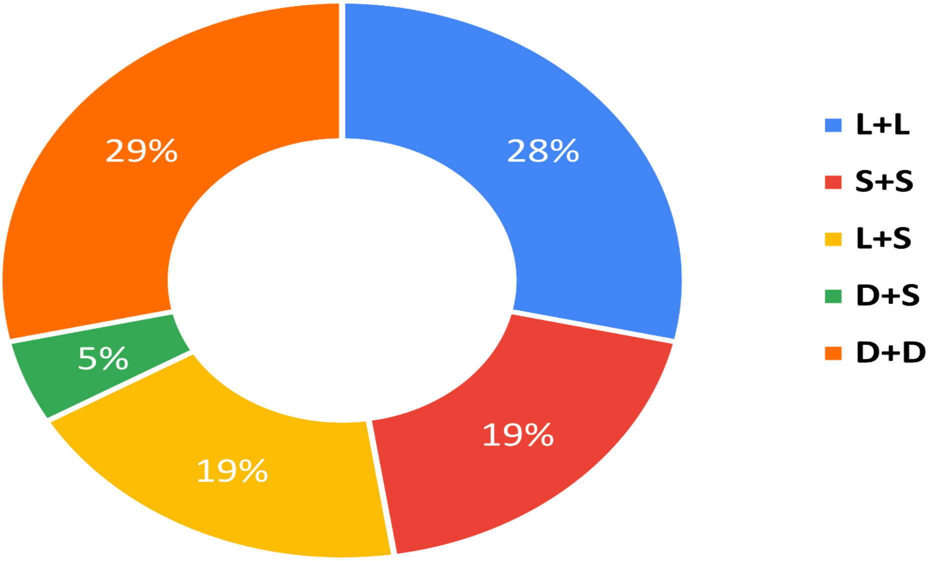

Fig. 11 Distribution of the fusion approaches used across the selected studies

6.1 Decision fusion

Decision fusion involves employing predictions from multiple models to make a final prediction. Decision fusion was only employed in six of the forty-seven studies included in the analysis (Table 9). Given the limited number of papers focusing on decision fusion, the combinations of modalities explored are notably diverse, with five distinct combinations indicating innovative applications in each study. The studies implemented distinct aggregation strategies such as averaging, weighted voting, and employing a meta-classifier. Liu (2021) integrates predictions from a DNN model focused on gene expression, along with CNV, and a CNN based on the VGG16 architecture for histopathological images. The fusion process involves meticulous adjustment of the fusion parameters and employs a weighted linear aggregation method to ascertain the weights assigned to the probability of each modality (Sun et al. 2019). integrates the predictions from three separate DNNs, each trained on a distinct modality—clinical data, gene expression, and CNV and adopts a similar aggregation method as Liu (2021). On the other hand Arya and Saha (2020), replaces DNNs with CNNs and employs a different approach by training a metaclassifier to establish the connection between the input probabilities and the ultimate output. Various classifiers were tested: random forest, naive Bayes, logistic regression, and support vector machine (SVM), with random forest emerging as the most effective in this context (Mullen et al. 2023). extracted image texture features from both MR and DBT images and inputted them into an SVM, generating likelihoods or cancer-likeness scores. These scores from each modality were then fed into a joint Bayesian classifier for data integration. Subsequently, the joint Bayesian classifier was employed to calculate the probability of malignancy. The authors of Holste et al. (2021) integrate the output probabilities from two models. The initial model is a DNN that takes clinical data as input, while the second model is a pretrained ResNet50 designed for MRI images. The probabilities generated by these models serve as inputs to a fully connected layer, ultimately producing the final prediction. Instead of relying on a single classifier for each modality to obtain the final class, the authors of Rabinovici-Cohen et al. (2022) relied on the ensemble learning method to generate the final decision by calculating the mean value of the six classifiers. They implemented three classifiers for each modality: three CNNs with different hyperparameters for MRI and logistic regression, random forest, and XGBoost for clinical data.

6.2 Feature fusion

Table 8 provides a detailed overview of feature-level multimodal fusion research for BC detection, encompassing modality combinations, datasets used, feature extraction methods (images and non-images), and specifics regarding fusion and classification models. In the realm of feature extraction, CNNs, recognized for their capacity to automatically learn hierarchical representations, have emerged as the predominant method, particularly for imaging-type modalities, with 23 papers employing CNNs out of the 34 papers related to such modalities. For non-imaging modalities, CNNs remained prevalent (9 papers), followed by DNNs (7 papers) and autoencoders (5 papers). Various other methods, including decision trees, logistic regression, random forest, and XGBoost, were employed sporadically. Notably, 12 papers integrated an attention mechanism into the feature extraction process, selecting the best features for each modality, and multi-instance learning (Li 2021) was explored in 3 papers (Xu et al. 2021; Wang et al. 2023; Li et al. 2021) mainly to address patches of WSIs of BC.

Following the feature fusion process, the majority of papers (29 out of 47) applied DNNs, primarily leveraging FCLs (21 papers), regular DNNs, MLPs, and ANNs for classification. Figure 12 illustrates the application of these models. Furthermore, random forest was the most commonly used ML classifier (10 papers), followed closely by SVM (9 papers). Other classifiers (naive Bayes, k-NN, GNNs, fuzzy classifier, XGBoost, decision tree, CNNs, extreme learning machine, extra random trees, and LSTM) were employed sporadically, with some papers comparing multiple classifiers to assess performance. The

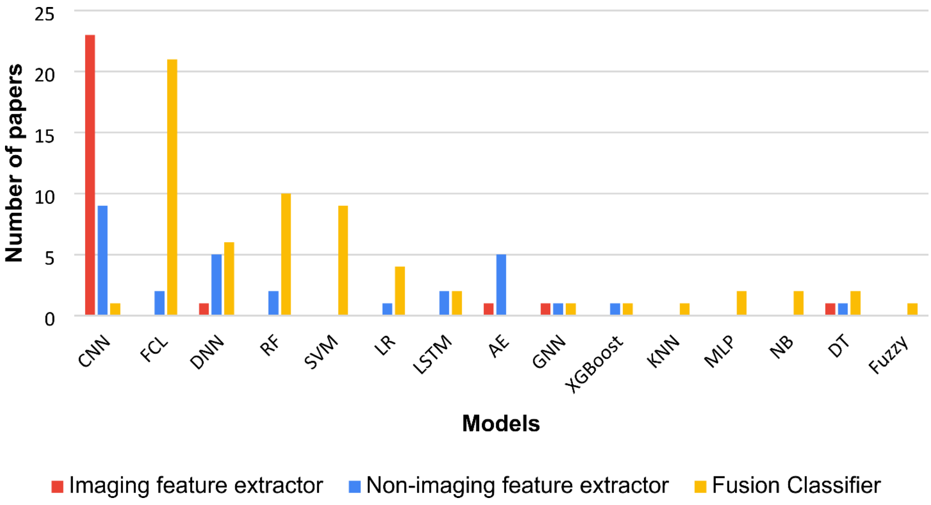

Fig. 12 Distribution of the different ML and DL models across the selected studies

distribution of models across selected papers reveals prevalent usage, with certain models emerging as dominant choices. This underscores the diverse landscape and popularity of specific approaches within the studied literature. In the selected studies, ML models were mainly used as meta-classifiers for feature-level and decision-level fusion, whereas DL models were primarily employed as feature extractors. This complex mix of ML and DL techniques, each used for a different reason, provides a sophisticated way of dealing with problems in multimodal feature classification in the context that was studied.

In the realm of multimodal fusion, the selected studies exhibit a diverse array of techniques for feature fusion. The predominant method across studies involves concatenation, a widely adopted strategy. However, a subset of papers, including Holste et al. (2021), Wang et al. (2023), and Mokni et al. (2021), explore alternative operations to integrate information from multiple modalities. Notably, variants of the learned feature fusion model implement elementwise addition and multiplication, demonstrating comparable performance to that of the concatenation-based version. Moreover, distinct feature fusion approaches are evident in the selected studies. For instance Yang et al. (2022), employs multimodal compact bilinear logic, emphasizing a compact representation strategy (Li 2020). opts for a conditional autoencoder, leveraging its capacity for correlated feature extraction. In contrast Yuan and Xu (2023), incorporates the Gated Multimodal Unit, showcasing a specialized unit for effective fusion. Additionally Wang et al. (2020), adopts a similarity network fusion algorithm, further diversifying the feature fusion methodologies explored across the reviewed studies.

6.3 Hybrid fusion

In the domain of multimodal fusion, an impactful hybrid strategy is evident, as illustrated by the research detailed in Wu et al. (2023), where the authors introduce the innovative multimodal DL radiomic nomogram (DLRN) crafted for predicting metastatic BC. The DLRN seamlessly integrates dual-modality US images, radiomic features, and clinical information, revealing the fusion of decision-making processes and feature extraction. Specifically, the DL score and radiomic score are derived from the deep CNN and the radiomic model, respectively. Importantly, clinical information, confined to age and tumor size, was harmoniously integrated with the DL score and radiomics score to formulate the multimodal DLRN. Subsequent evaluation underscored the superior performance of the DLRN model in comparison to three unimodal models constructed solely from clinical features, the DL score, and the radiomics score. This pioneering hybrid fusion approach represents a promising avenue for advancing the ability of multimodal models to predict the prognosis of BC patients.

6.4 DL models

In examining recent trends in multimodal fusion for BC classification, our analysis of selected studies revealed a significant shift toward the utilization of DL models, with 70% of the studies relying on DL techniques, as shown in Fig. 13. This dominance underscores the increasing preference for DL in the medical domain. Notably, DL models are strategically employed by researchers facing the intricate challenges of handling multimodal data in BC classification (Akkus et al. 2023). This trend underscores the prevalent preference for DL techniques, which are renowned for their success in automatically learning intricate rep-

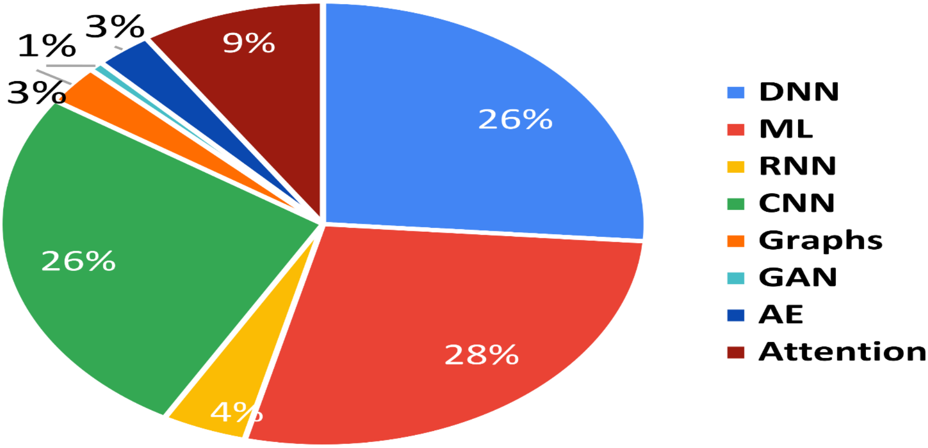

Fig. 13 Distribution of DL models for BC classification across the selected studies

resentations from large and complex datasets across various domains (Akkus et al. 2023). It is crucial to recognize that the choice between DL and ML models hinges on factors such as the nature of the specific problem, dataset characteristics, and interpretability importance (Hakkoum et al. 2022). While traditional ML models remain relevant in scenarios with smaller datasets or where interpretability is paramount, the dominance of DL models persists in tasks demanding complex data analysis.

The findings from our pie chart not only highlight the prevailing influence of DL in current research but also pave the way for a nuanced understanding of the specific DL models contributing to advancements in BC classification:

- FCL and MLP emerged as the most commonly used models for multimodal feature classification across the selected studies. This indicates the strategic utilization of DNN classifiers for feature fusion when combining information from various modalities, and the prevalence of FCL in multimodal feature classification underscores its effectiveness in capturing intricate relationships within diverse datasets.
- Notably, 70% of the papers opt for pretrained CNN models, such as ResNet50 (Szegedy 2017), trained on ImageNet. The straightforward fine-tuning procedure through transfer learning, which facilitates the ease of adaptation to medical imaging, favors this option (Hemant Kumar et al. 2021). While pretraining provides an initial convergence direction by leveraging knowledge from a broader dataset, it is essential to consider potential conflicts with the goal of representing diverse data aspects in information fusion methodologies (Stahlschmidt et al. 2022). Several papers (Guo et al. 2021; Atrey et al. 2023; Kayikci et al. 2023), and Yuan and Xu (2023) recognizing this trade-off choose to construct the CNN architecture from scratch, emphasizing a tailored approach.
- RNNs were employed in 4 distinct studies. In Othman et al. (2023), LSM and GRU were employed to classify stacked features, while Li (2020)utilized an LSTM network to derive a fixed-length image feature from TR and ROI features. Additionally, in Atrey et al. (2023), the LSTM, which is equipped with the capability to handle variable-length inputs and capture long-term dependencies through the incorporation of a forget gate, proved well suited for the classification task. Consequently, LSTM was implemented

for classification using features extracted by the CNN. Moreover Mustafa et al. (2023), adopted an LSTM module for feature extraction of gene expression data. Furthermore Yuan and Xu (2023), explored the effectiveness of a Bi-LSTM model, revealing its superior performance compared to that of 1D-CNN in processing CNA and gene expression data at the feature extraction layer.

- One notable application of autoencoders is seen in Yan et al. (2021), where denoising autoencoders (DAEs) are utilized to increase the dimensionality of the clinical data. While basic autoencoders are commonly used for dimensionality reduction, DAE, by training on noisy versions of raw data, enhances robust feature representation and prevents overfitting. Another instance is observed in Li (2020), which employs a conditional autoencoder for pathological and genome feature fusion. In this context, a conditional autoencoder was established to extract correlated features from tumor regions (TRs), regions of interest (ROIs), and gene expression data. Furthermore, the authors of Arya et al. (2023)explores variational autoencoders (VAEs), designing them with log-cosh loss for all modalities. VAEs introduce explicit regularization during training, ensuring the regularity of the latent space, and are shown to extract features of very low dimensions. Finally, the authors of Furtney et al. (2023) leverages an autoencoder to represent genomic variant results or microarray expression features in a condensed latent space, with the low-dimensional vectors extracted from the latent space providing inputs into the GNN.
- The paper Du et al. (2023) is the sole one among the selected studies that employed a GAN to create a multimodal adversarial representation framework for predicting BC prognosis. The feature projector, a key component for representation learning, generates an invariant representation across different modalities within a common subspace. Its primary objective is to perplex a modality classifier acting as an adversary, which attempts to differentiate items based on their modalities and guides the learning of the feature projector. The inclusion of the modality classifier in adversarial training enhances the alignment of representation distributions across modalities, fostering effective modality invariance. The optimized representation subspace, achieved through the convergence of this process, ensures cross-modal associations. Additionally, a classifier learns to categorize encoded representations into correct labels, preserving the underlying cross-modal semantic structure for predictive tasks. Furthermore, individual decoders for each modality prevent the loss of unimodal information.
- Notable applications of GNNs include Furtney et al. (2023), which employs a Relational GNN to predict molecular subtypes in BC patient data. Additionally Wang et al. (2020), integrates GNNs for cancer survival prediction by combining multiple genomic and clinical data. Furthermore Fu et al. (2023), introduces a deep multimodal graph-based network (DMGN) with a GAT to consider relationships between spatial phenotype information and clinical variables in medical imaging. The GAT approach optimizes information prioritization by emphasizing the contributions of neighboring nodes. These examples exemplify the effectiveness and versatility of GNNs, offering solutions across a spectrum of domains, from cancer prognosis to multimodal graph-based reasoning in medical data.

7 Multimodal fusion architectures

After scrutinizing the 47 selected papers, several distinct patterns in learned feature fusion emerged, deviating from the traditional feature fusion approach outlined in Sect. 2. Four noteworthy patterns are identified:

– Feature selection based on the attention mechanism.

  The attention module, a pivotal component in neural network architectures, dynamically focuses on relevant features of input data by assigning varying weights to different segments of the input sequence (Jetley 2018). Traditional applications in NLP, such as machine translation and speech recognition, have demonstrated the adaptability and effectiveness of attention mechanisms (Wan et al. 2019). However, challenges arise when dealing with multimodal data, as illustrated by concerns about the traditional attention mechanism's ignorance of data heterogeneity (Stahlschmidt et al. 2022). To address this, a recent study applied shallow attention nets to each feature, effectively extracting key information from multimodal data while accounting for the distinction and uniformity of heterogeneous data. Additionally, various papers have leveraged attention mechanisms in innovative ways. For example Guo et al. (2021), employs a shallow attention net for each feature, accommodating the need for multimodal data by effectively extracting key information. In Arya et al. (2021), CNNs incorporated sigmoid-gated attention, enhancing feature maps and discarding irrelevant features for improved classifications. The authors of Chen et al. (2019) proposed a DNN model (AMND) based on the attention mechanism, which demonstrated enhanced connections between patient clinical and gene expression data for BC prognosis. In Yao et al. (2022), an attention module was used to score each area of H&E-stained images, combining image features with clinical and gene expression data to predict the risk of recurrence and metastasis. Finally, a spatial attention convolution network automatically was developed in Luo et al. (2023) to focus on key regions of images, while a fully connected neural network performed feature correlation coding and extraction, combining the scoring data with image features in high-dimensional space. These examples underscore the versatility and widespread applicability of attention mechanisms across various domains and tasks.

– Multimodal Feature Selection

  In the realm of feature selection, a distinctive strategy emerges in which attention mechanisms come into play after the concatenation of features. This approach contrasts with the conventional method of employing attention mechanisms independently for each modality, giving rise to the notions of intra-selection and inter-selection. This enhancement is logical, given that selecting features from different modalities representing the same patient and disease introduces redundancy, which can be addressed only through feature selection on fused multimodal features. Papers such as Luo et al. (2023) exemplify this approach by addressing key challenges in classification tasks involving US images and BIRADS features. This research addresses the challenge of redundant information that arises when directly concatenating features extracted from US images and human knowledge descriptions. To address this issue, a channel attention mechanism-based aggregation method is introduced. This approach involves correlation analysis and feature selection between the features derived from images and BIRADS scoring data, automatically assigning weights to different channels for effective feature fusion. Similarly, a novel 'fuzzy' approach for BC prognosis was introduced in Chharia et al. (2021), framing the task as an incremental learning problem. The study employs a fuzzy classifier on a stacked multimodal feature vector for each patient, allowing the model to continually update its learned feature space on a nonstationary multimodal data stream. This unique approach demonstrates the model's ability to learn complex relationships between different multimodal attributes. The authors of Wang et al. (2021) proposed a genomic and pathological deep bilinear network (GPDBN) to establish a unified framework for BC prognosis prediction by integrating both genomic data and pathological images. Notably, the GPDBN introduces an inter-modality bilinear feature encoding module to model complex relations across different modalities. It also captures intra-modality relations through two intramodality bilinear feature encoding modules. The combination of inter- and intramodality bilinear features using a multilayer DNN further enhances the overall prognostic prediction performance. These examples illustrate the significance of post-concatenation attention mechanisms in optimizing the fusion of features from different modalities, leading to improved predictive capabilities in diverse applications. Additionally, two key attention mechanisms, namely, intramodality attention and inter-modality attention, were suggested by Zhang et al. (2023) for enhancing the overall feature representation. Intramodality attention focuses on refining features within the same modality. Concurrently, inter-modality attention extends its influence beyond a single modality. This attention module is designed to interactively fuse information across diverse modalities, fostering synergy among features extracted from mammography and US. By leveraging attention mechanisms, both intra- and inter-synergistic, the MDL-IIA model adeptly refines and fuses features, ensuring a more comprehensive and discriminative representation that contributes to the accurate prediction of BC molecular subtypes. This dual attention strategy allows the model to capture nuanced patterns within each modality and simultaneously exploit synergies across different modalities, resulting in a robust and effective multimodal feature representation.

– Multimodal Feature Extraction

In certain cases, the extraction of features from a single modality is influenced by interactions with another modality, such as through loss (before actual fusion occurs). First, to address the assumption of equal contributions from all modalities, the authors of Arya et al. (2021) designed a cross-modality attention-based architecture called the SiGaAtCNN Bi-Attention for three modalities. The model employs a bimodal-based attention mechanism, with cross-modal attention performed pairwise, prioritizing distinct features based on their significance in prediction tasks. In Li (2020), an autoencoder-based model is proposed for the correlation of image features with gene expression. To enhance the correlation, a gene encoder for dimension reduction and a conditional autoencoder for correlated feature extraction are introduced. Furthermore, the authors of Chen et al. (2019) introduce the attention-based multimodal neural network (AMND) for BC prognosis. This DNN fuses gene expression and clinical data using an attention mechanism. The gene expression profile was decomposed using five algorithms, and the attention mechanism calculates the weight of each representation based on clinical data. The weighted summation of these representations is concatenated with clinical data to create the final feature representation, which is subsequently input into the DNN for the classification task. This approach, by considering the intricate relationships between modalities, allows for a more holistic understanding of the data, leading to improved performance and richer feature representation.

– Multimodal Feature Representation

Another approach in feature fusion involves not just feeding fused features directly to a classifier for the final decision. Instead, various sublevels of fusion are employed, distributing features to different methods (e.g., DNNs and graph-based methods) to obtain diverse representations. These distinct representations, encompassing the same modalities, are then fused before being presented to the classifier. This approach includes what can be termed intermediate fusion, where multiple fusion steps are performed before multimodal representations are ultimately fused. Two studies have employed this methodology. In the study by Qiao et al. (2022), the authors present a discriminative feature learning method utilizing a min–max feature loss training strategy for BC classification. This method ensures the separation of modality-agnostic and modality-specific features, focusing on tumor regions and context sensitivity, respectively. Additionally, they introduced a feature fusion module that incorporates adversarial learning and nearest neighbor selection to enhance inter-modality affinity, ultimately contributing to improved predictive performance for lymph node metastasis, histological grade, and Ki-67 expression. Similarly, within Guo et al. (2021), a model for multimodal data fusion, named the multimodal affinity fusion network (MAFN), was introduced for predicting BC survival to integrate gene expression, CNA, and clinical data. By utilizing a stack-based shallow self-attention network, the model amplifies survival-related features within lesion regions. The affinity fusion module enhances structured information mapping between patients and multimodal data, leading to a robust fusion feature representation. The resulting fusion feature embedding, combined with a specific feature embedding from a triple modal network, is employed for accurate classification of long-term or short-term survival. This method not only considers multimodal data but also integrates multiple feature extraction methods, enhancing the prognostic performance of BC. Collectively, these studies highlight the significance of fusing different feature representations from various modalities to achieve improved predictive capabilities. However, the challenge of obtaining better feature representations and considering the intricate relationships among multimodal data is still being overcome, urging further exploration in this area.

8 Strengths and weaknesses of fusion approaches

In this section, we aim to provide a nuanced understanding of multimodal fusion for BC by delving into its various approaches, each characterized by distinct strengths and weaknesses.

8.1 Decision fusion

Decision fusion has been recognized for its various advantages: it minimizes the risk of suboptimal predictions that might result from a model trained on a single modality, leveraging the fact that errors from different modalities are typically uncorrelated (Baltrušaitis et al. 2019). This approach enhances algorithm performance, providing the flexibility to make predictions even when a specific modality is unavailable and facilitating training in the absence of parallel data. However, decision fusion may neglect low-level interactions between modalities (Huang et al. 2020). According to the selected studies, decision fusion is preferred when signals from different modalities do not complement each other, meaning that input modalities independently inform the final prediction without inherent interdependency. Decision fusion mitigates the "curse of dimensionality" by using specialized models for each modality, limiting the input feature vector size for each model (Aremu et al. 2020). In scenarios with missing or incomplete data, decision fusion retains the ability to make predictions. It employs separate models for each modality, allowing for aggregation functions even when predictions from a modality are missing. Decision fusion is advantageous when dealing with modalities having different numbers of features, as it considers each modality separately and can be tuned to mitigate differences in feature numbers. Two studies explored multiple fusion approaches and conducted comparisons between them (Arya and Saha 2020) and (Holste et al. 2021). In line with expectations, the decision fusion approach was found to achieve less than optimal performance in comparison with other fusion approaches, as reported in another review (Huang et al. 2020).

8.2 Feature fusion

In various applications, feature fusion is the initial choice for multimodal learning since it can learn shared representations, facilitating the model's ability to understand correlations across modalities and ultimately leading to improved overall performance (Singh et al. 2022). It is a straightforward approach that does not require training multiple models when semantic features are used; in that case, it was found that the combined modalities have the same dimension (different types of medical images) or can be represented using the same format (as for clinical data with genetic data). However, challenges arise when input modalities have different dimensions, especially when combining clinical data (1D) with imaging data (2D or 3D); in that case, studies tend to fuse learned features extracted using DL models. CNNs can be employed for feature extraction and are commonly chosen for integrating information, particularly at the feature level (Huang et al. 2020). While this preference is attributed to the quality of their features, these networks encounter a significant decline in performance when confronted with shifts in dataset features, especially with smaller datasets. The authors of Muramatsu et al. (2022) fused two imaging modalities using semantic features, and the learned features revealed that the fusion of learned features outperformed the fusion of raw images since the extracted features represent high-level abstractions learned by the DL models, highlighting discriminative patterns that are crucial for accurate diagnosis or analysis. On the other hand, when raw data from different modalities are fused directly, the model may struggle to discern the distinct features and relevant information within each modality, leading to potential information loss or interference between modalities, which can explain why the majority of papers using feature fusion rely on learned features of different modalities and not semantic features.

8.3 Hybrid fusion

The decision to employ a hybrid fusion approach, as demonstrated in Wu et al. (2023), stems from thoughtful consideration of the characteristics and dimensions of the involved modalities—dual-modality US images and radiomic features. In many cases, these imaging modalities generate high-dimensional data, and directly combining all the features could lead to challenges related to the curse of dimensionality. The curse of dimensionality refers to the increased complexity and computational demands as the number of features or dimensions grows (Aliper et al. 2016). By extracting DL scores and radiomic scores separately from the two imaging modalities, the study aimed to capture the essential information and patterns present in each modality independently. The subsequent integration of these scores with clinical features is a strategic decision to strike a balance and prevent potential issues associated with dimensionality. If all the raw features from the imaging modalities were directly combined, the resulting high-dimensional representation might overshadow or dominate the clinical features, leading to suboptimal model performance.

The introduction of scores allows for a condensed yet informative representation of the complex information contained in the imaging modalities. The fusion of these scores with clinical features provides a comprehensive prediction model that leverages the strengths of each modality while mitigating challenges related to dimensionality. This strategic fusion approach enables the model to effectively utilize the unique contributions of each modality without being overwhelmed by their respective dimensions, thereby enhancing the overall predictive capabilities for BC.

9 Model evaluation

The validation methods employed to assess the effectiveness of the ML techniques in the selected papers were categorized into three main approaches: K-fold cross-validation, data splitting (testing and training sets), and the holdout technique. Among the 47 papers, the following distributions were utilized for these methods: 49% (24) for K-fold cross-validation, 49% (23) for Data split, and 2% (1) for the Holdout technique. Additionally, one paper utilized both K-fold cross-validation and the holdout technique simultaneously. In the case of K-fold cross-validation, various K-folds were implemented, including K=2, 3, 4, 5, and 10. The most prevalent choices were 5-fold (8 papers) and 10-fold (13 papers), while other values were each represented in just one paper. For the data split method, multiple splits were employed, such as 90/10, 80/20, 70/30, or 60/40. The majority of papers opted for 80/20 (13 papers) and 70/30 (5 papers), while other split ratios were each represented in 2 or 1 paper, respectively. Regarding the evaluation of ML/DL models, the performance metrics are classified into two categories: single scalar and graphical. Single-scalar criteria, such as accuracy, recall (sensitivity), and specificity, are commonly used in contrast to measures such as the Cohen kappa, Jaccard index, G-measure, t value and p value. While these measures are straightforward, they may be less comprehensive in covering various aspects of the evaluation process. On the other hand, graphical evaluation criteria, exemplified by the receiver operating characteristic (ROC) curve, confusion matrix or Kaplan–Meier survival curves, are recognized for their complexity but are deemed more efficient. In Fig. 14, the prevalent performance metrics, including the AUC, accuracy, ROC curve, recall, precision, specificity, F1-score, MCC and c-index, are illustrated, and 40, 39, 37, 32, 21, 19, 13, 10 and 4 papers are shown. Notably, many studies utilized different performance measures to assess their findings.

In the field of healthcare, the significance of a model extends beyond its quantitative performance, as it is equally crucial to understand the rationale behind its relevance. Clinicians are unlikely to embrace a system that they cannot comprehend; thus, the interpretability of the model plays a crucial role in persuading medical professionals to trust the recommendations provided by the predictive system (Nunnari 2021). In our SLR, we found that out of 47 papers, only 13 considered interpretability, with feature importance for clinical data or gene expression using SHAP or random forest being the most common approach (6 papers). Other methods, such as Grad-CAM (3 papers) or the heatmap generated by the attention mechanism (2 papers), were also employed for histopathological, MRI, US, or mammography studies, highlighting the diverse strategies used to enhance model interpretability in the examined studies. A unique case involved the development of an interpretable classifier based on a flexible utility kernel-based SVM. In general, there is growing awareness of the importance of understanding and explaining the inner workings of ML/DL models. Notably, the absence of interpretability was recognized as a limitation in studies where XAI was not employed. This heightened awareness underscores the significance of ensuring interpretability in model predictions.

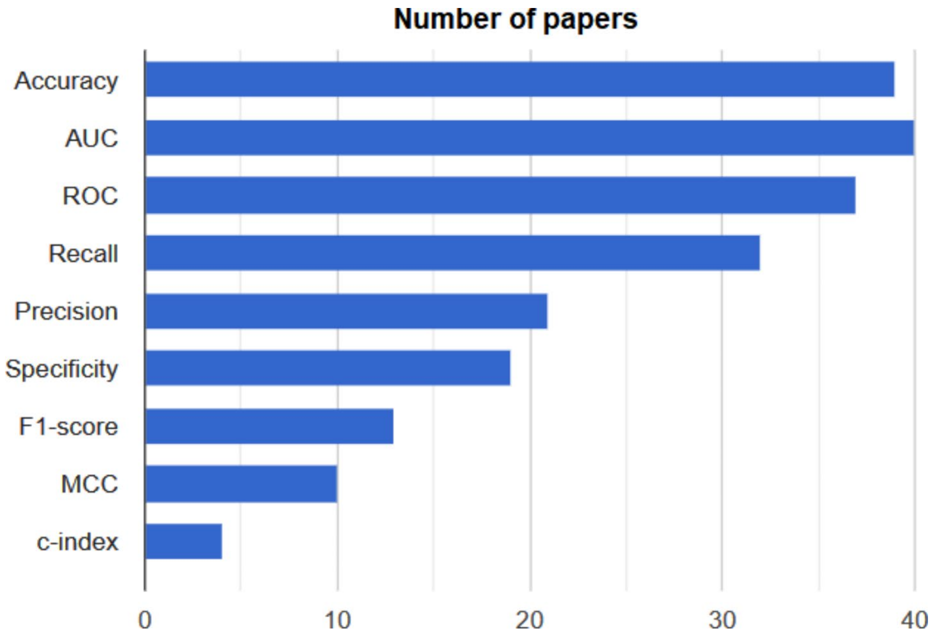

Fig. 14 Distribution of the different validation metrics across the selected studies

10 Overall performance

Figure 15 illustrates the quantity of papers at each performance level; the x-axis represents accuracy intervals ranging from 65 to 100%, with each interval indicating the range of model performance (with the values on the left included and the values on the right excluded). The y-axis displays the number of papers falling within each accuracy interval. The majority of the papers are concentrated in higher accuracy intervals, notably approximately 80% and above, suggesting a prevailing trend toward achieving robust performance. These findings suggest that many models achieve high accuracy in BC classification, reflecting the effectiveness of multimodal fusion approaches. Notably, a peak is observed at an interval of approximately 90% accuracy, where a substantial number of papers demonstrate commendable results. Conversely, only four papers were found in the lower accuracy ranges (between 80% and 85% accuracy), and one paper had 68% accuracy, indicating a focus on achieving higher precision and efficiency in the multimodal fusion models examined in the literature. Additionally, the variability in the number of papers across intervals highlights the diversity in model performances, capturing both successful and challenging instances that will further be discussed.

In all the selected papers, multimodal fusion consistently yielded superior results compared to models employing only a single modality. Figure 16 illustrates the substantial difference in accuracy (if not available AUC) between the multimodal models and their single-modality counterparts, highlighting the significant improvement achieved through the integration of diverse modalities.

Of the 47 papers analyzed, 38 specifically addressed the performance of models in the context of a single modality. In this examination, we sought to compare the effectiveness of the best-performing single-modality model with that of the optimal multimodal model proposed in each paper. Notably, our findings consistently demonstrated that multimodal models consistently outperformed their single-modality counterparts. Remarkably, the per-

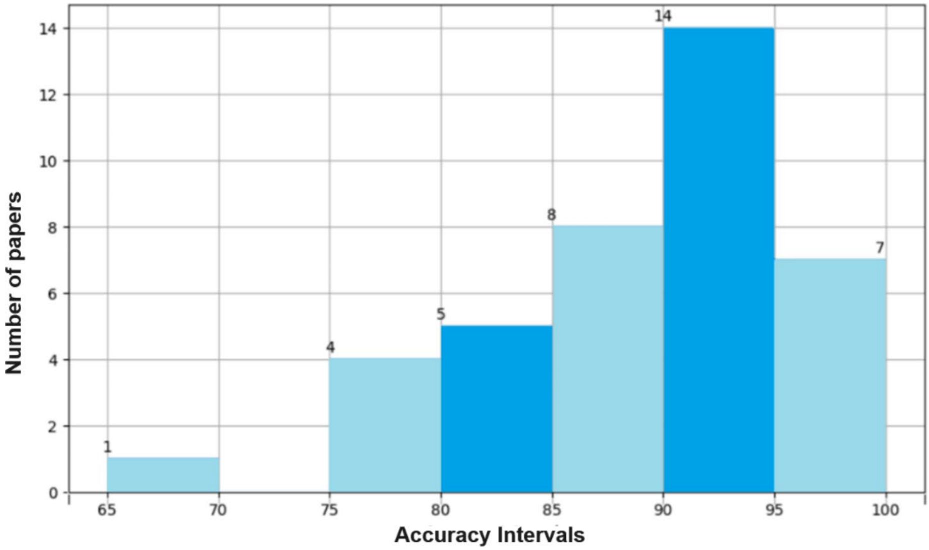

Fig. 15 Distribution of the accuracy of multimodal models across the selected studies

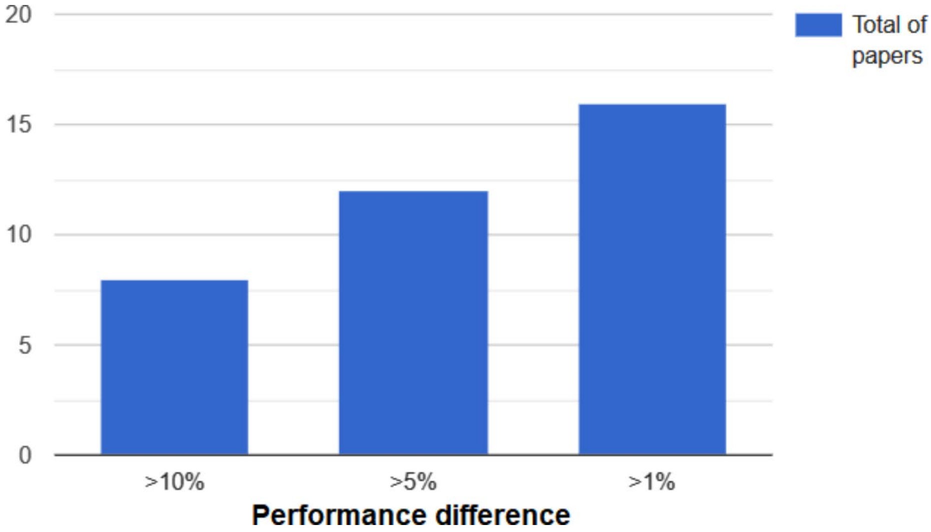

Fig. 16 Distribution of the difference in performance between the multimodal model and the model based on a single modality across the selected studies

formance difference exceeded 10% in 8 instances, demonstrating a substantial improvement attributed to multimodal learning fusion. Additionally, for 12 papers, the performance differential fell between 5% and 10%, underscoring the benefit that multimodal approaches offer. For the remaining papers, the performance variance ranged from 1 to 5%. The significance of these differences underscores the potential of multimodal learning fusion to enhance the predictive capabilities of ML and DL models. By leveraging information from multiple modalities, these models can achieve superior performance compared to relying on a single modality alone, opening avenues for more robust and accurate predictions in various applications. Notably, among the papers employing multiple fusion approaches, feature fusion consistently outperformed decision fusion.

11 Challenges and recommendations for researchers

The field of multimodal BC research is on the rise, though it requires further investigation (Lahat et al. 2015). The authors in Abhisheka et al. (2023) highlight a significant volume of research, comprising more than 200 papers from 2014 to 2023, that leverages a single imaging medical modality for predicting BC. In contrast, our focused approach specifically examines papers that involve the fusion of two medical imaging modalities or integrate a medical image with another modality. The relatively small number of papers in our cohort (14 papers) underscores the limited scope and exploration of multimodal approaches in BC research, emphasizing the need for more comprehensive investigations into the synergistic benefits and challenges posed by the fusion of diverse modalities. This work aims to shed light on the complexities researchers encounter when integrating diverse data modalities for BC classification. By addressing these challenges, the present review aims to provide a comprehensive resource that not only identifies hurdles but also offers valuable recommendations for researchers venturing into the realm of multimodal data fusion for cancer prediction.

- DL algorithms rely heavily on large, high-quality datasets to operate effectively. A primary hurdle lies in the scarcity of extensive datasets that encompass various modalities for each patient, including genetic information, age, images, therapy, and clinical records for training the DL algorithms. The creation of such comprehensive datasets poses significant challenges given the expense associated with collecting sufficient data, especially when dealing with multiple modalities (Lahat et al. 2015). Notably, studies involving mammography, MRI, or US modalities predominantly utilized private datasets. This emphasizes the need for increased attention to these modalities in the creation of future public datasets, fostering broader collaboration and benchmarking opportunities. The limited adoption of this approach may be attributed to potential challenges in interactions and collaborations between academic researchers and physicists, possibly necessitating robust data security measures. To address this challenge, fostering multi-institutional alliances has become a promising approach, enabling the creation of large, diverse datasets that encompass various patient demographics, clinical histories, imaging modalities, and treatment procedures (Tan et al. 2022). Consequently, future research should focus on handling the problem of missing modalities in incomplete data. The availability of an open-source multimodal BC dataset would significantly propel advancements in this area (Luo et al. 2024).
- The predominant reliance on private datasets, often treated as confidential, poses a challenge when attempting to compare the effectiveness of models across different studies. To address the issue of limited data, many analyses have turned to transfer learning, leveraging pretrained models to extract features from image and subsequently training DL models to perform specific tasks. While several studies have utilized pretrained model parameters as a starting point, followed by fine-tuning on medical image datasets, the efficacy of the target model depends heavily on the disparities between the features of the source and target datasets. In the case of smaller datasets, transfer learning has demonstrated effectiveness, albeit contingent on the dissimilarities between the source and target dataset features. Several studies have adopted data augmentation methods (Holste et al. 2021; Muramatsu et al. 2022; Atrey et al. 2023; Li et al. 2020; Zhang et al. 2023; Zhang et al. 2023; Qiao et al. 2022; Zhang 2024; Rabinovici-Cohen et al. 2022) to artificially expand the dataset, enhancing the prediction results. However, unlike new independent images, data augmentation contributes only limited additional information to the DL model.
- An inherent challenge in multimodal fusion is the disparate distribution of heterogeneous data originating from different modalities (Tong et al. 2020). This discrepancy poses difficulties in extracting additional information crucial for an overall interpretation of multimodal data. Despite numerous attempts to integrate multimodal data for cancer prediction, much of the prior research neglects the effort to learn a modality-invariant embedding space that aligns diverse modality distributions. In contrast, some approaches involve assigning a subnetwork to each modality and immediately proceeding with the merger (James et al. 2014). Consequently, the typical modality deviation significantly impacts the effectiveness of the fusion. To address this issue, recent studies have concentrated on feature fusion, primarily based on learned features, using

innovative architectures that efficiently combine modalities, thereby maximizing the utilization of information, as detailed in Sect.7.

- The realm of treatment prediction in AI tends to receive less focus when juxtaposed with the emphasis placed on BC detection and diagnosis (Sugimoto 2023). This emphasis on early detection stems from its potential to improve patient outcomes and survival rates (Ginsburg et al. 2020). However, predicting how an individual patient will respond to various treatment options is a complex endeavor (García-Aranda et al. 2019). Clinical validation and the need for extensive patient outcome data make developing accurate treatment prediction models challenging (El Haji et al. 2023). While numerous studies have made significant strides in the prediction of BC diagnosis and prognosis using multimodal methods, a notable disparity exists in regard to addressing treatment outcome prediction. This approach provides a promising perspective for highlighting the advantages of multimodal models compared to those based on a single modality and for identifying the shortcomings that multimodal models may exhibit.
- The limitations of the existing body of related work underscore the need for a more comprehensive approach to multimodal data fusion analysis; these studies have restricted the exploration of fusion types and techniques. In particular, the majority of the papers combined two modalities, and only 4 papers considered the three modalities at once. Furthermore, the scarcity of extensive combinations when integrating more than two modalities reveals a significant gap in the literature. Out of the 18 studies that involved more than two modalities, only 11 papers used the comprehensive approach of exploring all possible combinations. This indicates a notable trend where a subset of studies has systematically examined the performance of models with varying combinations of multiple modalities, showcasing a thoughtful and thorough approach in the realm of multimodal research. Thus, we propose that researchers address these deficiencies by offering a comprehensive and complete evaluation, comparison, and analysis of various fusion types and modality combinations.
- The use of ensemble methods as powerful tools for combining different data modalities has gained momentum since ensembles are considered optimal for creating more robust architectures and enhancing the overall performance and generalizability of models (Ganaie 2104). The advantage of this approach is that it can potentially reduce the loss of information caused by using a single modality or a single classifier (Osman and Aljahdali 2020), resulting in a more efficient and effective prediction of BC. In our exploration of the selected papers, we found that only 15 out of 47 considered ensemble learning. RF emerged as the most popular choice and was employed in 12 papers, while XGBoost was utilized twice and extreme learning machines (ELMs) were utilized once. Notably, one paper (Othman et al. 2023) innovatively employed a distinctive ensemble learning technique, specifically stacking, for making the final prediction. The proposed method utilizes a hard voting classifier with specific voting criteria, integrating LSTM and a GRU. This ensemble model, trained by separate models, predicts the output class label by consolidating the majority of projected class votes obtained for each class label. Overall, the findings of these studies underscore the versatility of ensemble methods in enhancing classification outcomes in comparison with single classifiers. However, research employing ensemble learning for multimodal classification predominantly centers on the application of the widely used “random forest” technique. This focus overlooks the exploration and evaluation of alternative ensemble methods utilizing

different base learners or different ensemble methods (stacking (Yang et al. 2021), bagging (Nakach 2023) and boosting (Nakach 2024)).

- An observed limitation in the selected papers that use XAI is a tendency to focus on explaining individual modalities, either within the context of a fused study or by explaining each modality separately. This observation raises a crucial question: Can we develop methods to interpret or explain features from different modalities simultaneously?The need for such an approach is paramount for gaining a comprehensive understanding of why specific predictions are made. It also becomes essential for unraveling the intricate relationships that exist between different modalities within a multimodal learning framework. Addressing this challenge can lead to more holistic interpretability (Hakkoum et al. 2022), providing insights not only into the predictions themselves but also into the interplay and synergies between various modalities, contributing to a deeper understanding of the multimodal learning process. This emerging trend highlights the evolving landscape of interpretability, emphasizing the need for advancements that can offer simultaneous insights into the interpretability of diverse modalities within a single, integrated framework.
- Throughout the review, the focus has primarily been on a specific type of multimodal ML fusion. However, during our selection process, we identified papers that employed alternative approaches, broadening the scope of multimodal learning. One such approach is translation in multimodal learning, which involves the conversion or mapping of information between modalities to facilitate joint understanding (Baltrušaitis et al. 2019). An example of translation can be found in He et al. (2020), where the authors propose a BC immunohistochemical benchmark attempting to synthesize immunohistochemical technique data directly with paired hematoxylin and eosin-stained images. Moreover He et al. (2020), demonstrated the development of a DL algorithm that can predict local gene expression from histopathology images, allowing for the identification of spatially resolved molecular biomarkers in BC without the need for retraining on external datasets. Additionally, co-learning is another approach in multimodal learning and is defined as the joint learning of representations from multiple modalities simultaneously. This method aims to leverage complementary information from different modalities to enhance overall model performance (Rahate et al. 2022). Notably, one of the selected papers in our review (Li et al. 2020) aligns with the co-learning approach, indicating the relevance and applicability of such methods in the field of multimodal ML. In the future, there is a potential avenue for researchers to employ translation and co-learning approaches in multimodal learning scenarios for BC classification. The scarcity of studies utilizing these methods adds to their intrigue, making them promising and compelling avenues for exploration.

12 Conclusion and future work

This SLR aimed to provide a comprehensive review of the available literature on multimodal DL techniques for BC classification and revealed a surge in interest and advancements in the field. Our study makes a novel contribution to the domain, as it is the first of its kind. A total of 47 papers published before December 2023 and sourced from six digital libraries were carefully selected, analyzed and classified. The exploration of the literature has unveiled a dynamic landscape in the realm of BC classification, where the fusion of different modalities stands at the forefront of innovative research. Across all the studies, the multimodal models based on data fusion demonstrated enhanced performance compared to models utilizing only single modalities; leveraging multiple modalities enhances the prediction accuracy and provides additional data, contributing to a more comprehensive understanding of the topic. Notably, each fusion approach has its own set of advantages and drawbacks, and understanding these nuances is crucial for selecting the most suitable method for a given task. Decision fusion is acknowledged for its ability to enhance algorithm performance by minimizing errors from different, typically uncorrelated modalities; it excels in scenarios where a specific modality is unavailable, enabling training in the absence of parallel data. However, learning associations among multiple heterogeneous data distributions poses a significant challenge. The semantic correlation and complementary nature of information from these various modalities underscore the need for capturing high-level associations through a compact set of latent variables that decision fusion may fall short in capturing. Thus, when fusing features, the conventional approach of concatenating data descriptors from different sources into a single high-dimensional feature vector has proven ineffective, as it tends to prioritize within-modality correlations over inter-modality correlations. To overcome this challenge, the majority of studies employ DL methods to represent the modalities and extract the features. These methods address the issue by learning joint representations shared across multiple modalities at higher layers of the DNN, following modality-specific network layers. This strategic use of DL mitigates within-modality correlation issues in raw features, facilitating the capture of patterns across data modalities. In this context, researchers have innovatively proposed solutions and architectures to enhance the fusion of learned features. These innovative architectures, although diverse, can be categorized into four distinct patterns. The integration of previously untapped modalities and the pursuit of diverse task objectives promise to contribute significantly to ongoing efforts in advancing cancer prediction methodologies. Nevertheless, researchers still grapple with the challenge of handling diverse multimodal data through the integration of various DL networks at different levels. A primary challenge remains in the limited interpretability of multimodal models, particularly in providing insufficient information about the modality or feature responsible for specific predictions. As the field progresses, addressing this interpretability challenge will be crucial for the broader adoption and ethical deployment of multimodal ML techniques in BC research and healthcare. Consequently, the current study underscores potential avenues for future research, mainly the use of ensemble learning for multimodal fusion, and outlines challenges associated with the adoption of new DL-based methodologies. By narrowing the research scope and delving deeper into other specific multimodal DL aspects, our future work will aim to provide a comprehensive exploration of the literature on translation and co-learning methods based on multiple modalities for BC research.

Appendix

Details on the extracted data from the selected studies and abbreviations list. See Tables 8 and 9.

Table 8 Data extracted from selected papers using feature fusion

Paper	Comb	Dataset	Feature extraction for image	Feature extraction for nonimage	Selection of multimodal features	Fusion strategy	Classifier	Performance
(Yan et al. 2021)	C2	PathoEMR	VGG16	Denoising autoencoder	Feature maps of 3 FCLs	Concat.	FCL	Accuracy: 91% AUC: 94%

(Arya et al. 2023)	C8	TCGA-BRCA	VAE;
			PCA;
				VAE; PCA		Concat.	SVM with different kernels; RF	Accuracy: 77% Precision: 76% Recall: 100% F1-Score: 87%

(Yang et al. 2022)	C2	TCGA-BRCA	ResNet50	RF for feature importance		Multimodal Compact Bilinear	FCL	AUC: 72% Recall: 67% Specificity: 83%
(Guo et al. 2021)	C6	TCGA-BRCA METABRIC	-	DNN and affinity fusion module for genomics	Attention module	Concat.	FCL; SVM; LR; RF	Accuracy: 89% AUC: 94% Precision: 91% Recall: 94% F1-Score: 92%
(Arya and Saha 2020)	C6	TCGA-BRCA METABRIC	-	CNN		Concat.	RF; SVM; NB; LR	Accuracy: 88% AUC: 97% Precision: 95% Recall: 95%
(Arya et al. 2021)	C6	TCGA-BRCA METABRIC	-	CNN	Gated-attention layer	Concat.	RF; SVM; NB; LR	Accuracy: 91% AUC: 95% Precision: 84% Recall: 80% C-Index: 76%
(Chen et al. 2019)	C1	METABRIC	-	DNN	Nonnegative Matrix Factorization and Attention module	Concat.	DNN; SVM; RF; LR	Accuracy: 84% AUC: 87% Precision: 85% Recall: 91% MCC: 93%
(Li 2020)	C11	TCGA-BRCA	CNN; DNN; LSTM	AE	-	Conditional autoencoder	FCL	Accuracy: 97% C-Index: 76%
(Mokni et al. 2021)	C13	Private	Gradient Local Information Pattern	-	Canonical Correlation Analysis	Concat.; Summation	RBFNN; SVM; RF; ANN; KNN	Accuracy: 98% AUC: 99% Recall: 86%

Table 8 (continued)

Paper	Comb	Dataset	Feature extraction for image	Feature extraction for nonimage	Selection of multimodal features	Fusion strategy	Classifier	Performance
(Yan et al. 2019)	C2	Private	VGG16		Feature maps of 3 FCLs	Concat.	FCL	Accuracy: 90% AUC: 94%
(Muramatsu et al. 2022)	C15	Private	EfficientNetB3	-		Concat.	FCL	Accuracy: 77% F1-score: 59%
(Atrey et al. 2023)	C15	Private	CNN	-	Statistical significance analysis	Concat.	LSTM; SVM with various kernels	Accuracy: 95% AUC: 95% Recall: 94% Specificity: 96% MCC: 89%
(Othman et al. 2023)	C6	METABRIC	-	CNN	50% dropout layer for CN and GE	Concat.	Ensemble of LSTM and GRU with hard voting	Accuracy: 98% AUC: 98% Recall: 99% Specificity: 99% C-Index: 94%
(Li et al. 2020)	C19	Private	VGG16			Concat.	VGG16	Accuracy: 89% AUC: 95% Recall: 87% Specificity: 90%
(Wang et al. 2021)	C11	TCGA-BRCA	Bilinear feature encoding module		Capture the complex intra and inter-modality relations	Concat.	Multilayer DNN	Accuracy: 80% AUC: 82% C-Index: 72%
(Yao et al. 2022)	C7	Private	DNN;		Attention module	Concat.	FCL	Accuracy: 80% AUC: 75%
(Joo et al. 2021)	C4	Private	3D-ResNet	FCL		Concat.	FCL	Accuracy: 85% AUC: 89% Recall: 67% Specificity: 93%

Table 8 (continued)

Paper	Comb	Dataset	Feature extraction for image	Feature extraction for nonimage	Selection of multi-modal features	Fusion strategy	Classifier	Performance
(Zhang et al. 2023)	C15	Private	Resnet50	-	Self-attention inter & intra modality	Concat.	FCL	Accuracy: 89% Precision: 88% Recall: 85% C-Index: 84%
(Kayikci et al. 2023)	C6	TCGA-BRCA METABRIC	-	CNN	Bimodal attention	Concat.	FCL	Accuracy: 91% AUC: 95% Recall: 45%
(Luo et al. 2023)	C5	Private	Spatial attention with Xception	FCL with output similar to imaging features	Channel attention	Concat.	FCL	Accuracy: 91% AUC: 95% Recall: 93% Specificity: 89% MCC: 92%
(Wang et al. 2023)	C2	Private TCGA-BRCA	ResNet50; K-means	Multiple instance learning	Graph attention networks; Multihead attention networks	Concat.; Linear add; Outer product	FCL	C-Index: 73%
(Xu et al. 2021)	C2	BCNB	VGG16	Multi-instance learning		Concat.	FCL	Accuracy: 76% AUC: 83% Recall: 89% C-Index: 67%
(Du et al. 2023)	C6	TCGA-BRCA; METABRIC;	-	Multiscale bilinear CNN	Decoder and discriminator	Concat.	Extreme random Trees	A ccuracy: 93% Precision: 91% Recall: 81% F1-Score: 86% MCC: 79%
(Sun et al. 2018)	C10	TCGA-BRCA	Multiple kernel learning	Multiple kernel learning		Concat.	Multiple kernel learning	Accuracy: 83% Specificity: 95%
(Zhang et al. 2023)	C20	Private	VGG11	-		Concat.	FCL	Accuracy: 83% AUC: 95% Recall: 95% Specificity: 60%

Table 8 (continued)

Paper	Comb	Dataset	Feature extraction for image	Feature extraction for nonimage	Selection of multi-modal features	Fusion strategy	Classifier	Performance
(Yuan and Xu 2023)	C6	METABRIC		CNN; LSTM;	The attention mechanism	Gated multimodal unit	MLP; RF	Accuracy: 94% AUC: 96% Recall: 85% C-Index: 87%
(Qu et al. 2023)	C12	Private	Quantum CNN	-	Quantum ML model	Concat.	Variational quantum classifier	Accuracy: 97% AUC: 97% Recall: 97%
(Wang et al. 2020)	C6	TCGA-BRCA	-	Sample similarity; Feature matrix;		Similarity network fusion algorithm	Graph convolutional network	Accuracy: 79% AUC: 93% Precision: 77% Recall: 99%
(Mustafa et al. 2023)	C6	METABRIC	-	CNN; DNN; LSTM		Concat.	Random forest	Accuracy: 98% AUC: 97% Recall: 100% Specificity: 95% F1-Score: 99%
(Chharia et al. 2021)	C1	TCGA-BRCA	-	CNN	Weighted k-NN and mRMR Feature Selection	Concat.	Fuzzy classifier	Accuracy: 87% AUC: 93% Recall: 89% Specificity: 91%
(Dhillon et al. 2020)	C11	TCGA-BRCA	Cell Profiler	FSelector		Concat.	ELM with gradient-based boosting	Accuracy: 85% AUC: 85% Recall: 83% Specificity: 75% MCC: 56%
(Barros et al. 2023)	C3	Private	CNN	XGBoost + SHAP		Concat.	XGBoost	AUC: 85%
(Kayikci and Khoshgoftaar 2022)	C6	METABRIC	-	3 CNNs		Concat.	RF; DT; SVM	Accuracy: 90% C-Index: 67% MCC: 93%

Table 8 (continued)

Paper	Comb	Dataset	Feature extraction for image	Feature extraction for nonimage	Selection of multimodal features	Fusion strategy	Classifier	Performance
(Arya et al. 2023)	C8	TCGA-BRCA	ResNet152 ; PCA			Concat.	SVM Utility Kernel	C-Index: 94%
(Qiao et al. 2022)	C14	Private	2D-ResNet; 3D-ResNet		Feature fusion module based on affinity model	Concat.	MUM-Net	Accuracy: 76% AUC: 86% Precision: 92% Recall: 56% Specificity: 95%
(Wu et al. 2023)	C5	Private	D-BUS-NET	T test		Concat.	DNN	Accuracy: 93% AUC: 98% Recall: 94% Specificity: 92%
(Yala et al. 2019)	C3	Private	ResNet18	LR; RF;	Hybrid DL model that used both traditional risk factors and mammograms	Concat.	FCL	Accuracy: 70%
(Li et al. 2021)	C2	Private	EfficientNet-B0	TabNet	Attention-based Multi-instance Learning fusion module	Concat.	FCL	AUC: 88% Precision: 75% Recall: 83% MCC: 79%
(Jadoon et al. 2023)	C6	Private	CNN	CNN; DNN;		Concat.	RF	Accuracy: 97% AUC: 90% Precision: 98% Recall: 97% F1-score: 98%
(Zhang 2024)	C15	Private	ResNet50	-	Attention	Concat.	MLP	AUC: 86% MCC: 79%
(Fu et al. 2023)	C18	Private	CNN	Clinical embedding module	Graph attentional layer	Concat.	FCL	C-index: 75%
(Furtney et al. 2023)	C9	TCGA-BRCA	Graph neural network	Autoencoder; GNN	DL feature representation		GNN	AUC: 87%

Table 8 (continued)

Paper	Comb	Dataset	Feature extraction for image	Feature extraction for nonimage	Selection of multimodal features	Fusion strategy	Classifier	Performance
(Holste et al. 2021)	C4	Private	ReseNet50	DNN		Concat.; Elementwise addition and multiplication	FCL	AUC: 90% Specificity: 95% C-Index: 49%

Table 9 Data extracted from selected papers using decision-fusion and hybrid fusion

Paper	MC	Target class	Dataset	Classifier		Fusion method	Performance (%)
				Imaging	Nonimaging		
(Liu 2021)	C10	4 molecular subtypes	TCGA	VGG16	DNN	Weighted linear aggregation	Accuracy: 88% AUC: 94%
(Sun et al. 2019)	C6	Short-term/long-term survival	METABRIC TCGA	-	DNN	Weighted linear aggregation	Accuracy: 83% AUC: 85% Precision: 75% Specificity: 95% F1-Score: 49%
(Arya and Saha 2020)	C6	Short-term/long-term survival	TCGA METABRIC	- -	CNN CNN	RF	Accuracy: 90% AUC: 93% Precision: 84% Recall: 75% F1-Score: 73%
(Mullen et al. 2023)	C16+C17	Short-term/long-term survival	Private	SVM		Joint Bayesian classifier	Accuracy: 98% AUC: 91%
(Holste et al. 2021)	C4	Benign/malignant tumors	Private	ResNet50	DNN	FCL	AUC: 90% Specificity: 95% C-Index: 49%
(Rabinovici-Cohen et al. 2022)	C4	Reccurence or no-reccurence	Private	CNN base on ResNet	RF; LR; XGBoost	Ensemble: Mean value of 3 models for each modality	Accuracy: 63% AUC: 75% Specificity: 90% C-Index: 57%

Acknowledgements This work was conducted under the research project "ML based Breast Cancer Diagnosis and Treatment", 2023–2025. The authors would like to thank the Moroccan Ministry of Higher Education and Scientific Research, Digital Development Agency (ADD), CNRST, and UM6P for their support.

Author contributions All authors contributed equally to this article.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Abdou MA (2022) Literature review: efficient deep neural networks techniques for medical image analysis. Neural Comput Appl 34(8):5791–5812

Abdullakutty F, Akbari Y, Al-Maadeed S, Bouridane A, Hamoudi R (2024) Advancing histopathology-based breast cancer diagnosis: insights into multi-modality and explainability. arXiv. http://arxiv.org/abs/2406.12897. Accessed 7 Jul 2024

Abhisheka B, Biswas SK, Purkayastha B (2023) A comprehensive review on breast cancer detection, classification and segmentation using deep learning. Arch Computat Methods Eng 30(8):5023–5052

Akkus C, Chu L, Djakovic V, Jauch-Walser S, Koch P, Loss G et al (2023) Multimodal deep learning. arXiv. http://arxiv.org/abs/2301.04856. Accessed 21 Sep 2023.

Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Mol Pharm 13(7):2524–2530

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O et al (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8(1):53

Aremu OO, Hyland-Wood D, McAree PR (2020) A machine learning approach to circumventing the curse of dimensionality in discontinuous time series machine data. Reliab Eng Syst Saf 195:106706

Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M et al (2022) Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast 66:15–23

Arya N, Saha S (2020) Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. IEEE/ACM Trans Comput Biol Bioinf 1–1

Arya N, Saha S (2021) Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowl Based Syst 221:106965

Arya N, Saha S, Mathur A, Saha S (2023) Improving the robustness and stability of a machine learning model for breast cancer prognosis through the use of multi-modal classifiers. Sci Rep 13(1):4079

Arya N, Mathur A, Saha S, Saha S (2023) Proposal of SVM utility kernel for breast cancer survival estimation. IEEE/ACM Trans Comput Biol Bioinf 20(2):1372–1383

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 Signal Process Control 86:104919

Bahl M (2022) Updates in artificial intelligence for breast imaging. Sem Roentgenol 57(2):160–167

Baltrušaitis T, Ahuja C, Morency LP (2019) Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell 41(2):423–443

Barros V, Tlusty T, Barkan E, Hexter E, Gruen D, Guindy M et al (2023) Virtual biopsy by using artificial intelligence–based multimodal modeling of binational mammography data. Radiology 306(3):e220027

Battleday RM, Peterson JC, Griffiths TL (2021) From convolutional neural networks to models of higher-level cognition (and back again). Ann N Y Acad Sci 1505(1):55–78

Brito-Sarracino T, Rocha dos Santos M, Freire Antunes E, Batista de Andrade Santos I, Coelho Kasmanas J (2019) Ponce de Leon Ferreira de Carvalho AC. Explainable machine learning for breast cancer diagnosis. In: 2019 8th Brazilian Conference on Intelligent Systems (BRACIS). Salvador, Brazil: IEEE. pp. 681–6. https://ieeexplore.ieee.org/document/8923961/. Accessed 24 Apr 2022

cBioPortal for Cancer Genomics. https://www.cbioportal.org/study/summary?id=brca_metabric. Accessed 26 Dec 2023

Chen H, Gao M, Zhang Y, Liang W, Zou X (2019) Attention-based multi-NMF deep neural network with multimodality data for breast cancer prognosis model. Biomed Res Int 2019:e9523719

Cheng J, Gao M, Liu J, Yue H, Kuang H, Liu J et al (2021) Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading. IEEE J Biomedical Health Inf

Chharia A, Kumar N (2021) Foreseeing survival through 'fuzzy intelligence': a cognitively-inspired incremental learning based de novo model for breast cancer prognosis by multi-omics data fusion. In: Rekik I, Adeli E, Park SH, Schnabel J (eds) Predictive Intelligence in Medicine. Springer International Publishing, Cham, pp 231–242. (Lecture Notes in Computer Science).

Chu C, Zhmoginov A, Sandler M CycleGAN, a Master of Steganography. arXiv; 2017. http://arxiv.org/abs/1712.02950. Accessed 28 Dec 2023

Choi Y, El-Khamy M, Lee J (2019) Variable rate deep image compression with a conditional autoencoder. pp. 3146–54. https://openaccess.thecvf.com/content_ICCV_2019/html/Choi_Variable_Rate_Deep_Image_Compression_With_a_Conditional_Autoencoder_ICCV_2019_paper.html. Accessed 28 Dec 2023

Dey R, Salem FM (2017) Gate-variants of gated recurrent unit (GRU) neural networks. In: 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). pp. 1597–600. https://ieeexplore.ieee.org/abstract/document/8053243. Accessed 28 Dec 2023

Dhillon A, Singh A (2020) eBreCaP: extreme learning-based model for breast cancer survival prediction. IET Syst Biol 14(3):160–169

Du X, Zhao Y (2023) Multimodal adversarial representation learning for breast cancer prognosis prediction. Comput Biol Med 157:106765

El Haji H, Souadka A, Patel BN, Sbihi N, Ramasamy G, Patel BK et al (2023) Evolution of breast cancer recurrence risk prediction: a systematic review of statistical and machine learning–based models. JCO Clin Cancer Inf. (7):e2300049

ElOuassif B, Idri A, Hosni M, Abran A (2021) Classification techniques in breast cancer diagnosis: a systematic literature review. null 9(1):50–77

Fu X, Patrick E, Yang JYH, Feng DD, Kim J (2023) Deep multimodal graph-based network for survival prediction from highly multiplexed images and patient variables. Comput Biol Med 154:106576

Furtney I, Bradley R, Kabuka MR (2023) Patient graph deep learning to predict breast cancer molecular subtype. IEEE/ACM Trans Comput Biol Bioinform 20(5):3117–3127

Ganaie MA, Hu M, Tanveer* M, Suganthan* PN (2021) Ensemble deep learning: a review. arXiv:210402395. http://arxiv.org/abs/2104.02395. Accessed 9 Sep 2021.

García-Aranda M, Redondo M (2019) Immunotherapy: a challenge of breast cancer treatment. Cancers 11(12):1822

Ginsburg O, Yip CH, Brooks A, Cabanes A, Caleffi M, Yataco JAD et al (2020) Breast cancer early detection: a phased approach to implementation. Cancer 126(S10):2379–2393

Gondara L (2016) Medical image denoising using convolutional denoising autoencoders. In: IEEE 16th international conference on data mining workshops (ICDMW). pp. 241–6. https://ieeexplore.ieee.org/abstract/document/7836672. Accessed 28 Dec 2023

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al (2020) Generative adversarial networks. Commun ACM 63(11):139–144

Guo Y, Shi H, Kumar A, Grauman K, Rosing T, Feris R (2018) SpotTune transfer learning through adaptive fine-tuning.10

Guo W, Liang W, Deng Q, Zou X (2021) A multimodal affinity fusion network for predicting the survival of breast cancer patients. Front Genet. 12. https://www.frontiersin.org/articles/https://doi.org/10.3389/fgene.2021.709027. Accessed 28 Oct 2023

Hakkoum H, Abnane I, Idri A (2022) Interpretability in the medical field: a systematic mapping and review study. Appl Soft Comput 117:108391

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674

He B, Bergenstråhle L, Stenbeck L, Abid A, Andersson A, Borg Å et al (2020) Integrating spatial gene expression and breast tumour morphology via deep learning. Nat Biomed Eng 4(8):827–834

Hemant Kumar AVS, Tripathi R, Agrawal S, Kumar (2021) Transfer learning and supervised machine learning approach for detection of skin cancer: performance analysis and comparison. dcth 10(1):1845–1860

Holste G, Partridge SC, Rahbar H, Biswas D, Lee CI, Alessio AM (2021) End-to-end learning of fused image and non-image features for improved breast cancer classification from MRI. In: 2021 IEEE/CVF international conference on computer vision workshops (ICCVW). pp. 3287–96

Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP (2020) Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. Npj Digit Med 3(1):136

Ito FT, Caseli H, de Moreira M J (2018) The effects of unimodal representation choices on multimodal learning. 8

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

Jain AK, Ross A (2004) Multibiometric systems. Commun ACM 47(1):34–40

James AP, Dasarathy BV (2014) Medical image fusion: a survey of the state of the art. Inform Fusion 19:4–19

Jetley S, Lord NA, Lee N, Torr PHS (2018) Learn to pay attention. arXiv; 2018. http://arxiv.org/abs/1804.02391. Accessed 20 Mar 2023

Joo S, Ko ES, Kwon S, Jeon E, Jung H, Kim JY et al (2021) Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Sci Rep 11(1):18800

Kayikci S, Khoshgoftaar T (2022) A stack based multimodal machine learning model for breast cancer diagnosis. In: 2022 International congress on human-computer interaction, optimization and robotic applications (HORA). pp. 1–5

Kayikci S, Khoshgoftaar TM (2023) Breast cancer prediction using gated attentive multimodal deep learning. J Big Data 10(1):62

Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. 2

Lahat D, Adali T, Jutten C (2015) Multimodal data fusion: an overview of methods, challenges, and prospects. Proc IEEE 103(9):1449–1477

Li X, Qin G, He Q, Sun L, Zeng H, He Z et al (2020) Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. Eur Radiol 30(2):778–788

Li H, Yang F, Xing X, Zhao Y, Zhang J, Liu Y et al (2021) Multi-modal multi-instance learning using weakly correlated histopathological images and tabular clinical information. In: Medical image computing and computer assisted intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII. Berlin, Heidelberg: Springer-Verlag. pp. 529–39. https://doi.org/10.1007/978-3-030-87237-3_51. Accessed 5 Nov 2022

Li B, Oka R, Xuan P, Yoshimura Y, Nakaguchi T (2022) Robust multi-modal prostate cancer classification via feature autoencoder and dual attention. Inf Med Unlocked 30:100923

Li X, Zhou Y, Wang J, Lin H, Zhao J, Ding D et al (2021) Multi-modal multi-instance learning for retinal disease recognition. In: Proceedings of the 29th ACM international conference on multimedia. New York, NY, USA: Association for Computing Machinery; pp. 2474–82. https://doi.org/10.1145/3474085.3475418. Accessed 2 May 2022

Li S, Shi H, Sui D, Hao A, Qin H (2020) A Novel Pathological Images and Genomic Data Fusion Framework for Breast Cancer Survival Prediction. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1384–7. https://ieeexplore.ieee.org/document/9176360. Accessed 7 Jan 2024

Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D et al (2022) Artificial intelligence for multimodal data integration in oncology. Cancer Cell 40(10):1095–1110

Liu T, Huang J, Liao T, Pu R, Liu S, Peng Y (2021) A Hybrid Deep Learning Model for Predicting Molecular Subtypes of Human Breast Cancer Using Multimodal Data. IRBM. https://www.sciencedirect.com/science/article/pii/S1959031820301858. Accessed 31 Aug 2021

Logan R, Williams BG, Ferreira da Silva M, Indani A, Schcolnicov N, Ganguly A et al (2021) Deep convolutional neural networks with ensemble learning and generative adversarial networks for Alzheimer's disease image data classification. Front Aging Neurosci 13:497

Lu J, Steeg PS, Price JE, Krishnamurthy S, Mani SA, Reuben J et al (2009) Breast cancer metastasis: challenges and opportunities. Cancer Res 69(12):4951–4953

Luo Y, Lu Z, Liu L, Huang Q (2023) Deep fusion of human-machine knowledge with attention mechanism for breast cancer diagnosis. Biomed Signal Process Control 84:104784

Luo L, Wang X, Lin Y, Ma X, Tan A, Chan R et al (2024) Deep learning in breast cancer imaging: a decade of progress and future directions. IEEE Rev Biomed Eng 1–20

Madani M, Behzadi MM, Nabavi S (2022) The role of deep learning in advancing breast cancer detection using different imaging modalities: a systematic review. Cancers 14(21):5334

Mahmood T, Li J, Pei Y, Akhtar F, Imran A, Rehman KU (2020) A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities. IEEE Access 8:165779–165809

Mathur A, Arya N, Pasupa K, Saha S, Roy Dey S, Saha S (2024) Breast cancer prognosis through the use of multi-modal classifiers: current state of the art and the way forward. Brief Funct Genomics. elae015

Metzger-Filho O, Sun Z, Viale G, Price KN, Crivellari D, Snyder RD et al (2013) Patterns of recurrence and outcome according to breast cancer subtypes in lymph node–negative disease: results from international breast cancer study group trials VIII and IX. J Clin Oncol 31(25):3083–3090

Mokni R, Gargouri N, Damak A, Sellami D, Feki W, Mnif Z (2021) An automatic computer-aided diagnosis system based on the multimodal fusion of breast cancer (MF-CAD). Biomed Signal Process Control 69:102914

Mullen LA, Walton WC, Williams MP, Peyton KS, Porter DW (2023) Breast cancer detection with upstream data fusion, machine learning, and automated registration: initial results. J Med Imaging (Bellingham) 10(Suppl 2):S22409

Muramatsu C, Iwasaki T, Oiwa M, Kawasaki T, Fujita H (2022) Classification of intrinsic subtypes and histological grade for breast cancers by multimodality images. In: 16th International Workshop on Breast Imaging (IWBI2022). SPIE. pp. 228–33. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12286/122860Y/Classification-of-intrinsic-subtypes-and-histological-grade-for-breast-cancers/https://doi.org/10.1117/12.2625871.full. Accessed 7 Jan 2024

Murtaza G, Shuib L, Abdul Wahab AW, Mujtaba G, Mujtaba G, Nweke HF et al (2020) Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artif Intell Rev 53(3):1655–1720

Mustafa E, Jadoon EK, Khaliq-uz-Zaman S, Humayun MA, Maray M (2023) An ensembled framework for human breast cancer survivability prediction using deep learning. Diagnostics 13(10):1688

Nakach FZ (2024) Hybrid deep boosting ensembles for histopathological breast cancer classification. Health Technol 18

Nakach FZ, Idri A, Zerouaoui H (2023) Deep hybrid bagging ensembles for classifying histopathological breast cancer images. pp. 289–300. https://www.scitepress.org/Link.aspx?doi=10.5220/0011704200003393. Accessed 23 May 2023

Nassif AB, Talib MA, Nasir Q, Afadar Y, Elgendy O (2022) Breast cancer detection using artificial intelligence techniques: a systematic literature review. Artif Intell Med 127:102276

Nunnari F, Sonntag D (2021) A software toolbox for deploying deep learning decision support systems with XAI capabilities. In: Companion of the 2021 ACM SIGCHI symposium on engineering interactive computing systems. New York, NY, USA: Association for Computing Machinery. pp. 44–9. (EICS '21). https://doi.org/10.1145/3459926.3464753. Accessed 26 Apr 2022

Osman A, Aljahdali HM (2020) An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model. IEEE Access

Othman NA, Abdel-Fattah MA, Ali AT (2023) A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction. Big Data Cogn Comput 7(1):50

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71

Pei X, Zuo K, Li Y, Pang Z (2023) A review of the application of multi-modal deep learning in medicine: bibliometrics and future directions. Int J Comput Intell Syst 16(1):44

Qiao M, Liu C, Li Z, Zhou J, Xiao Q, Zhou S et al (2022) Breast tumor classification based on MRI-US images by disentangling modality features. IEEE J Biomed Health Inf 26(7):3059–3067

Qu Z, Li Y, Tiwari P (2023) QNMF: a quantum neural network based multimodal fusion system for intelligent diagnosis. Inform Fusion 100:101913

Rabinovici-Cohen S, Fernández XM, Grandal Rejo B, Hexter E, Hijano Cubelos O, Pajula J et al (2022) Multimodal prediction of five-year breast cancer recurrence in women who receive neoadjuvant chemotherapy. Cancers (Basel) 14(16):3848

Rahate A, Walambe R, Ramanna S, Kotecha K (2022) Multimodal co-learning: challenges, applications with datasets, recent advances and future directions. Inform Fusion 81:203–239

Romeo V, Accardo G, Perillo T, Basso L, Garbino N, Nicolai E et al (2021) Assessment and prediction of response to neoadjuvant chemotherapy in breast cancer: a comparison of imaging modalities and future perspectives. Cancers 13(14):3521

Salvi M, Loh HW, Seoni S, Barua PD, García S, Molinari F et al (2024) Multi-modality approaches for medical support systems: a systematic review of the last decade. Inform Fusion 103:102134

Singh LK, Khanna M, Pooja (2022) A novel multimodality based dual fusion integrated approach for efficient and early prediction of glaucoma. Biomed Signal Process Control 73:103468

Stahlschmidt SR, Ulfenborg B, Synnergren J (2022) Multimodal deep learning for biomedical data fusion: a review. Brief Bioinform 23(2):bbab569

Steyaert S, Pizurica M, Nagaraj D, Khandelwal P, Hernandez-Boussard T, Gentles AJ et al (2023) Multimodal data fusion for cancer biomarker discovery with deep learning. Nat Mach Intell 5(4):351–362

Sugimoto M, Hikichi S, Takada M, Toi M (2023) Machine learning techniques for breast cancer diagnosis and treatment: a narrative review. Ann Breast Surg 7(0). https://abs.amegroups.org/article/view/7085. Accessed 22 Sept 2023

Sun D, Li A, Tang B, Wang M (2018) Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome. Comput Methods Programs Biomed 161:45–53

Sun D, Wang M, Li A (2019) A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. IEEE/ACM Trans Comput Biol Bioinf 16(3):841–850

Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In: Thirty-First AAAI Conference on Artificial Intelligence. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806. Accessed 25 Feb 2022

Tan XJ, Cheor WL, Lim LL, Ab Rahman KS, Bakrin IH (2022) Artificial Intelligence (AI) in breast imaging: a scientometric umbrella review. Diagnostics 12(12):3111

Taud H, Mas JF (2018) Multilayer perceptron (MLP). In: Camacho Olmedo MT, Paegelow M, Mas JF, Escobar F (Eds) Geomatic approaches for modeling land change scenarios (Lecture notes in geoinformation and cartography). Cham: Springer International Publishing. pp. 451–5. https://doi.org/10.1007/978-3-319-60801-3_27

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. https://doi.org/10.1007/s11042-023-16634-w

The Cancer Genome Atlas Program (TCGA)—NCI. https://www.cancer.gov/ccg/research/genome-sequencing/tcga. Accessed 1 Oct 2023

Tong L, Mitchel J, Chatlin K, Wang MD (2020) Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis. BMC Med Inf Decis Mak 20(1):225

Wan Y, Shu J, Sui Y, Xu G, Zhao Z, Wu J et al (2019) Multi-modal attention network learning for semantic source code retrieval. In: Proceedings of the 34th IEEE/ACM international conference on automated software engineering. San Diego, California: IEEE Press. pp. 13–25. (ASE '19). https://doi.org/10.1109/ASE.2019.00012. Accessed 23 May 2022

Wang F, Han J (2009) Multimodal biometric authentication based on score level fusion using support vector machine. Opto-Electron Rev 17:59–64

Wang C, Guo J, Zhao N, Liu Y, Liu X, Liu G et al (2020) A cancer survival prediction method based on graph convolutional network. IEEE Trans Nanobiosci 19(1):117–126

Wang Z, Li R, Wang M, Li A (2021) GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction. Bioinformatics. btab185

Wang Y, Zhang L, Li Y, Wu F, Cao S, Ye F (2023) Predicting the prognosis of HER2-positive breast cancer patients by fusing pathological whole slide images and clinical features using multiple instance learning. Math Biosci Eng 20(6):11196–11211

Weng L (2019) From From GAN to WGAN. arXiv. http://arxiv.org/abs/1904.08994. Accessed 28 Dec 2023

Wu Y, Wei L, Duan Y (2019) Deep spatiotemporal LSTM network with temporal pattern feature for 3D human action recognition. Comput Intell 35(3):535–554

Wu P, Jiang Y, Xing H, Song W, Cui X, Wu X et al (2023) long,. Multimodality deep learning radiomics nomogram for preoperative prediction of malignancy of breast cancer: a multicenter study. Phys Med Biol 68(17):175023

Xu F, Zhu C, Tang W, Wang Y, Zhang Y, Li J et al (2021) Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides. Front Oncol. https://doi.org/10.3389/fonc.2021.759007

Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R (2019) A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 292(1):60–66

Yan R, Ren F, Rao X, Shi B, Xiang T, Zhang L et al (2019) Integration of multimodal data for breast cancer classification using a hybrid deep learning method. In: Huang DS, Bevilacqua V, Premaratne P (Eds) Intelligent computing theories and application (Lecture Notes in Computer Science; vol. 11643). Cham: Springer International Publishing. pp. 460–9. https://doi.org/10.1007/978-3-030-26763-6_44

Yan R, Zhang F, Rao X, Lv Z, Li J, Zhang L et al (2021) Richer fusion network for breast cancer classification based on multimodal data. BMC Med Inf Decis Mak 21(1):134

Yang Y, Wei L, Hu Y, Wu Y, Hu L, Nie S (2021) Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning. J Neurosci Methods 350:109019

Yang J, Ju J, Guo L, Ji B, Shi S, Yang Z et al (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

Yao Y, Lv Y, Tong L, Liang Y, Xi S, Ji B et al (2022) ICSDA: a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data. Brief Bioinform 23(6):bbac448

Yassin NIR, Omran S, El Houby EMF, Allam H (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. Comput Methods Programs Biomed 156:25–45

Yin C, Zhu Y, Fei J, He X (2017) A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access 5:21954–21961

Yuan H, Xu H (2023) Deep multi-modal fusion network with gated unit for breast cancer survival prediction. Comput Methods Biomech BioMed Eng 0(0):1–14

Yuan Y, Giger ML, Li H, Bhooshan N, Sennett CA (2010) Multimodality computer-aided breast cancer diagnosis with FFDM and DCE-MRI. Acad Radiol 17(9):1158–1167

Zerouaoui H, Idri A (2021) Reviewing machine learning and image Processing Based decision-making systems for breast Cancer imaging. J Med Syst 45(1):8

Zhang D, Zou L, Zhou X, He F (2018) Integrating feature selection and feature extraction methods with Deep Learning to predict clinical outcome of breast cancer. IEEE Access 6:28936–28944

Zhang T, Tan T, Han L, Appelman L, Veltman J, Wessels R et al (2023) Predicting breast cancer types on and beyond molecular level in a multi-modal fashion. NPJ Breast Cancer 9(1):16

Zhang M, Xue M, Li S, Zou Y, Zhu Q (2023) Fusion deep learning approach combining diffuse optical tomography and ultrasound for improving breast cancer classification. Biomed Opt Express 14(4):1636–1646

Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z et al (2020) Graph neural networks: a review of methods and applications. AI Open 1:57–81

Zhang T, Han L, Gao Y, Wang X, Beets-Tan R, Mann R (2024) Predicting molecular subtypes of breast cancer using multimodal deep learning and incorporation of the attention mechanism

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Fatima-Zahrae Nakach[1] · **Ali Idri**[2] · **Evgin Goceri**[3]

✉ Fatima-Zahrae Nakach
fatimazahra.nakach@um6p.ma

Ali Idri
ali.idri@um6p.ma

Evgin Goceri
evgin@akdeniz.edu.tr

[1] Faculty of Medical Sciences, UM6P-Mohammed VI Polytechnic University, Hay Moulay Rachid, Ben Guerir, Marrakech-Safi 43150, Morocco

[2] ENSIAS, UM5-Mohammed V University, Rabat-Salé-Kénitra, Av. Regragui, Rabat 10000, Morocco

[3] Biomedical Engineering Department, Akdeniz University, Antalya 07070, Turkey