Discover Oncology

Review

Harnessing artificial intelligence for predicting breast cancer recurrence: a systematic review of clinical and imaging data

Jaqueline Alvarenga Silveira[1] · Alexandre Ray da Silva[1] · Mariana Zuliani Theodoro de Lima[1,2]

Received: 17 October 2024 / Accepted: 3 February 2025

Abstract

Breast cancer is a leading cause of mortality among women, with recurrence prediction remaining a significant challenge. In this context, artificial intelligence application and its resources can serve as a powerful tool in analyzing large amounts of data and predicting cancer recurrence, potentially enabling personalized medical treatment and improving the patient's quality of life. Thus, the systematic review examines the role of AI in predicting breast cancer recurrence using clinical data, imaging data, and combined datasets. Support Vector Machine (SVM) and Neural Networks, especially when applied to combined data, demonstrate strong potential in improving prediction accuracy. SVMs are effective with high-dimensional clinical data, while Neural Networks in genetic and molecular analysis. Despite these advancements, limitations such as dataset diversity, sample size, and evaluation standardization persist, emphasizing the need for further research. AI integration in recurrence prediction offers promising prospects for personalized care but requires rigorous validation for safe clinical application.

Keywords Breast cancer recurrence prediction · Breast cancer survival analysis model · Deep learning in breast cancer · Machine learning in breast cancer

1 Introduction

Breast cancer, challenging both global and local health spectrum, remains a leading cause of mortality among women despite recent declines in mortality rates. In addition to the health implications, breast cancer imposes a decreased quality of life for patients despite a considerable socioeconomic burden, evidenced by lost work hours and high health care costs. [1].

Facing the complexity of breast cancer recurrence, about 30% of early-stage patients re-experience the disease within a decade, often within the first five years after initial diagnosis. [2]. Accurate prediction of this recurrence is hampered by conventional methods, highlighting the need for technological and methodological innovations.

In this sense, Artificial Intelligence (AI) techniques emerge as a promise, driven by significant advances in the collection and processing of medical data [3–5]. These techniques, which include Machine Learning and Convolutional Neural

---

Alexandre Ray da Silva and Mariana Zuliani Theodoro de Lima have contributed equally to this work.

✉ Mariana Zuliani Theodoro de Lima, mariana.lima@mackenzie.br | [1]OncoAI, Oncologia Inteligência Artificial, Cel Jose Eusebio, 95, Sao Paulo, Sao Paulo 01239–030, Brazil. [2]Engineering School, Mackenzie Presbyterian University, Consolacao street, 930, Sao Paulo, Sao Paulo 01302–907, Brazil.

Discover Oncology (2025) 16:135 | https://doi.org/10.1007/s12672-025-01908-6

Networks, not only improve breast cancer diagnosis and prognosis [6, 7], but they also promote treatment personalization and efficiency in patient care.

Artificial intelligence techniques are emerging to solve medical questions such as diagnosis, prognosis, drug design and testing [3–5] in different specialties. Specifically in breast cancer, artificial intelligence techniques have been used for diagnosis [6] and prognosis [7] of breast cancer, offering the opportunity for personalized care, improved therapy according to response rates, reduced adverse effects and decreased costs of unnecessary treatment. However, the adoption of AI in clinical practice is accompanied by ethical and practical challenges, including issues of privacy, data quality and model interpretability [8]. Furthermore, the results of this review have the potential to influence public health policies and screening strategies by highlighting the practical impact and relevance of the research [9].

When considering patients' perspectives, it is critical to recognize how these technologies can improve quality of life and experience during treatment [10]. Finally, the review will address current limitations in the field of AI in predicting breast cancer recurrence, establishing the need for continued research and innovation [11].

Given the quickly advancement of AI techniques, the present systematic review aims to provide an updated analysis of their application in predicting breast cancer recurrence. This work differentiates itself from previous reviews by focusing on the most recent developments and innovations [12], seeking to evaluate the effectiveness of the techniques used, the training and testing methodologies, and the evaluation metrics. The goal is to significantly contribute to the understanding of this evolving field and its impact on improving clinical decisions and patient outcomes.

2 Methodology

This systematic literature review was conducted to identify, analyze, and interpret available evidence pertinent to a specific research question. The methodology adopted strictly followed the criteria established by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [13]. This approach involved the clear definition of eligibility criteria, thus establishing the objectives of the review, and the inclusion and exclusion criteria for studies. Furthermore, this segment details the databases consulted, the search strategy implemented and the sequential stages of study selection, ensuring a systematized and replicable approach.

2.1 Eligibility criteria

In this systematic review, eligibility criteria were established to ensure the inclusion of relevant and high-quality studies. The choice of the time range from 2003 to May 2023 for the selection of studies was based on key considerations. The year 2003 represents a milestone in the advancement of Artificial Intelligence, marking the beginning of significant developments in machine learning techniques and the increasing availability of large data sets. This period also coincides with the increased digitization of medical records and images, expanding the practical applications of AI in medicine. The inclusion of studies through May 2023 ensures analysis of the latest discoveries and innovations in the application of AI for the classification and predictive analysis of breast cancer recurrence, offering a complete view of historical and contemporary advances in the field.

The following were included: (1) studies that explore the use of Artificial Intelligence approaches in the classification and predictive analysis of breast cancer recurrence, using both medical images (such as chest x-rays, chest computed tomography scans, or ultrasounds) and clinical and laboratories; and (2) studies that validate or evaluate methodologies and models for classifying recurrence in breast cancer. Furthermore, inclusion was restricted to works written in English, in scientific format, and available in full.

In the process of this systematic review, exclusion criteria were applied to ensure the relevance and focus of the studies analyzed. The following were excluded: (1) studies that address diseases other than breast cancer recurrence; (2) research using data types other than medical images or clinical information, such as meteorological or regional data; (3) studies whose main focus is not breast cancer recurrence; and (4) types of publications such as reviews, letters, conference summaries, book chapters or studies that were limited to describing classification models without a detailed evaluation of them.

Additionally, other exclusion criteria were applied to ensure the methodological robustness of the included studies: (5) studies that did not describe in detail and transparently the Artificial Intelligence method used, including data treatment and analysis, were excluded; and (6) studies that failed to provide an adequate and detailed description of the database used. This requirement ensures that only research with clear and replicable methodologies, essential for validating and understanding results in Artificial Intelligence, were considered in this review.

2.2 Database and search strategy

To identify relevant studies, we developed an individual search strategy for each of the following electronic databases: Scopus, Embase, PubMed and Web of Science (Clarivate Analytics). The choice of these bases was strategic, aiming to cover a wide spectrum of scientific literature. Each offers a unique contribution: PubMed is essential for health research, Embase offers broader coverage in biomedicine, Scopus stands out for its multidisciplinary coverage, and Web of Science is known for its diversity of disciplines and high-impact studies. This selection ensures a balanced and comprehensive approach.

The search strategy was meticulously planned using specific keywords and their respective synonyms. The terms chosen included: (i.) artificial intelligence, (ii.) machine learning, (iii.) deep learning, (iv.) breast cancer, (v.) recurrence and (vi.) recurrence free survival, combined with Boolean operators, as detailed in the Table 1. This search string was applied to the Title/Abstract fields to focus on studies directly related to the objectives of this review, optimizing the relevance and quality of the identified publications.

Duplicate references were carefully removed using Mendeley® software. This crucial step avoids multiple counting of the same data and ensures the validity of the review results. Furthermore, all searches in the electronic databases were carried out on a specific date, May 2, 2023. This uniformity in the search date ensures consistency in data collection and clearly delimits the temporal scope of the review.

2.3 Selection steps

The selection of studies for this systematic review was carried out meticulously in two phases. Initially, in the First Phase, three researchers independently reviewed the titles and abstracts of studies identified in the electronic databases, each responsible for a portion of the articles. They selected those who appeared to meet the inclusion criteria. In the Second Phase, the same researchers read the pre-selected articles in full. At this stage, studies that did not fully meet the inclusion criteria were excluded. Any divergences found in the two phases were discussed and resolved through consensus between all researchers involved. When necessary, a third researcher was consulted to reach an impartial final decision.

The initial search identified 459 articles. After removing duplicates with Mendeley®, 168 unique articles remained. Review of titles and abstracts reduced this number to 88. A detailed assessment and thorough reading by a second group of reviewers resulted in the final selection of 62 articles that met the objectives of this review. The Fig. 1, illustrates the PRISMA application flowchart [13] for each stage of this selection process, from initial identification to the final choice of articles, ensuring transparency and clarity of the procedure adopted.

The flowchart outlines the study search and selection process conducted in this systematic review, following the PRISMA model. Initially, 459 studies were identified across four databases (Embase, PubMed, Scopus, and Web of Science). After removing duplicates, 168 unique studies remained. During the screening stage, 21 studies were excluded based on their titles, and 59 were discarded after abstract analysis, leaving 88 studies selected. No additional studies were included from reference lists. Subsequently, the 88 studies were assessed for eligibility, of which 26 were excluded for not meeting the inclusion criteria. Finally, 62 studies were deemed eligible and included in the systematic review.

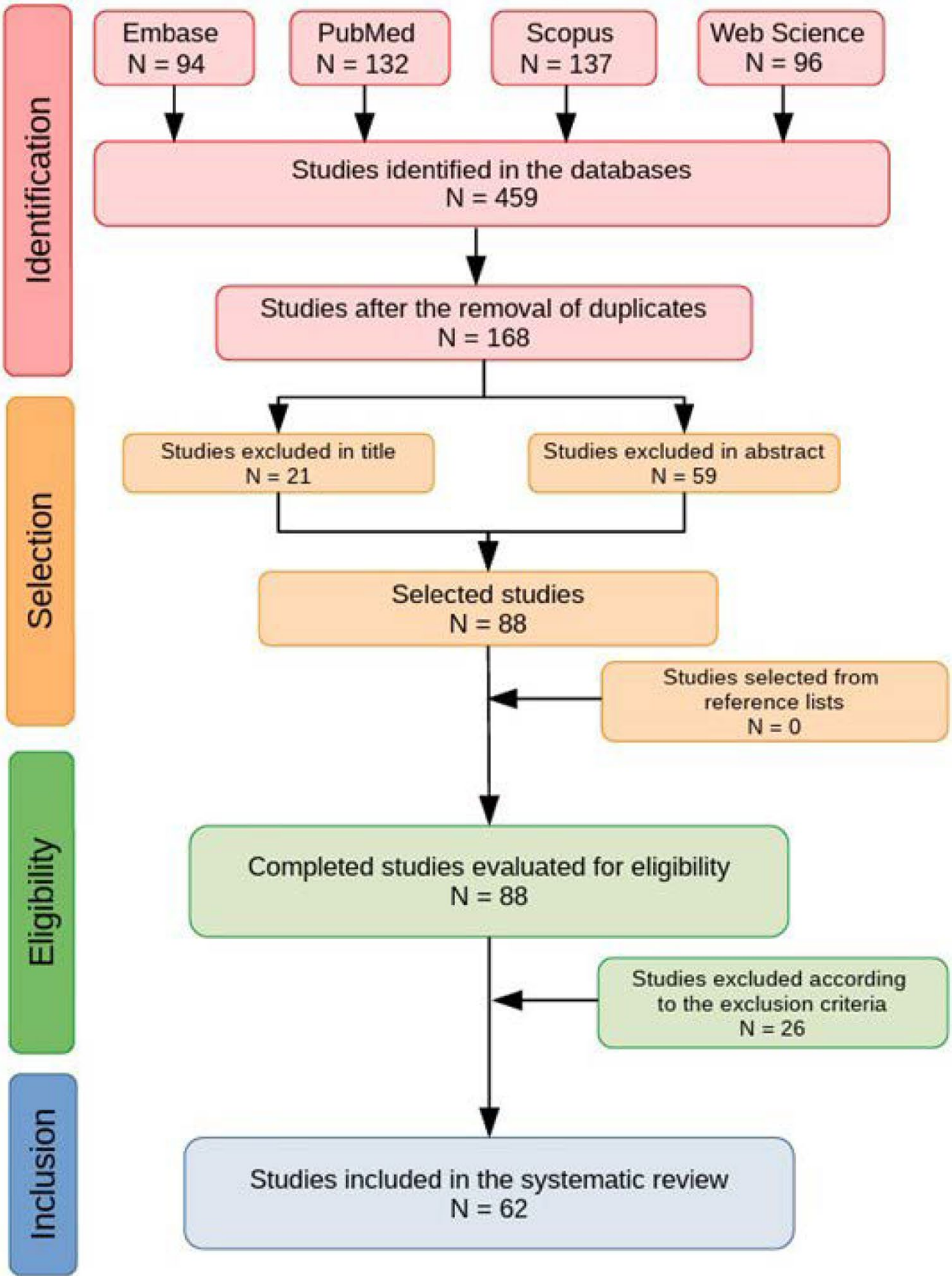
2.4 Study characteristics

In this section, we perform a detailed analysis of the studies selected for this systematic review. The Table 2 was designed to facilitate the understanding and comparison of these works. This table includes columns that describe the bibliographic references (Ref.), the nature of the data used (Data type), the Artificial Intelligence methods used (AI method), the performance of the algorithms (Perform), and information about the databases used (Base information).

Table 1 Search string applied to selected databases

Search string
("artificial intelligence" OR "machine learning" OR "deep learning")
AND ("recurrence" OR "recurrence free survival") AND "breast cancer"

Fig. 1 Flowchart illustrating the study search and selection process, adapted from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [13]. It details each stage of the selection process, from initial identification to the final inclusion of articles



The organization of this data provides a clear and comparative view of the different approaches and results in the use of AI in the classification of breast cancer recurrence.

The analysis of the studies presented in Table 2 indicates a prevalence of Machine Learning techniques. However, one can note the growing study of Deep Learning (DL) techniques, with emphasis on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs, a fundamental class in DL, are particularly effective at analyzing visual data, learning to identify and extract important features and correlating them with desired results [73, 74]. On the other hand, RNNs stand out in managing data with temporal dependencies, demonstrating their ability to process sequences of information [75–77]. Advances in DL have brought innovations to RNNs, improving their ability to handle longer and more complex data sequences, as evidenced in recent literature [78, 79].

Another critical aspect in AI studies for cancer recurrence classification is the size and diversity of the datasets. Figure 2, a histogram, illustrates the distribution of the sizes of the data sets used. Using the histogram, it is possible to identify patterns in the volume of data used, providing insights into the necessary amount of data for the effectiveness of AI models. Furthermore, the diversity of data, fundamental for the generalization of results, is visually

Table 2 Summary of an analysis of the selected studies

Ref	Data type	AI method	Perform	Base information
[14]	Clinical	SVM	82.0 (Q^2)	117 patients
[15]	Clinical	Naive Bayes	70.0 (A)	1.035 patients
[16]	Clinical	SVM	85.0 (A)	733 patients
[17]	Clinical	SVM	89.9 (S)	1.466 patients
[18]	Clinical	BDT	90.0 (A)	1.085 patients
[19]	Clinical	Naive Bayes	81.0 (AUC)	679 patients
[20]	Clinical	Logistic Regression	71.0 (AUC)	49.538 patients
[21]	Clinical	Random Forest	97.0 (S)	328 patients
[22]	Clinical	BDT	78.5 (AUC)	776 patients
[23]	Clinical	SVM	93.3 (AUC)	701 patients
[24]	Clinical	SVM	90.1 (S)	298 patients
[25]	Image	Random Forest	91 (S)	159 patients
[26]	Clinical	Ensenble of Classifiers	88.0 (AUC)	114 patients
[27]	Clinical	Random Forest	91.7 (AUC)	301 patients
[28]	Image	XGBoost	94.0 (AUC)	38 patients
[29]	Clinical	Random Forest	74.6 (AUC)	302 patients
[30]	Clinical/Image	SVM and CNN	95.0 (AUC)	6.899 patients
[31]	Clinical	kNN	67.1 (A)	749 patients
[32]	Clinical	SVM and ANN	95.2 (S)	1.021 patients
[33]	Clinical	XGBoost	91.8 (A)	217 patients
[34]	Clinical	CNN	88.0 (AUC)	6.899 patients
[35]	Clinical	ANN	97.62 (AUC)	1.140 patients
[36]	Clinical	XGBoost	75.0 (AUC)	2.000 samples
[37]	Clinical	MLP	75.0 (AUC)	320 patients
[38]	Clinical/Image	SVM	95.0 (S)	80 patients
[39]	Image	SVM	82.0 (A)	83 patients
[40]	Image	X Cept	91.0 (A)	138 patients
[41]	Image	CNN	84.9 (Recall)	101 patients
[42]	Clinical	SVM	81.0 (S)	256 patients
[43]	Image	CNN	75.1 (AUC)	1.343 patients
[44]	Clinical	AdaBoost	88.9 (AUC)	83 patients
[45]	Clinical	RNN	90.0 (AUC)	13.117 patients
[39]	Clinical	SVM	87.0 (S)	13.117 patients
[46]	Clinical/Image	Random Forest	94.0 (AUC)	130 patients
[47]	Clinical	Fuzzy Logic	≈ 92.4 (S)	695 samples
[48]	Clinical	Bagging	92.25 (A)	1.475 patients
[49]	Clinical	LSTM	94.0 (AUC)	894 patients
[50]	Image	XGBoost	75.0 (AUC)	127 patients
[51]	Image	ResNet34	80.0 (S)	706 patients
[52]	Clinical/Image	ResNet18	75.0 (AUC)	196 patients
[53]	Clinical/Image	ResNet18	87.0 (S)	1.727 patients
[54]	Clinical	SVM	81.94 (A)	484 patients
[55]	Clinical	SVM	90.0 (A)	511 patients
[56]	Clinical	DNN	80.0 (S)	677 patients
[57]	Clinical	XGBoost	77.7 (A)	529 patients
[55]	Clinical	SVM and kNN	91.1 (A)	511 patients
[58]	Clinical	Random Forest	91.0 (S)	511 patients
[59]	Clinical	Logistic Regression	65.0 (AUC)	2.251 patients
[60]	Clinical	Random Forest	96.3 (A)	3.736 patients
[44]	Clinical/Image	CNN	72.0 (AUC)	127 patients
[61]	Clinical/Image	Random Forest	93.3 (AUC)	147 patients

Table 2 (continued)

Ref	Data type	AI method	Perform	Base information
[62]	Clinical	Random Forest	81.46 (AUC)	6.486 patients
[63]	Clinical	SVM	96.5 (S)	569 patients
[64]	Clinical	SVM	79.62 (S)	286 patients
[65]	Clinical/Image	SVM	85.0 (A)	83 patients
[66]	Image	ResNet50	81.1 (AUC)	151 patients
[67]	Clinical	XGBoost	93.7 (AUC)	486 patients
[68]	Clinical	Naive Bayes	92.31 (S)	286 patients
[69]	Clinical	LSTM	89.0 (A)	1.841 patients
[70]	Image	Logistic Regression	66.0 (AUC)	248 patients
[71]	Image	Linear Regression	76.0 (AUC)	62 patients
[72]	Clinical	Random Forest	86.0 (AUC)	79.483 patients

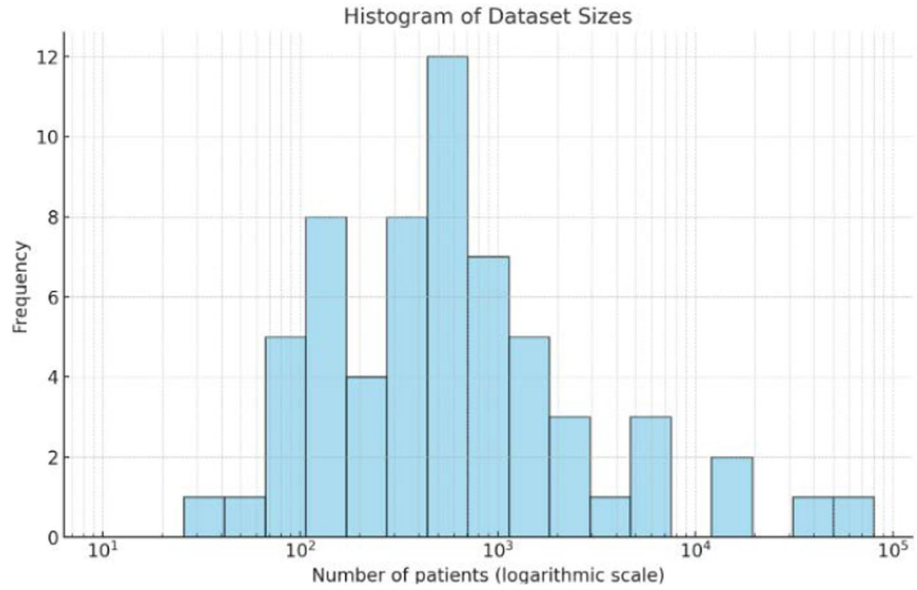


Fig. 2 Figure illustrating the distribution of dataset sizes used in the studies selected for this review

emphasized, highlighting the influence of the size and variety of data sets on the performance of AI models in medicine.

Therefore, this section offers a comprehensive overview of the methods and data used in AI studies applied to breast cancer recurrence, highlighting how these factors contribute to the effectiveness and applicability of these techniques.

3 Results and technical explanation

The systematic review conducted in this study aimed to investigate the application of AI techniques in predicting breast cancer recurrence, providing a detailed analysis of clinical and imaging data, as well as the combination of these two types of data. Below are the main insights obtained AI algorithms applied to clinical data showed varied results depending on the model used. SVM emerged as one of the most effective methods, with an average performance of around 87.03%. This performance is attributed to SVM's ability to handle high-dimensional spaces and provide robust classifications, making it a solid choice for predictive models with complex clinical data. Other algorithms like Naive Bayes and Decision Trees showed moderate performance, indicating that while useful, they may lack the sophistication needed to handle the inherent complexity of clinical data in all cases.

Studies utilizing artificial neural networks, particularly the MLP and RNN, demonstrated superior performance, especially in genetic and molecular datasets. A high average performance of 97.62% was observed in studies applying these techniques, highlighting the effectiveness of neural networks in capturing complex patterns in high-dimensional data.

The application of AI to imaging data, such as MRI and QUS, showed that CNNs are the most suitable models, with an average performance of around 80.0%. CNNs in recognizing spatial patterns in images, which is crucial for analyzing medical visual data. However, it was observed that the complexity of medical images often requires hybrid or specialized approaches to achieve optimal performance. In addition to CNNs, the ResNet architecture, particularly ResNet-34, was frequently used and showed a good balance between computational complexity and performance in computer vision tasks, being particularly useful in contexts where computational resources are limited. Combined data, integrating clinical records with genetic or imaging information, resulted in a significant improvement in the performance of predictive models. Approaches that combined SVM with ANN or SVM with CNNs achieved an average performance between 95.0% and 95.2%. These results indicate that combining different data types and AI techniques can offer a more holistic and accurate view, improving the robustness and accuracy of models.

A critical factor identified was the impact of the size and diversity of datasets on the performance of AI models. Studies employing large volumes of data, as well as a diversity of data sources, showed better performance overall. This suggests that to maximize the effectiveness of AI models in predicting breast cancer recurrence, the use of broad and diverse datasets is essential.

The results of this review reveal that the choice of AI algorithm and the nature of the data used are key factors in predicting breast cancer recurrence. SVM-based methods and Neural Networks demonstrated high effectiveness, especially when applied to combined data. This study highlights the importance of a personalized approach in choosing AI techniques, considering the specific characteristics of the available data.

Furthermore, the review suggests that despite the advancement and effectiveness of AI models, there is significant room for improvement, especially concerning the integration of multimodal data and overcoming challenges related to the complexity of medical imaging data. For the future, continued development of hybrid approaches and studies exploring the clinical applicability of these techniques in real healthcare settings are recommended.

The adoption of AI practices in predicting breast cancer recurrence has the potential to significantly transform clinical care, offering more accurate and personalized diagnoses. However, it is crucial that these technologies be accompanied by rigorous and ongoing validation to ensure their safety, efficacy, and acceptance in clinical practice.

4 Clinical data articles

Within the set of studies selected for this systematic review, forty-three focus specifically on the use of clinical data for the classification of breast cancer recurrence. For this analysis, "clinical data" refers to patients' laboratory information, details of prescriptions and treatments, and records of disease progression. This data provides information regarding patient conditions and the progression of breast cancer, acting as a fundamental basis for the development and improvement of AI models. Figure 3 illustrates the words most frequently used to describe features in articles, providing an overview of the most common clinical variables and their relevance in the studies analyzed.

As illustrated in Fig. 3, The most frequently used keywords in clinical data studies include age, tumor grade, lymph node, estrogen receptor, size, node, invasion, histological and ER. These characteristics are often chosen because of their

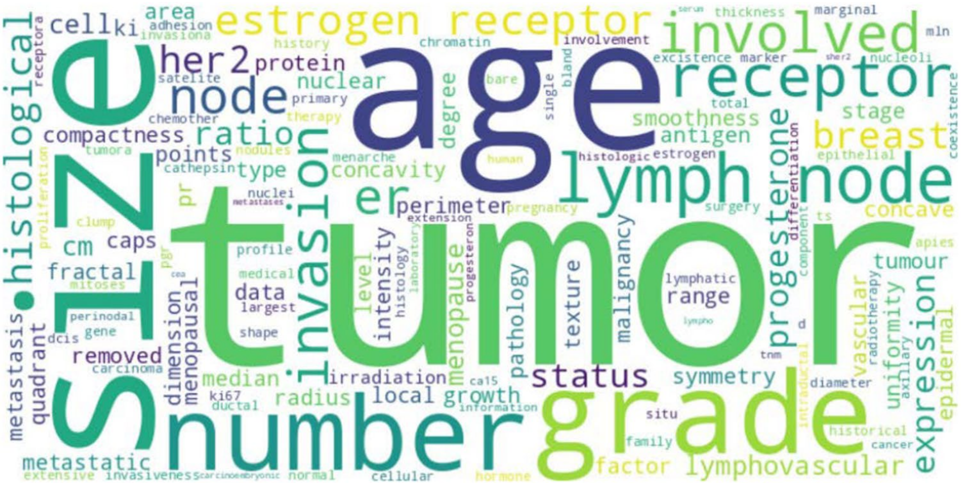


Fig. 3 Most frequently used terms to describe the clinical features in the selected articles that relied solely on clinical data

strong cor- relation with breast cancer recurrence. Their selection in studies is commonly guided by feature selection algorithms, which identify variables with statistical significance or clinical relevance for the predictive outcome [15, 18–20]. This approach allows researchers to focus on factors that have the greatest potential to influence the effectiveness of AI models in predicting recurrence, ensuring that the models are both accurate and clinically relevant.

After identifying the most relevant features for breast cancer recurrence, studies advanced in the application of various machine learning algorithms to develop predictive models. This step in modeling involves selecting specific techniques that best fit the nature of the data and forecasting objectives. To facilitate the discussion about algorithms, we propose a categorization based on their characteristics and methodological approaches. The defined categories reflect the different strategies and potential of each type of algorithm in addressing the problem of breast cancer recurrence, namely:

- Linear Algorithms: This category includes algorithms that establish a linear relationship between the independent variables and the dependent variable, characterized by modeling relationships using linear functions, which makes them widely adopted in various analytical contexts. The preference for linear algorithms is due to their simplicity, computational efficiency and, above all, the ease of interpreting the results. An example used in the studies analyzed is Logistic Regression [20], which stands out for its ability to deal with the classification of binary data, as is common in predicting events such as the recurrence of breast cancer.
- Tree-Based Algorithms: This category uses tree structures to model a series of decisions and their possible implications, including uncertainty and cost variables. They are effective in classification and regression tasks and stand out for their ability to identify non-linear patterns in data. Tree-based algorithms are recognized for their robustness and flexibility, capable of dealing with a wide range of clinical data in complex contexts, such as breast cancer recurrence. Examples found in studies include Decision Tree [22], which offers a simple and intuitive model; Random Forest [21, 29, 58, 60, 62, 72], known for its high accuracy and ability to handle large data sets; XGBoost [33, 36, 57, 67], efficient for optimizing the performance of machine learning models; and AdaBoost [44], a boosting algorithm that improves the accuracy of prediction models.
- Algorithms Based on Support Vector Machines (SVMs): These algorithms are effective in identifying the optimal hyperplane that distinctly separates classes of data in a high-dimensional space. They are particularly useful in scenarios where data is not linearly separable. In these cases, SVMs employ kernel functions, such as polynomial or Radial Basis Function (RBF), to transform the input space, thus enabling effective linear separation of classes. In the studies selected for this review, SVMs were used, demonstrating their relevance and effective- ness in modeling the complexity of clinical data associated with breast cancer recurrence [14, 16, 23, 24, 39, 42, 54, 55, 63, 64].
- Neighborhood-Based Algorithms: These algorithms are characterized by making decisions based on the proximity or similarity between data instances. They are particularly effective in tasks where patterns and relationships can be discerned by analyzing the neighborhood of data. K-Nearest Neighbors (KNN) is one of the most prominent examples of this category, establishing its predictions considering the k nearest neighbors of an instance, as observed in breast cancer recurrence studies [31].
- Probabilistic Algorithms: This category of algorithms applies probabilistic concepts to deal with uncertainties in scenarios where data may be uncertain or incomplete, a common situation in clinical data. Algorithms such as Naive Bayes and Gaussian Naive Bayes [15, 19, 68], exemplify this approach well. These algorithms use simplified assumptions about the independence between features to calculate probabilities, facilitating modeling and prediction in complex medical data contexts. Its use in the analysis of breast cancer recurrence demonstrates the feasibility of applying probabilistic methods to interpret patterns and trends in data where absolute certainty is not always possible.
- Artificial Neural Networks: These computational models are designed to estimate or approximate complex functions that may depend on a large number of inputs and are generally unknown. Each type of neural network has unique characteristics, making them suitable for different types of problems and data sets. For example, the Multi- layer Perceptron Neural Network (MLPNN) is used for general supervised learning tasks; Recurrent Neural Networks (RNNs) are excellent for sequential data; Long Short-Term Memory Networks (LSTMs) are a specialization of RNNs, designed to handle long-term dependencies; and Deep Neural Networks (DNNs) are employed to learn complex features in large datasets. In the studies analyzed [35, 37, 45, 49, 56], these varieties of neural networks were applied, reflecting the diversity and potential application of these models in predicting breast cancer recurrence.
- Special Approaches: In addition to the more common techniques, some studies in the systematic review employed specialized methods, focusing on specific aspects of breast cancer recurrence classification. Among these, the use of the Multi-objective Multi-Classifier stands out, where multiple models are trained, each focused on a different objective, improving the accuracy and specialization of the results. [26]. Another example is Two-class Decision Jungle,

a variation of decision tree algorithms, adapted for binary classification and compression in directed acyclic graph structures. [27]. Cox model was also applied [35], a statistical technique for survival analysis, investigating the relationship between survival and predictor variables. Finally, eTumorMetastasis [47] represents a breakthrough in the use of whole exome sequencing data, transforming functional mutations into network profiles to identify critical genetic signatures, demonstrating its effectiveness in distinguishing between recurrent and non-recurrent tumors.

Figure 4 offers an overview of the distribution of selected articles, categorized according to the machine learning algorithms they employed. This visual overview provides a clear understanding of which approaches are most prevalent in studies and highlights trends in the use of specific algorithms in the area of breast cancer recurrence prediction.

As illustrated in Fig. 4, the XGBoost, Random Forest, and SVM algorithms stand out as the most frequently utilized in the analyzed studies. Among these, SVM has emerged as the preferred algorithm in the majority of cases, primarily due to its ability to achieve effective generalization, as demonstrated in several reviewed studies [14, 16, 23, 24]. Moreover, studies employing SVM consistently report superior performance when compared to other classifiers evaluated [42, 54, 55, 63].

Similarly, the XGBoost algorithm is widely favored for its ability to deliver high efficiency and shorter processing times, as highlighted in multiple studies [33, 36, 57, 67]. Random Forest also remains a popular choice, renowned for its robust and consistent performance across various datasets, often outperforming other classifiers in comparative evaluations [60, 62, 72].

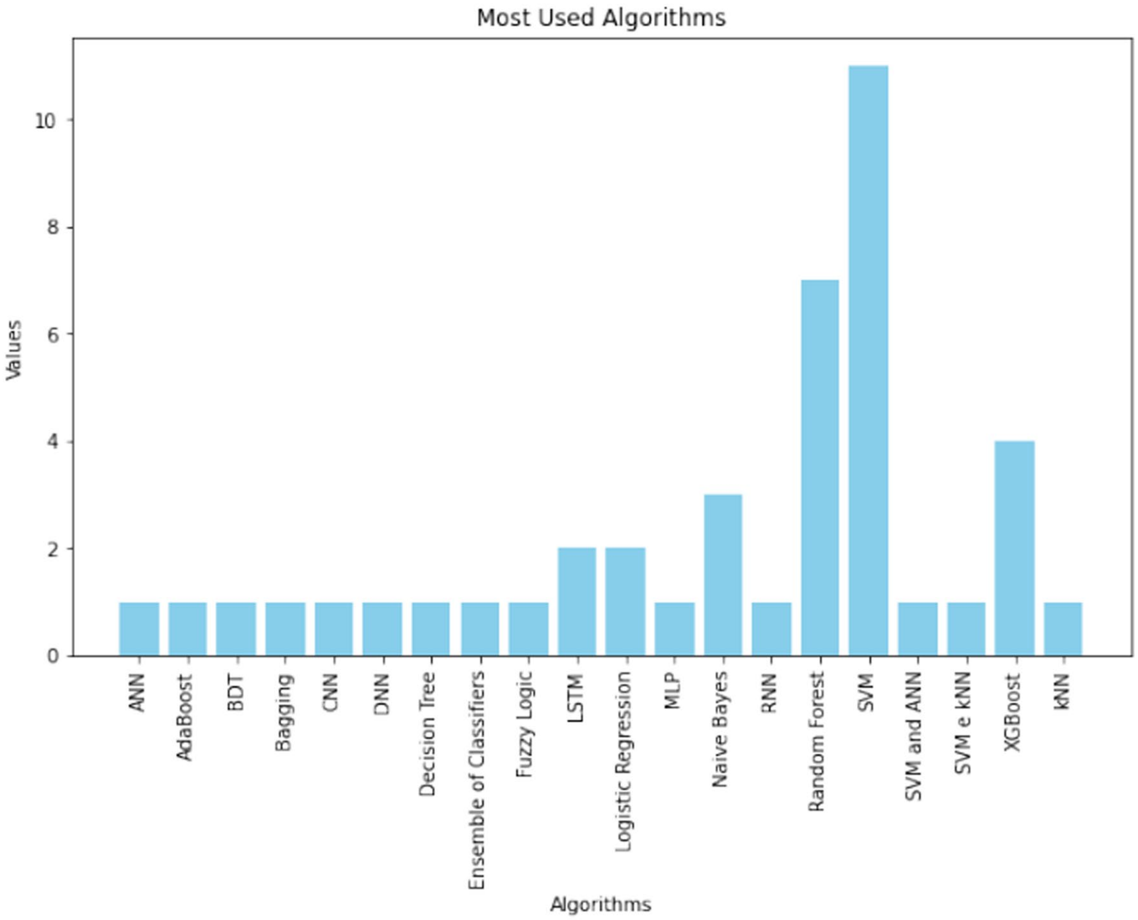


Fig. 4 Algorithms used for predicting recurrence in the articles selected from the review that utilized clinical features

5 Image articles

In the context of imaging data, eleven articles utilized various types of images. The types of imaging data employed include Digitized Whole Slide Images (WSI) [40, 46, 47, 80–82], Multiparametric Magnetic Resonance Imaging (mpMRI), which holds the potential to enhance prediction accuracy [30], Magnetic Resonance Imaging (MRI) [83, 84], and Quantitative Ultrasound (QUS) [39], which captures raw radiofrequency data that can provide a more detailed characterization of tissue microstructure.

The authors of these studies applied a range of machine learning algorithms to develop predictive models for breast cancer recurrence. To provide a clear overview of the algorithms used in the selected works, they were grouped into the following categories:

Convolutional Neural Network (CNN): This class of deep neural networks is specifically designed to recognize spatial patterns within data, making it particularly effective for tasks such as image and video recognition, image classification, object detection, and semantic segmentation. The basic structure of a CNN includes convolutional lay- ers, which are central to the network's operation, activation functions that introduce non-linearity to the model, pooling layers that progressively reduce the spatial dimensions of feature maps, and dense layers where every neuron is connected to every neuron in the subsequent layer, commonly used for classification tasks. Additionally, a flattening layer is employed to convert 2D feature maps into a 1D vector, enabling them to be processed by the dense layers. This approach is seen in this study [82].

ResNet-34: This variation of the ResNet architecture is a convolutional neural network designed for image recognition and other computer vision tasks. ResNet introduced a breakthrough solution to the vanishing gradient problem, allowing for much deeper networks than previously possible. The"34″ in its name refers to its 34 convolutional layers. ResNet-34 utilizes residual blocks, which help mitigate the vanishing gradient problem. While not the deepest ResNet variant, it strikes a commend- able balance between computational complexity and performance in computer vision tasks. This architecture is particularly advantageous when computational resources are limited or when fast processing is necessary, such as in real-time applications or resource-constrained environments. This approach is employed in study [47].

Other algorithms applied to imaging data have been discussed in subsection 3.1, which covers Clinical Data.

Figure 5 highlights that CNN and XGBoost are the most frequently used algorithms across the analyzed articles. CNN, in particular, is often selected for its ability to achieve high performance, as demonstrated in previous studies [82]. However, it is crucial to emphasize that an algorithm's performance is not solely dependent on its inherent capabilities. Factors such as data collection methods, the quality and type of images used, and other contextual elements play a significant role in influencing the overall outcome.

6 Clinical data articles and images

In the context of studies that involve both clinical and imaging data, nine articles utilized a combination of these data types. Figure 6 highlights the key terms that characterize the features used in the models described in these studies. Variables such as age, estrogen receptor status, progesterone receptor status, tumor stage and grade, Ki67, and histological type are identified as the most important and frequently used features in studies involving both clinical and imaging data.

Regarding imaging features, the characteristics extracted depend on the type of imaging modality used, resulting in limited overlap of features across different studies, as shown in Fig. 7.

The types of images employed in these studies include Digitized Whole Slide Images (WSI) [40, 46, 47, 82, 85], Multiparametric Magnetic Resonance Imaging (mpMRI), which has demonstrated potential for improving predictive accuracy [30], Magnetic Resonance Imaging (MRI) [58], and Quantitative Ultrasound (QUS) [39], which captures raw radiofrequency data that can provide better characterization of tissue microstructure. Various machine learning algorithms were employed by the authors to develop predictive models for breast cancer recurrence. To better discuss the algorithms used in the selected studies, they were categorized as follows:

ResNet-50: This architecture consists of 50 convolutional layers, beginning with an initial convolutional layer followed by 16 residual blocks that form the backbone of the network. ResNet-50 employs a"bottleneck" design in its

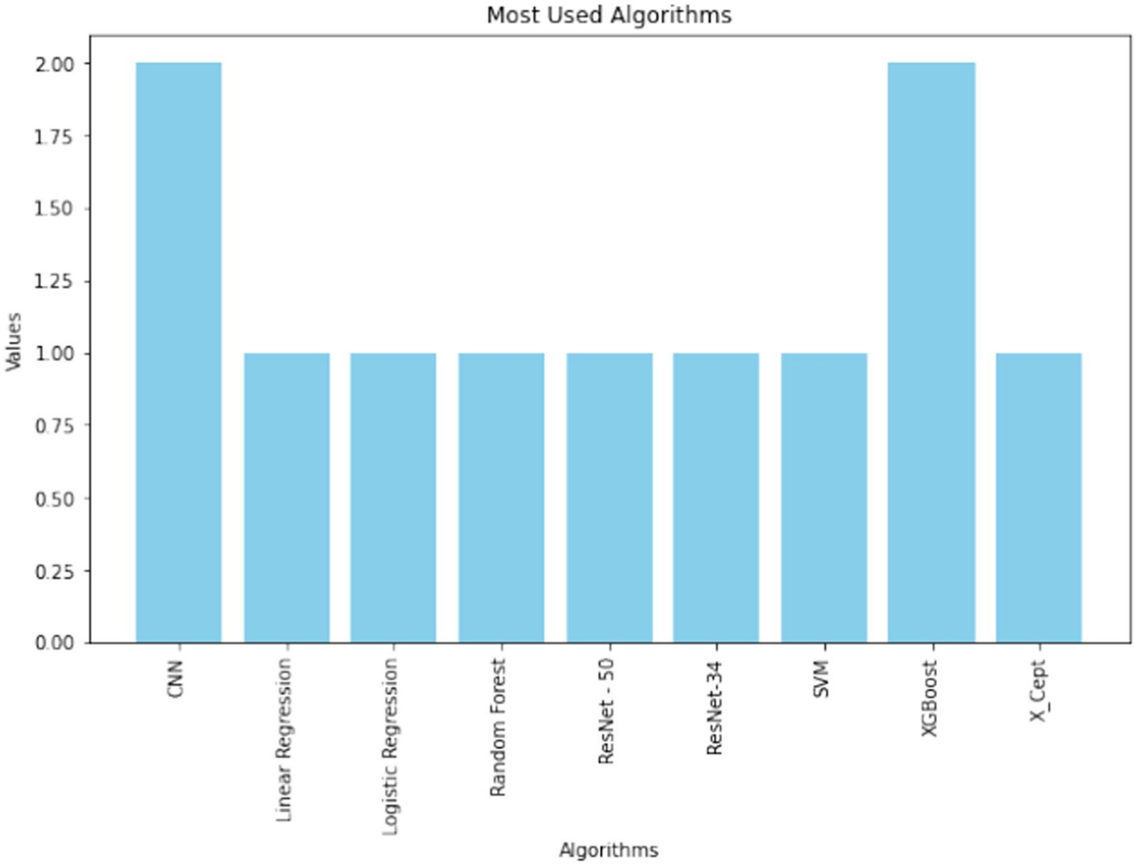
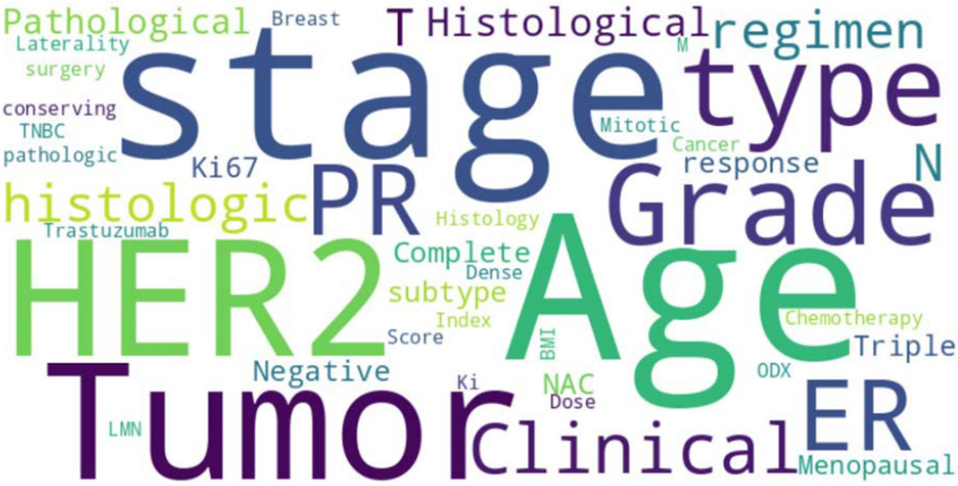


Fig. 5 Algorithms used for predicting recurrence in the articles selected from the review that relied solely on image data

Fig. 6 The most frequently used terms to describe clinical-related features in articles that incorporate both clinical data and images



residual blocks, with each block comprising three layers: a 1 × 1 convolutional layer for dimensionality reduction, a 3 × 3 convolutional layer for spatial processing, and another 1 × 1 convolutional layer to restore the dimensions. This structure helps to reduce computational complexity while maintaining the network's effectiveness.

Figure 8 highlights that the most frequently used algorithms in these studies were Random Forest, ResNet18, and SVM. Notably, approaches that combined SVM with ANN or SVM with CNN achieved superior performance, as detailed in Table 2. It is important to emphasize that a model's performance is not solely determined by the choice of algorithm; other factors, such as data collection methods, image quality, and related variables, also play a critical role in influencing the outcomes.

Fig. 7 The most frequently used terms to describe image-related features in articles that utilize both clinical data and images

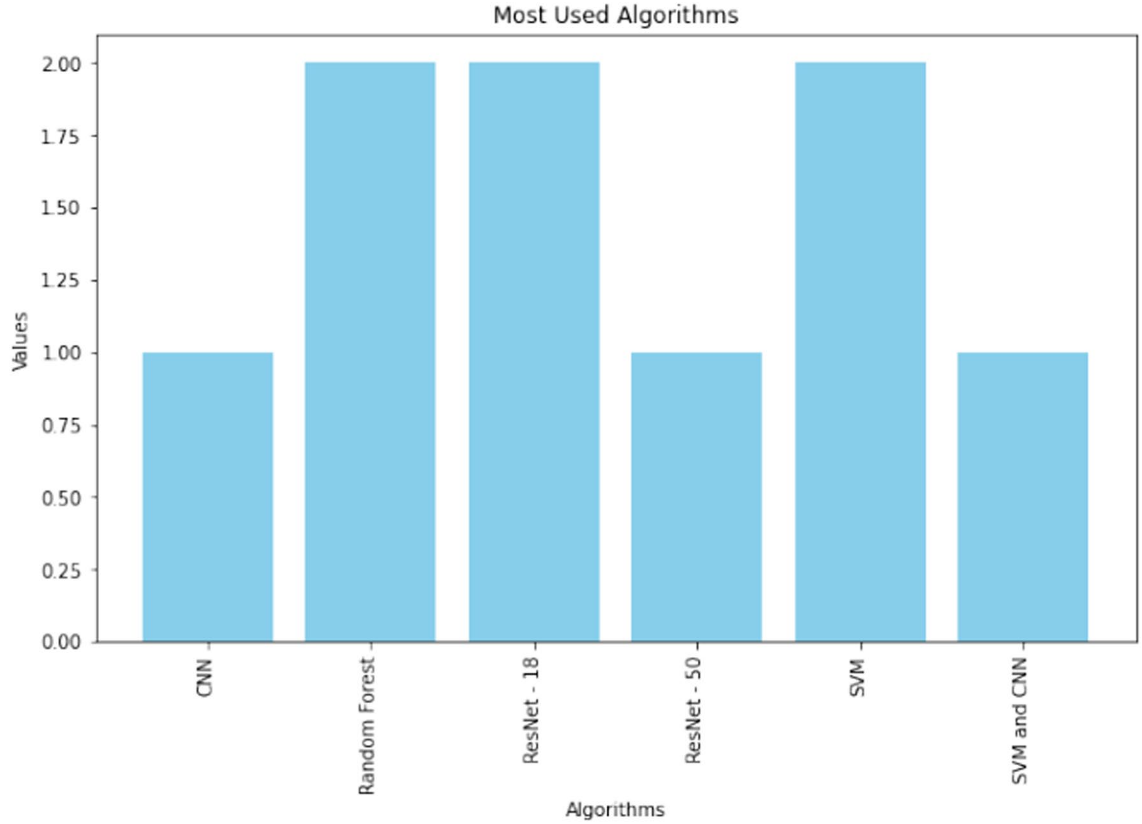


Fig. 8 Algorithms used for predicting recurrence in the articles selected from the review that utilized both clinical data and images

7 Limitations

While this systematic review has offered valuable insights into the use of AI techniques for predicting breast cancer recurrence, several limitations must be acknowledged, as they may influence the interpretation and generalizability of the findings.

A primary limitation identified is the lack of diversity in the datasets used across the analyzed studies. Most studies relied on data from a single institution or a specific database, which may not adequately reflect the diversity of populations and clinical contexts found in different geographical regions. This homogeneity limits the applicability of AI models to broader populations beyond those studied, raising concerns about their generalizability. Another significant limitation is the relatively small sample size in some of the reviewed studies. AI models, particularly those based on deep learning, typically require large datasets to achieve robust and reliable performance. Small sample sizes can lead to overfitting, where the model captures patterns unique to the training data but fails to generalize effectively to new, unseen data.

While a variety of AI algorithms were explored, the review may not have fully encompassed all potential methods and their variations that could be relevant for predicting breast cancer recurrence. Some advanced or novel techniques might have been underrepresented due to the time frame of the review or the exclusion criteria applied to certain studies. Furthermore, the analysis of the included studies may have been affected by confounding factors that were either not considered or insufficiently controlled within the AI models. Clinical variables, such as the specific treatments received, treatment responses, and cancer stage, are critical influences on outcomes but may not have been consistently included or appropriately weighted in the models analyzed.

Despite advancements in AI model development, interpretability remains a significant challenge. Many of the studies reviewed employed complex algorithms, such as deep neural networks, which, while powerful, function as" black boxes"—obscuring the reasoning behind their predictions. This opacity can undermine the confidence of healthcare professionals and hinder the integration of these models into clinical practice. Additionally, the heterogeneity in evaluation metrics and tools used to validate AI models poses another limitation. The lack of standardized validation methods complicates direct comparisons between studies and impedes the accurate assessment of the true effectiveness of different models. Moreover, many studies relied on internal validation processes, which may overstate the model's performance, lacking the rigor of external validation.

These limitations highlight the need for caution when interpreting the findings of this review and when considering the practical implementation of AI models for predicting breast cancer recurrence. To enhance the generalizability and applicability of future results, researchers should prioritize increasing dataset diversity, incorporating larger sample sizes, exploring a broader spectrum of algorithms, and developing methods that improve AI model interpretability. Additionally, it is crucial to adopt standardized evaluation criteria and perform rigorous external validations to ensure the robustness of models across diverse clinical settings.

8 Conclusion

This systematic review explored the emerging role of AI techniques in predicting breast cancer recurrence, analyzing clinical data, imaging data, and combined data. The findings indicate that AI techniques, particularly SVM and Neural Networks, have shown great potential in improving the accuracy and personalization of recurrence prognostics.

The application of AI algorithms to clinical data proved effective, especially with the use of SVMs, which provided consistent results in high-dimensional scenarios. However, the use of Neural Networks, particularly in genetic and molecular data, stood out as a promising approach, capturing complex patterns that other algorithms might not detect.

When applied to imaging data, CNNs and architectures like ResNet excelled, though the complexity of medical images suggests the need for hybrid or specialized approaches to achieve optimal performance. The combination of clinical and imaging data resulted in enhanced performance of predictive models, suggesting that multi-modal approaches can provide a more comprehensive and accurate view for predicting breast cancer recurrence.

Despite these promising developments, several limitations were identified, including the limited diversity of data, small sample sizes, and the lack of standardization in evaluation tools. These limitations highlight the need for future research to address these challenges, aiming to improve the generalization and applicability of AI models in different clinical contexts.

In summary, the adoption of AI techniques in predicting breast cancer recurrence represents a significant advancement, with the potential to transform clinical care by offering more accurate and personalized diagnostics. However, to ensure the safety and effectiveness of these technologies, it is essential that they are accompanied by rigorous and continuous validation, along with careful consideration of the ethical and practical issues associated with their use in clinical settings.

Acknowledgements I would like to thank CNPq and Softex for their financial support, and DeepTech OncoAI for their institutional support in the development of this project.

Author contributions All authors wrote the main manuscript. J.A.S. and M.Z.T.L. reviewed the manuscript. J.A.S. and A.R.S. has written the main manuscript and reviewed the artificial intelligence models and the algorithms applications in the manuscript. J.A.S. prepared the tables and figures. M.Z.T.L. raised funding.

Funding Conselho Nacional de Desenvolvimento Científico e Tecnológico, Nº 54/2022 Process: 409643/2022-0, Nº 54/2022.

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

Declarations

Competing interests The authors declare no competing interests.

References

1. Shachar SS, Muss HB. Internet tools to enhance breast cancer care. NPJ Breast Cancer. 2016;2:16011.
2. Ahmad A. Pathways to breast cancer recurrence. ISRN oncology. 2013;2013:1–16.
3. Hwang EJ, et al. Development and validation of a deep learning-based auto- mated detection algorithm for major thoracic diseases on chest radiographs. JAMA Netw Open. 2019;2: e191095.
4. Geras KJ et al. High-resolution breast cancer screening with multi-view deep convolutional neural networks. 2017.
5. Chilamkurthy S, et al. Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. Lancet. 2018;392:2388–96.
6. Sadoughi F, et al. Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review. Breast Cancer. 2018;10:219–30.
7. Jain A, Jain A, Jain S, Jain L. Artificial intelligence techniques in breast cancer diagnosis and prognosis, vol. 39. Singapore: World Scientific Publishing Co., Pte. Ltd; 2000.
8. Amedior NC. Ethical implications of artificial intelligence in the healthcare sector. Adv Multidiscip Sci Res J. 2023;36:1–12.
9. Merson MH, Black RE, Mills A. International public health: diseases, programs, systems, and policies. 2nd ed. Sudbury: Jones & Bartlett Learning; 2006.
10. Wang B, Asan O, Mansouri M. Patients' perceptions of integrating ai into healthcare: Systems thinking approach, IEEE, 2022; 1–6.
11. Kolla L, Parikh RB. Uses and limitations of artificial intelligence for oncology, IEEE 2024; 1–6.
12. Abreu PH, Santos MS, Abreu MH, Andrade B, Silva DC. Predicting breast cancer recurrence using machine learning techniques. ACM Comput Surv (CSUR). 2016;49:1–40.
13. Moher D, et al. Preferred reporting items for systematic reviews and meta- analyses: the prisma statement. PLoS Med. 2009;6: e1000097.
14. Chiu SH, Chen CC, Lin TH. Using support vector regression to model the correlation between the clinical metastases time and gene expression profile for breast cancer. Artif Intell Med. 2008;44:221–31.
15. Strumbelj E, Bosnic Z, Kononenko I, Zakotnik B, Kuhar CG. Explanation and reliability of prediction models: the case of breast cancer recurrence. Knowl Inf Syst. 2010;24:305–24.
16. Kim W, et al. Development of novel breast cancer recurrence prediction model using support vector machine. J Breast Cancer. 2012;15:230–8.
17. Stoean R, Stoean C. Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selec- tion. Exp Syst Appl Int J. 2013;40:2677–86.
18. Mohebian MR, Marateb HR, Mansourian M, Man˜ana MA, Mokarian F. A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (hpbcr) using optimized ensemble learning. Computat Struct Biotechnol J. 2016;15:75–85.
19. Kim W, Kim KS, Park RW. Nomogram of naive bayesian model for recurrence prediction of breast cancer. Healthcare Inf Res. 2016;22:89–94.
20. Witteveen A, Nane GF, Vliegen IM, Siesling S, IJzerman MJ. Comparison of logistic regression and bayesian networks for risk prediction of breast cancer recurrence. Med Decis Making. 2018;38:822–33.
21. Thakur SS, et al. The use of automated ki67 analysis to predict oncotype dx risk-of-recurrence categories in early-stage breast cancer. PLoS ONE. 2018;13: e0188983.
22. Takada M, et al. Prediction of postoperative disease-free survival and brain metastasis for her2-positive breast cancer patients treated with neoadjuvant chemotherapy plus trastuzumab using a machine learning algorithm. Breast can cer Res Treat. 2018;172:611–8.
23. Zeng Z, et al. Using natural language processing and machine learning to identify breast cancer local recurrence. BMC Bioinform. 2018. https://doi.org/10.1186/s12859-018-2466-x.
24. Ramkumar C, et al. Development of a novel proteomic risk-classifier for prognostication of patients with early-stage hormone receptor-positive breast cancer. Biomarker Insights. 2018. https://doi.org/10.1177/1177271918789100.

25. Klimov S, et al. A whole slide image-based machine learning approach to predict ductal carcinoma in situ (dcis) recurrence risk. Breast Cancer Res BCR. 2019. https://doi.org/10.1186/s13058-019-1165-5.
26. Chen X. et al. A reliable multi-classifier multi-objective model for predicting recurrence in triple negative breast cancer. *Proceedings of the Annual Inter- national Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,* 2019; 2182–2185.
27. Kim I, et al. A predictive model for high/low risk group according to oncotype dx recurrence score using machine learning. Eur J Surg Oncol. 2019;45:134–40.
28. Tahmassebi A, et al. Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients. Invest Radiol. 2019;54:110–7.
29. Tseng YJ, et al. Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. Int J Med Inf. 2019;128:79–86.
30. Zeng Z, et al. Identifying breast cancer distant recurrences from electronic health records using machine learning. J Healthcare Inf Res. 2019;3:283–99.
31. Mac´ıas-Garc´ıa L, et al. Autoencoded dna methylation data to predict breast cancer recurrence: machine learning models and gene-weight significance. Artif Intell Med. 2020;110:101976.
32. Boeri C, et al. Machine learning techniques in breast cancer progno- sis prediction: a primary evaluation. Cancer Med. 2020;9:3234.
33. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast cancer recurrence prediction. Artif IntellMed. 2020;107:101858.
34. Wang H, Li Y, Khan SA, Luo Y. Prediction of breast cancer distant recurrence using natural language processing and knowledge-guided convolutional neural network. Artif Intell Med. 2020;110:101977.
35. Lou SJ, et al. Machine learning algorithms to predict recurrence within 10 years after breast cancer surgery: a prospective cohort study. Cancers. 2020;12:1–15.
36. Ravkin HD, Givton O, Geffen DB, Rubin E. Direct comparison shows that mrna-based diagnostics incorporate information which cannot be learned directly from genomic mutations. BMC Bioinform. 2020. https://doi.org/10.1186/s12859-020-3512-z.
37. Baltres A, et al. Prediction of oncotype dx recurrence score using deep multi-layer perceptrons in estrogen receptor-positive, her2-negative breast cancer. Breast Cancer. 2020;27:1007–16.
38. Jacobs MA, et al. Integrated multiparametric radiomics and informatics system for characterizing breast tumor characteristics with the oncotypedx gene assay. Cancers. 2020;12:1–23.
39. Dasgupta A, et al. Radiomics in predicting recurrence for patients with locally advanced breast cancer using quantitative ultrasound. Oncotarget. 2021;12:2437–48.
40. Phan NN, Hsu C-Y, Huang C-C, Tseng L-M, Chuang EY. Prediction of breast cancer recurrence using a deep convolutional neural network without region-of-interest labeling. Front Oncol. 2021;11:734015.
41. Cheung RA, Kraft AM, Petty HR. Relocation of phosphofructoki- nases within epithelial cells is a novel event preceding breast cancer recurrence that accurately predicts patient outcomes. Am J Physiol Cell Physiol. 2021;321:C654–70.
42. Massafra R, et al. A clinical decision support system for predicting invasive breast cancer recurrence: preliminary results. Front Oncol. 2021. https://doi.org/10.3389/fonc.2021.576007.
43. Cho SY, et al. Deep learning from he slides predicts the clinical benefit from adjuvant chemotherapy in hormone receptor-positive breast cancer patients. Sci Rep. 2021. https://doi.org/10.1038/s41598-021-96855-x.
44. Yang J, et al. Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. Comput Struct Biotechnol J. 2021;20:333–42.
45. Kim JY, et al. Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer cohort in tertiary cancer center reg- istry. Front Oncol. 2021. https://doi.org/10.3389/fonc.2021.596364.
46. Eun NL, Kang D, Son EJ, Youk JH, Kim JA, Gweon HM. Tex- ture analysis using machine learning–based 3-t magnetic resonance imaging for predicting recurrence in breast cancer patients treated with neoadjuvant chemotherapy. Eur Radiol. 2021;31:6916–28.
47. Milanese J-S, et al. Etumormetastasis: a network-based algorithm predicts clin- ical outcomes using whole-exome sequencing data of cancer patients. Genom Proteom Bioinform. 2021;19:973–85.
48. Alzu'bi A, Najadat H, Doulat W, Al-Shari O, Zhou L. Predicting the recurrence of breast cancer using machine learning algorithms. Multimed Tools Appl. 2021;80:13787–800.
49. Sanyal J, Tariq A, Kurian AW, Rubin D, Banerjee I. Weakly super- vised temporal model for prediction of breast cancer distant recurrence. Sci Rep. 2021;11:1–11.
50. Liu X, et al. Predicting breast cancer recurrence and metastasis risk by inte- grating color and texture features of histopathological images and machine learning technologies. Comput Biol Med. 2022;146:105569.
51. Wetstein SC, et al. Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. Sci Rep. 2022;12:1–12.
52. Yao Y, et al. Icsda: a multi-modal deep learning model to predict breast can- cer recurrence and metastasis risk by integrating pathological, clinical and gene expression data. Brief Bioinform. 2022. https://doi.org/10.1093/bib/bbac448.
53. Rabinovici-Cohen S, et al. Multimodal prediction of five-year breast cancer recur- rence in women who receive neoadjuvant chemotherapy. Cancers. 2022;14:3848.
54. Alwohaibi M, Alzaqebah M, Alotaibi NM, Alzahrani AM, Zouch M. A hybrid multi-stage learning technique based on brain storming optimization algo- rithm for breast cancer recurrence prediction. J King Saud Univ Comput Inf Sci. 2022;34:5192–203.
55. Merouane E, Said A, Nour-eddine EF. Prediction of metastatic relapse in breast cancer using machine learning classifiers. Int J Adv Comput Sci Appl. 2022;13:176–81.
56. Azman BM, Hussain SI, Azmi NA, Ghani MZAA, Norlen NID. Prediction of distant recurrence in breast cancer using a deep neural network. *Revista Internacional de M´etodos Num´ericos para C´alculo y Disen˜o en Ingenier´ıa*, 2022; 38
57. Massafra R, et al. A machine learning ensemble approach for 5- and 10-year breast cancer invasive disease event classification. PLoS ONE. 2022;17: e0274691.

58. Ertel M, Sadqui A, Said A, El Faddouli, N.-e. Prediction of the most effective adjuvant therapeutic combinations for breast cancer patients using multinomial classification. *Journal of Theoretical and Applied Information Technology,* 2022; **100**
59. Ni C, Warner JL, Malin BA, Yin Z. Predicting hormonal therapy medica- tion discontinuation for breast cancer patients using structured data in electronic medical records. AMIA Summits Transl Sci Proc. 2022;2022:359.
60. Pawloski KR, et al. Supervised machine learning model to predict oncotype dx risk category in patients over age 50. Breast Cancer Res Treat. 2022;191:423–30.
61. Ma M, et al. Radiomics features based on automatic segmented mri images: Prognostic biomarkers for triple-negative breast cancer treated with neoadjuvant chemotherapy. Eur J Radiol. 2022;146:110095.
62. Xin L, et al. Multicenter study of the clinicopathological features and recurrence risk prediction model of early-stage breast cancer with low-positive human epider- mal growth factor receptor 2 expression in china (chinese society of breast surgery 021). Chin Med J. 2022;135:697–706.
63. Gupta SR. Prediction time of breast cancer tumor recurrence using machine learning. Cancer Treat Res Commun. 2020;32:100602.
64. Mikhailova V, Anbarjafari G. Comparative analysis of classification algo- rithms on the breast cancer recurrence using machine learning. Med Biol Eng Comput. 2022;60:2589–600.
65. Bhardwaj D, et al. Early changes in quantitative ultrasound imaging parameters during neoadjuvant chemotherapy to predict recurrence in patients with locally advanced breast cancer. Cancers. 2022;14:1247.
66. Su Z, et al. Bcr-net: a deep learning framework to predict breast cancer recur- rence from histopathology images. PLoS ONE. 2023;18: e0283562.
67. Massafra R, et al. Analyzing breast cancer invasive disease event classification through explainable artificial intelligence. Front Med. 2023. https://doi.org/10.3389/fmed.2023.1116354.
68. Ganapathy S, Harichandrakumar K, Tamilarasu K, Penumadu P, Nair NS. An auxiliary approach to prediction of binary outcome with bayesian net- work model: exploration with data for recurrence of breast cancer. J Clin Diagn Res. 2023. https://doi.org/10.7860/JCDR/2023/59472.17598.
69. Zeng L, et al. The innovative model based on artificial intelligence algorithms to predict recurrence risk of patients with postoperative breast cancer. Front Oncol. 2023;13:1117420.
70. Romeo V, et al. Mri radiomics and machine learning for the prediction of onco- type dx recurrence score in invasive breast cancer. Cancers. 2023;15:1840.
71. Chiacchiaretta P, et al. Mri-based radiomics approach predicts tumor recurrence in er + /her2—early breast cancer patients. J Digit Imaging. 2023;36:1071–80.
72. Lauritzen AD, Berg T, Jensen MB, Lillholm M, Knoop A. Identify- ing recurrent breast cancer patients in national health registries using machine learning. Acta Oncol. 2023;62:350–7.
73. He K, Gkioxari G, Dollar P, Girshick R. *Mask r-cnn*, 2017. 2961–2969.
74. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86:2278–324.
75. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Learning internal represen- tations by error propagation. In:* Rumelhart DE, McClelland JL (eds). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. MIT Press, Cambridge, MA. 1986; 318–362.
76. Jordan MI. Serial order: A parallel distributed processing approach. Tech. Rep. ICS Report 8604, University of California, San Diego, Institute for Cognitive Science, La Jolla, CA, USA (1986).
77. Arathi CR, et al. Breast cancer recurrence prediction with deep neural network and feature optimization. Automatika. 2024;65:343–60.
78. Peng H, et al. An ai-based approach for modeling the synergy between radiotherapy and immunotherapy. Sci Rep. 2024;14:8250.
79. Bonacin R, Vechi SM, Dametto M, Ruppert GCS. Exploring deep learning techniques in the prediction of cancer relapse using an open brazilian tabular database. Berlin: Springer Nature; 2024. p. 331–8.
80. Palmer NR, et al. Using natural language processing to construct a metastatic breast cancer cohort from linked cancer registry and electronic medical records data. JAMIA Open. 2019;2:528–37.
81. Do TTN, et al. Comparison of the metastasis predictive potential of mrna and long non-coding rna profiling in systemically untreated breast cancer. Cancers. 2021;13:4907.
82. Ertel M, Amali S, Eddine El Faddouli N. Prediction of the most effective adjuvant therapeutic combinations for breast cancer patients using multinomial classification. *Research Square* 2024.
83. Makubate B, Donnan PT, Dewar JA, Thompson AM, McCowan C. Predicting hormonal therapy medication discontinuation for breast cancer patients using structured data in electronic medical records. Breast Cancer Res Treat. 2013;108:1515–24.
84. Ling AY, Kurian AW, Caswell-Jin JL, Sledge GW, Shah NH, Tamang SR. Using natural language processing to construct a metastatic breast cancer cohort from linked cancer registry and electronic medical records data. JAMIA Open. 2019;2:528–34.
85. Ni C, Warner J, Malin B, Yin Z. Predicting hormonal therapy medica- tion discontinuation for breast cancer patients using structured data in electronic medical records. *AMIA. Annual Symposium proceedings. AMIA Symposium* **2022**, 359–368