Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tciv20

Machine-learning methods in detecting breast cancer and related therapeutic issues: a review

Ali Jafari

To cite this article: Ali Jafari (2024) Machine-learning methods in detecting breast cancer and related therapeutic issues: a review, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 12:1, 2299093, DOI: 10.1080/21681163.2023.2299093

To link to this article: https://doi.org/10.1080/21681163.2023.2299093

Published online: 27 Jan 2024.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING: IMAGING & VISUALIZATION
2024, VOL. 12, NO. 1, 2299093
https://doi.org/10.1080/21681163.2023.2299093

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Machine-learning methods in detecting breast cancer and related therapeutic issues: a review

Ali Jafari

Department of Computer Science and Statistics, K.N.Toosi University of Technology, Tehran, Tehran Province, Iran

ABSTRACT
In 2020, the World Health Organization reported that breast cancer resulted in the deaths of 685,000 people worldwide, with 2.3 million women diagnosed with the disease. Breast cancer is the most common cancer globally, with 7.8 million women diagnosed in the past five years. Machine learning (ML) techniques can help identify breast cancer early and define its type by analyzing tumor size. ML models have been used for image classification and cancer prediction, and have been shown to be beneficial for breast cancer diagnosis. The current systematic review aims to highlight the gaps and shortcomings of previous works regarding the use of ML for breast cancer prediction based on image processing. The review updates publications to see the pros and cons of various ML and deep learning (DL) techniques, and can benefit medical practitioners seeking advanced therapies. The previous works mainly benefited from SVM, KNN, and DT in detecting BC; however, other techniques, especially the DL ones, can be useful.

ARTICLE HISTORY
Received 23 May 2023
Accepted 20 December 2023

KEYWORDS
Breast cancer; machine learning; cancer detection; therapeutic methods; artificial intelligence

1. Introduction

In the modern age of algorithms, machine learning (ML) (Chen et al. 2021) and deep learning (DL) (Chen and Jain 2020) tools have revolutionised multiple industries, especially manufacturing, transportation, and government. In recent years, DL has delivered cutting-edge performance in various fields (Kaul et al. 2022), like speech processing, text analytics, and computer vision. The widespread use of ML/DL algorithms across many industries (including social media) makes these technologies indispensable to daily life (Mittal and Hasija 2020). ML/DL algorithms are also beginning to impact healthcare, a historically resistant sector to significant technology disruptions (Latif et al. 2017). For example, the recognition of the internal organs from medical scans, the recognition of interstitial lung illnesses, the detection of lung nodules, the reconstruction of diagnostic images (Yan et al. 2016), and the classification of brain tumours (Anthimopoulos et al. 2016) can be conducted using the ML/DL algorithms have in the medical field (Havaei et al. 2017).

Over recent years, notable advancements have been seen in using ML across diverse sectors and scholarly investigations (Bhardwaj et al. 2017). ML applied to electronic health records (EHRs) may provide useful insights that can be used to streamline hospital operations, enhance patient risk score systems, and anticipate the development of disease (Shickel et al. 2017). ML algorithms have significant value in healthcare because they can effectively analyse vast quantities of daily healthcare data via electronic health records (Usmani and Jaafar 2022). Additionally, using ML techniques may facilitate the identification of optimal pharmaceutical dosages, reducing healthcare expenses for both patients and healthcare providers. ML has the potential to be used not only for dose determination but also for identifying the most suitable drug for individual patients. Statistical models that use the diversity and complexity of EHR-derived data remain comparatively uncommon and present an attractive field of study (Callahan and Shah 2017). A system that entails improving medical services to meet peoples' medical needs falls under the broad healthcare category. Patients, doctors, vendors, health organisations, and IT firms all work to protect and restore patient records in the healthcare industry. Indian healthcare has been one of the world's fastest-growing industries for the past 10 years (Sarwal et al. 2021). ML used in healthcare analyses can diagnose various illnesses, including cancer, diabetes, strokes, and other conditions. Lung cancer, BC, prostate cancer, stomach cancer, and various other kinds of cancer can be diagnosed using ML based on image processing. Each year, 12% of lung cancer diagnoses and 10% of deaths are documented. Similar statistics apply to BC, where 11% of incidences result in 9% of fatalities. This occurs in all cancer types. It is necessary to produce accurate and high-quality data to analyse cancer in the healthcare system. Healthcare must employ data in a world of competition to boost healthcare quality and reduce the costs associated with treatment (Dhillon and Singh 2019).

BC is a type of cancer that forms in the breast when cells grow out of control. BC cells usually create a tumour seen on X-rays or felt as a lump. It is imperative to understand that most breast lumps are benign and not cancerous or malignant (Sharma et al. 2010). Non-cancerous breast tumours grow abnormally but do not spread outside the breast. They are not life-threatening, but some benign breast lumps increase the risk of BC in women (Waks and Winer 2019). It should be

CONTACT Ali Jafari Ali.Jafari@email.kntu.ac.ir Department of Computer Science and Statistics, K.N.Toosi University of Technology, Mirdamad Blvd, No. 470, Tehran, Tehran Province, Iran

mentioned that setting up preventative initiatives at the primary level is too rigid. Therefore, systematic and thorough screening programmes that accurately and promptly identify the condition can reduce the significant resulting complications and deaths. Mammography, thermography, ultrasonography, and breast biopsy are frequently used screening techniques. However, many communities, particularly in developing nations, lack access to the aforementioned solutions since they are frequently too expensive and difficult to implement. As a result, BC incidence and death have increased to create communities. Additionally, due to time-oriented and cumulative data scarcity, errors in case identification are unavoidable.

The predominant manifestation of BC often presents as the emergence of a novel lump or mass; however, it is important to note that most breast lumps are non-cancerous (Graydon et al. 1997). A painless, firm lesion with irregular margins indicates malignancy; nevertheless, breast carcinomas may manifest as soft, rounded, sensitive, or even painful masses. The use of ML in image processing expedites the diagnosis process, and after that, the tests can confirm the initial diagnoses. Currently, the breast imaging modalities most often used are mammography, ultrasound, and magnetic resonance imaging (MRI) of the breast. Additional diagnostic procedures, such as computed tomography (CT) scans, bone scans, or positron emission tomography (PET) scans, may be used on occasion to facilitate the determination of metastasis in cases of breast cancer (Dye et al. 2012). There are different tests to diagnose BC. If the doctor finds signs of cancer during the test, or if you see signs that indicate the possibility of this type of cancer, more tests are needed to be sure of this issue. Tests such as mammograms, breast ultrasound or breast ultrasounds, and breast MRIs are mainly considered for detection purposes. Having regular check-ups is the best way to deal with this cancer. However, computer-aided diagnosis (CAD) can lead to better results as they reduce trial costs and have better accuracy. Machine learning is a subfield of artificial intelligence that combines a range of statistical, probabilistic, and optimisation techniques to enable computers to 'learn' from previous examples and find difficult-to-detect patterns in vast, noisy, or complex data sets. Consequently, machine learning is increasingly being employed in cancer detection and diagnosis (Osareh and Shadgar 2010). Solid evidence supports the efficacy of conventional laboratory methods like CT and MRI (Tabl et al. 2019). Nevertheless, they fail to disclose anything regarding the mechanism behind cancer development. However, developments in DNA microarray technology have made it possible to collect large numbers of specimens of gene expression. Treatment for breast cancer and survival rates can be used as training data for ML models. Hence, the ML models can be widely employed in medical fields and health assessment.

After analysing the input variables used to predict BC, the ML models can be considered for BC prediction using ML techniques. For this purpose, ML algorithms such as Naive Bayes (NB), Bayesian network (BNeT), random forest (RF), multilayer perceptron (MLP), SVM, eXtreme Gradient Boosting (XGBoost), and C4.5 decision tree can be implemented before. Ensemble learning is an effective technique for prediction enhancement and overfitting possibility reduction. Despite the remarkable interest of scientists in dealing with BC using the various ML techniques, many gaps and shortcomings still need to be tackled as soon as possible. Hence, the current review aims to examine the limitations of the presented ML methods to open a new way for more research. The ML methods' accuracy and speed for detecting BC are investigated here as an innovation. The highest level of accuracy, amounting to 78.6%, was attained through the utilisation of an MLP classifier to identify breast cancer. This classification task was performed using the Wisconsin Prognosis Breast Cancer dataset (WPBC), initially made available in 1970. The study conducted by Yue et al (Yue et al. 2018). Technologies utilised in the healthcare sector encompass the management and retrieval of electronic medical records pertaining to patients and the instruments employed in the process. The identification of cancer has consistently been a significant problem in the realm of diagnosing and developing treatment strategies for haematological illnesses. An overwhelming percentage of the population is affected by one or more diseases. In recent years, there have been significant advancements in medical science. Notwithstanding these developments, a substantial knowledge deficit persists among the general population regarding health and illness. A significant segment of the populace likely experiences health ailments, potentially including those of a life-threatening kind (Kamboj et al. 2021). Accordingly, the issue of accuracy and precision in detection has always been a striking topic that has yet to be comprehensively covered. As seen from the literature, this is the only review that has examined the challenges and benefits of the various ML techniques in terms of accuracy and speed for BC prediction.

The primary purpose of the work is to examine the limitations and strengths of the previous ML methods in detecting breast cancer. Hence, the suggestions obtained from the gaps in the previous works are given. The advent of ML has presented a promising prospect in the battle against breast cancer. ML approaches have exhibited remarkable potential in the timely detection and categorisation of different breast cancer types through the analysis of tumour sizes. ML has demonstrated significant use in the field of image classification, allowing precise prognostications for a range of cancer types, such as breast cancer. Although breast cancer is primarily observed in women, it is crucial to acknowledge that this condition can also afflict men. The importance of new technology in properly addressing breast cancer is evident when considering the anticipation of a specific death rate for this disease in 2025. To address this pressing matter, a thorough systematic evaluation is undertaken, specifically focusing on the numerous machine learning models employed in diagnosing breast cancer. The second purpose of this review is to evaluate the accuracy, precision, and recall of several machine learning models used in breast cancer prediction using image processing techniques. The third aim of the review is to examine recent papers, carefully analysing the advantages and disadvantages of various ML and DL approaches. The knowledge acquired from this extensive assessment can significantly transform the approaches of healthcare professionals in their pursuit of innovative treatments for breast cancer. Prior studies predominantly utilised support vector machines (SVM), k-nearest neighbours (KNN), and decision trees (DT) as primary methods

Figure 1. The growing trend for the applications of ML in the medical field.

for breast cancer diagnosis. However, this review emphasises the need to investigate and incorporate other strategies associated with deep learning methodologies. The use of these sophisticated methodologies has the potential to significantly enhance the precision and effectiveness of breast cancer prediction models. In summary, this study has the potential to facilitate significant progress in breast cancer detection and prognosis, providing a framework for incorporating state-of-the-art ML and DL methodologies. This research seeks to further the medical community's efforts in addressing breast cancer by identifying and discussing the shortcomings of prior methodology. The intention is to encourage the adoption of more rigorous, precise, and effective procedures.

The rest of the paper is organised as follows: The second section briefly reviews the related works and gives essential information. The role of ML in detecting BC is studied in the third section. The fourth section examines the various abilities of the ML method based on the related works. Finally, the fifth section draws on the significant findings and suggestions for future works.

2. Methodology

The use of ML techniques in medical applications is one of the significant achievements of the technology these days. The computer-aided techniques are proposed to tackle the problems and shortcomings of traditional methods. As shown in Figure 1, the term 'machine learning in medical application' has been widely used in the literature.

The review conducted here is systematic and particularly focuses on the applications of ML in the treatment of BC. Accordingly, about 120 were extracted from the main search engines, such as Google Scholar and Science Direct, in the beginning. After categorising the related papers based on the purposes and methods, about 76 papers were cited in this review. The primary sources of paper extraction were Springer and Elsevier. The review outlines the major gaps and shortcomings that can be starting points for future works.

3. Literature review

Many existing studies in the broader literature have examined the effectiveness of ML methods in BC detection (Gayathri et al. 2013; Alarabeyyat and Alhanahnah ; Bazazeh and Shubair). BC affects 8% of women over their lifetime; behind lung cancer, it is the second leading cause of fatalities in both the developed and developing worlds. BC is distinguished by gene mutation, continuous pain, and variations in the size, colour (redness), and skin texture of the breasts. Pathologists use BC classification to find a systematic and objective prognosis; the most common classification is binary (benign cancer/malignant cancer). ML techniques are now widely used in the classification of BC. They offer great classification accuracy as well as powerful diagnostic capabilities. In another study, two distinct classifiers were proposed for BC classification: the Naive Bayes (NB) classifier and the nearest neighbour (KNN) classifier (Amrane et al. 2018). A comparison of the two new solutions and the use of cross-validation were made to assess their accuracy. The outcomes indicated that KNN has the highest efficiency (97.51%) with the lowest error rate, followed by the NB classifier (96.19%). In another research, the support vector machine (SVM), KNN, random forests, artificial neural networks (ANNs), and logistic regression supervised machine learning algorithms were presented (Islam et al. 2020). The Wisconsin Breast Cancer dataset was taken from the UCI machine learning database, a well-known machine learning resource. The reliability, specificity, sensitivity, precision, adverse predictive value, false-negative rate, false-positive rate, F1 score, and Matthews Correlation Coefficient are used to assess the study's performance. Furthermore, the precision-recall area under curve and receiver operating characteristic curve of various approaches were evaluated. The results showed that ANNs have the greatest precision, accuracy, and F1 scores of 98.57%, 97.82%, and 0.9890, respectively, while SVM had the best accuracy, precision, and F1 scores of 97.14%, 95.65%, and 0.9777, respectively. In 2021, J. Wu and C. Hicks proposed using gene expression data to classify patients with triple-negative breast cancer and non-triple-negative breast cancer using an ML approach (Wu and Hicks 2021). To identify the features (genes) used in the construction and validation of the classification models, RNA-Sequence data was analysed from 110 triple negative and 992 non-triple negative BC tumour samples from the Cancer Genome Atlas. We tested four classification models for categorising the two forms of breast cancer, namely SVMs, KNN, NB, and Decision Tree (DT), using characteristics picked at varying threshold levels to train the models. The proposed approaches were utilised on independent gene expression datasets to assess effectiveness and validation. The SVM method categorised breast cancer more precisely into triple minus and non-triple opposite breast cancer and exhibited fewer misclassification errors than all three tested algorithms. The prediction results demonstrated that ML algorithms are practical for classifying breast cancer into triple negative and non-triple negative categories. In 2019, Ganggayah et al. (Ganggayah et al.

2019) employed various ML techniques to create models for detecting and visualising relevant BC survival rate prognostic factors.

Early disease diagnosis has been a critical challenge due to the recent population expansion in medical studies (Islam et al. 2020). With the tremendous population growth, the danger of death caused by breast cancer has increased rapidly. In recent years, many advances have been made to ML techniques, the primary medical tool (Allegra et al. 2012; Cardoso et al. 2016; Abou Tabl et al.). In 2019, Tabl et al. (Tabl et al. 2019) proposed a novel technique for detecting the symptoms of BC. Accordingly, the authors conducted clinical operations for 347 patients, considering the combination of feature selection techniques and a prediction method. The findings revealed that the proposed model can successfully detect the classes with high-performance measurements. Also, the ultrasonic scan is the most extensively utilised approach for diagnosing geological illness, i.e. BC. The initial stage for recognising the anomaly of breast cancer (malignant from benign) is removing the region of interest (ROI). For this purpose, a new strategy for breast ROI extraction was suggested to minimise false positive cases (FP) (Zeebaree et al. 2019). The efficacy of the suggested method was contrasted with the current methods utilised to divide up various types of images. In another work (Saber et al. 2021), an innovative DL model based on the transfer-learning (TL) technique was proposed to efficiently help automatically detect and diagnose the BC suspicious region based on 80–20 and cross-validation techniques.

Furthermore, Sharma et al (Sharma et al. 2018). compared three widely used machine learning algorithms and methodologies for breast cancer prediction: Random Forest (RF), kNN, and NB. The Wisconsin Diagnosis Breast Cancer data set was employed as a training set to examine the performance of several machine learning approaches in terms of essential characteristics such as accuracy and precision. The outcomes were extraordinarily comparable and might be applied to detection and therapy. In 2016, Asri et al (Asri et al. 2016). used the Wisconsin Breast Cancer (original) datasets to analyse the performance of the various ML methods, including SVM, DT, NB, and k-NN. The primary goal was to evaluate the correctness of data classification in terms of the efficiency and effectiveness of each method in terms of accuracy, precision, sensitivity, and specificity. The experimental findings demonstrated that SVM has the most excellent accuracy (97.13%) and the lowest error rate. All experiments were conducted in a simulated environment using the WEKA data mining tool. The Nottingham histological grade (NHG) for breast cancer is a well-established predictive indicator widely utilised in clinical decision-making. 50% of patients are classed as grade 2, an intermediate-risk group with little clinical value. In another study, a novel histological grade model (DeepGrade) was developed using digital whole-slide histopathology images (WSIs) and deep learning to improve risk stratification in NHG 2 breast cancer patients (Wang et al. 2022). More studies are also summarised in Table 1 based on the methods and accuracy.

4. Use of computer-aided techniques for BC diagnosis

The hypothesis of hormone reliance in breast cancer was initially postulated based on the observation of the disease's aggressive nature in younger women. In 1906, Beatson played a pivotal role in initiating the period of endocrine surgery, predating Jensen's discovery of oestrogen receptors in 1967 and the subsequent popularity of oophorectomy and adrenalectomy as methods of achieving castration (Horsley and Horsley 1962). Oestrogen receptor modulators, luteinising hormone-releasing agonists, and aromatase inhibitors progressively superseded the utilisation of these more aggressive approaches. The preservation of Halstead's legacy was temporarily upheld by Margottini and Veronesi in Milan, who also removed internal mammary nodes. Furthermore, other individuals expanded the concept of 'radicality' by including the removal of supraclavicular and mediastinal nodes. Nevertheless, during the late 19th and early 20th centuries, a progressive shift marked the decline of the belief that larger surgeons were more skilled, as indicated by their ability to make larger incisions and execute more extensive surgery. Patey and Handley, hailing from London, and Auchincloss Jr., based in

Table 1. A review of the related works based on the method and accuracy.

No	Reference	Method	Accuracy
1	(Allugunti 2022)	Convolutional Neural Network (CNN), SVM, RF, Convolutional Networks, SVM, RF	Processing mammography images before the use of ML techniques increased the accuracy
2	(Michael et al. 2022)	KNN, SVM, RF, Xgboost, Lightgbm	Lightgbm has the accuracy, precision, recall, and the F1 score of 99.86%, 100%, 99.6%, and 99.8%
3	(Siddiqui et al. 2021)	An Internet of Medical Things (IoMT) cloud-based model	The accuracies required for detecting ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma were 99.69%, 99.32%, 98.96%, and 99.32%
4	(Ak 2020)	LR, KNN, SVM, NV, and DT	A logistic regression model with a classification accuracy of about 98%
5	(Sha et al. 2020)	You Only Look Once (YOLO) and RetinaNet	The accuracy and precision of 79% and 91% for the detection
6	(Sha et al. 2020)	Image noise reduction, optimal image segmentation based on the convolutional neural network, a grasshopper optimization algorithm, and optimized feature extraction and feature selection based on the grasshopper optimization algorithm,	96% Sensitivity, 93% Specificity, 85% PPV, 97% NPV, 92% accuracy
7	(Zheng et al. 2020)	Deep Learning assisted Efficient Adaboost Algorithm (DLA-EABA)	Accuracy, sensitivity, and specificity of 97%, 98%, and 96%
8	(Wang et al. 2022)	Hybrid deep hybrid learning (CNN-GRU)	Accuracy, precision, and sensitivity, specificity of 86%, 85.5%, 85.6%, 84.7%

New York, spearheaded a transformative movement aimed at modifying the radical mastectomy procedure while ensuring the preservation of the pectoralis major muscle (Thornes 1967). The rapid progression of medical radiation techniques for cancer cell eradication, alongside the development of novel kinds of chemotherapy that accomplish the same objective and induce medical castration or target altered tumour receptors, has necessitated a re-evaluation of methodologies employed in cancer management. The growing understanding of the biological characteristics of breast cancer and the limited efficacy of surgery as a standalone treatment accompanied these findings. The introduction of mammography for the early diagnosis of tiny lesions has significantly enhanced the surgical management of cancer (O H C D Panel 1979).

It is projected that in the year 2022, an estimated 287,850 incidents of invasive breast cancer and 51,400 instances of ductal carcinoma in situ (DCIS) will be detected in women residing in the United States. Additionally, it is anticipated that 43,250 women will experience mortality as a result of breast cancer. The majority of invasive breast cancers, namely 83%, are detected in women who are 50 years old or above. Additionally, a significant proportion, 91%, of breast cancer-related fatalities are seen within this age demographic. Furthermore, half of all breast cancer deaths are reported among women aged 70 years or above. The median age of diagnosis for female breast cancer is generally seen to be 62 years; however, it tends to be slightly lower for Hispanic (57 years), Asian/Pacific Islanders (API) (58 years), Black (60 years), and AIAN (61 years) women compared to White women (64 years). This discrepancy can be attributed, at least in part, to variations in the age distribution of these respective populations. The median age at which individuals succumb to breast cancer is 69 years on average. However, this age varies across different racial and ethnic groups, with White women experiencing a median age of 70 years, Hispanic women at 62 years, and API and Black women at 63 years (Denise Jozwik et al. 2023). As of 1 January 2022, an estimated population of 4.1 million women in the United States were reported to have a documented medical history of breast cancer. Around 4% of these women are now suffering from metastatic illness, with over half of them first being diagnosed with early-stage (I-III) malignancies (Gallicchio et al. 2022).

The breast comprises several primary constituents, including: Lobules are integral components of the glandular system, serving as glandular structures responsible for breast milk production. Lobules are shown to be organised in clusters, together comprising a lobe. Ducts are tiny conduits that transport breast milk from the lobules to the nipple (Ma et al. 2019). Because breast cancer is women's second most significant cause of death, accurate early identification can significantly reduce breast cancer mortality rates (Houssein et al. 2021). Radiologists can detect abnormalities more efficiently using computer-aided detection. Medical images provide information that can be used to detect and diagnose various diseases and abnormalities. Several modalities allow radiologists to investigate the interior structure, which has sparked interest in various research areas. Each of the modalities, as mentioned earlier, is important in some medical domains.

Inspired by pattern and computational learning theory, ML examines the study and construction of algorithms that can learn and make predictions based on data (Shailaja et al. 2018; Wiens and Shenoy 2018; Siddique and Chow 2021). Such algorithms do not simply follow the program's commands and make predictions or decisions by modelling sample input data. ML is used in computing tasks where designing and programming explicit algorithms with proper performance is difficult or impossible. Some applications include email filtering, identification of Internet intruders or internal malware that intends to breach information, optical text reader, ranking learning, and machine vision. The need to automate decision-making and decision-making processes has increased with the expansion of information technology applications in various fields. As one of the leading solutions to meet these needs, artificial intelligence uses methods based on machine learning. ML is closely related to computational statistics and often overlaps with it. The focus of this branch is computer prediction, and it has a strong link with mathematical optimisation, which also brings methods, theories, and applications into the field. Machine learning is sometimes combined with data mining, and this subsection focuses on the exploratory analysis of data, known as unsupervised learning. ML can also be unsupervised and can be used to learn and recognise different organisms' initial forms of behaviour and then find meaningful anomalies. In data analysis, ML is a method for designing complex algorithms and models used for prediction; this is known as predictive analytics in the industry. These analytical models allow researchers, data science researchers, engineers, and analysts to obtain reliable and repeatable decisions and results and, by learning from the relationships and trends related to the past, reveal the hidden frosts (Swain et al. 2022).

Much attention has been devoted to the applications of ML in detecting breast cancer, as shown in the literature. In 2020, Vaka et al. proposed a novel way of detecting breast cancer using ML algorithms (Vaka et al. 2020). To assess performance, the authors conducted an experimental analysis on a dataset. Compared to existing methods, the proposed method generated precise and efficient results. Data mining methods are crucial in predicting early-stage breast cancer. A strategy was also provided that increases the precision and efficiency of three different classifiers: DT, NB, and Sequential Minimal Optimization (SMO) (Mohammed et al. 2020). The classifiers were also validated and compared on two benchmark datasets: Wisconsin Breast Cancer (WBC) and Breast Cancer dataset. Because the chance of cases falling within the majority class was very high, the ML models were considerably prone to categorise novel findings to the majority class. This work addressed such a difficulty. The authors employed the data level technique, which entailed resampling the data to offset the effect of class imbalance. 10-fold cross-validations were used for evaluation. Each classifier's efficiency was measured regarding true positive, false positive, Roc curve, standard deviation (Std), and accuracy (AC). Experiments demonstrated that applying a resample filter improves the classifier's performance, with SMO outperforming another in the WBC dataset and J48 outperforming competitors in the Breast Cancer dataset. In other work, two of the most prominent ML approaches were employed to classify the Wisconsin Breast Cancer

(Original) dataset. Their classification performance was compared using accuracy, precision, recall, and ROC Area values (Bayrak et al. 2019). The SVM approach produced the greatest results with the best accuracy. On the Wisconsin Breast Cancer Diagnostic (WBCD) dataset, Bayrak et al. compared the performance of five nonlinear machine learning algorithms: MLP, KNN, Classification and Regression Trees (CART), Gaussian Naive Bayes (NB), and SVM (Bayrak et al. 2019). The major goal was to assess each algorithm's performance in categorising data in terms of efficiency and effectiveness based on classification test accuracy, precision, and recall. An ML and image processing-based evolutionary strategy was presented to identify and detect breast cancer (Jasti et al. 2022). To aid in classifying and detecting skin diseases, this model integrated image preprocessing, feature extraction, feature selection, and ML approaches. A geometric mean filter was utilised to improve the image's quality. AlexNet was employed for feature extraction, and the relief algorithm was also considered to pick features. The model employed ML techniques such as least square SVM, KNN, RF, and NB for disease categorisation and detection. MIAS data collection was used in the experimental inquiry. Using image analysis, this proposed method was useful for reliably recognising breast cancer disease. Due to the complicated molecular variety, triple-negative breast cancer (TNBC) is challenging to diagnose and treat. To address these issues, employing artificial intelligence to forecast the cellular uptake of nanoparticles (NPs) against distinct cancer stages was suggested (Alafeef et al. 2020). For the first time, the authors showed that a ML method combined with distinctive cellular uptake responses for particular cancer types might successfully classify various cancer cell types. This method optimised nanomaterials to achieve the best structure-internalisation response for a specific particle (Alafeef et al. 2020).

Despite its high cost and numerous adverse effects, mammography is frequently employed as the most frequent laboratory approach for detecting breast cancer. ML prediction has demonstrated promising outcomes as an alternate strategy. Mojrian et al. provided a strategy for detecting breast cancer using an extreme learning machine (ELM) classification model linked with a radial basis function (RBF) kernel termed ELM-RBF, utilising the Wisconsin dataset (Mojrian et al. 2020). The proposed model's performance was then compared to a linear SVM model. The suggested model beat the linear-SVM model by having RMSE, R 2, and MAPE values of 0.1719, 0.9374, and 0.0539, respectively. Furthermore, both models' accuracy, precision, sensitivity, specificity, validity, true positive rate (TPR), and false-negative rate (FNR) were investigated. The ELM-RBF model performed better for these criteria than the SVM model. A thermogram-based breast cancer detection method was proposed in another research (AlFayez et al. 2020). This method was divided into four stages: (1) Image preprocessing with homomorphic filtering, top-hat transform, and adaptive histogram equalisation, (2) ROI segmentation with binary masking and K-mean clustering, (3) feature extraction with signature boundary, and (4) classification with Extreme Learning Machine (ELM) and Multilayer Perceptron (MLP) classifiers. The proposed method was tested using the general dataset DMR-IR. Various experiment situations (for example, integration of geometrical and textural feature extraction) were constructed and assessed using various measurements (for example, accuracy, sensitivity, and specificity). It was revealed that ELM-based outcomes outperformed MLP-based results by more than 19%. In a review of 2021 (Meenalochini and Ramkumar 2021), the methods and processes proposed for cancer tumour classification were examined. The performance of several classification techniques was compared. The authors stated that classification accuracy can be enhanced using hybrid techniques.

Notably, many databases like () can be used to assess the performance of the various ML techniques in detecting breast cancer. The features are derived from an image that has been scanned, which represents a fine needle aspirate (FNA) of a breast lump. The properties of the cell nuclei seen in the photograph are typically described in the sources of the considered databases. In order to verify the performance of the used ML or DL techniques, the effectiveness of the proposed methods in the various databases has been neglected so far. The breast dataset is a comprehensive collection of data encompassing a significant portion of the Prostate, Lung, Colorectal, and Ovarian (PLCO) data about breast cancer incidence and death analysis (). In the case of several women, the study records the occurrence of multiple instances of breast cancer. However, the present document encompasses data about the initial breast cancer diagnosis within the experiment. The dataset has entries for approximately 78,000 women participating in the PLCO experiment. The Breast Secondary dataset comprises data about supplementary breast malignancies recorded throughout the experiment and gathered using the Breast Cancer Supplement form. The collection comprises individual records for about 78,000 women participating in the PLCO experiment.

5. The various ML techniques

Medicine and health are very important for the continuation of human life, and the applications of artificial intelligence in medicine and health have increased in recent years (Chen et al. 2021). The research in the fields related to medicine, medicine, and services for people with disabilities indicates that artificial intelligence technology can create significant changes in fields such as disease diagnosis, treatment methods, drug disorders, and medical image processing (Jain and Chatterjee 2020). Among the other factors influencing the health of the human body, we can mention exercise and healthy nutrition (McCoy et al. 2020). Artificial intelligence, with the ability to analyse and process information quickly, as well as technologies such as machine vision, the Internet of Things, and robotics, has been able to provide software, platforms, and practical gadgets to improve the quality of human life in the field of sports and nutrition (Ngiam and Khor 2019). Healthcare companies have the potential to enhance healthcare efficiency and achieve cost savings via the utilisation of ML technology. One potential use of ML in the healthcare domain is the advancement of algorithms to enhance the management of patient information and the scheduling of appointments. This ML form can mitigate the inefficiencies associated with repetitive tasks in the healthcare system, optimising time and resource use.

Indeed, ML is a potent subfield within the realm of AI. However, it is essential to note that AI encompasses a broader range of approaches and methodologies, each with distinct merits and practical uses. The selection of an appropriate AI approach is contingent upon several factors, including the particular situation at hand, the characteristics of the data, the requirement for interpretability, and the intended objectives. Because it serves as the basis for both learning and generalisation, accessing an appropriate dataset is fundamental to accomplishing machine learning models. However, there may be obstacles in the way of fully realising the promise of ML approaches due to problems with data quality and availability, prejudice, and ethical implications. In ML, it is vital to have efficient data collecting, curation, and preprocessing in place in order to make the most of the benefits related to dataset needs and avoid the drawbacks. The primary objective of CAD is to enhance the precision and uniformity of diagnostic imaging via the use of image processing, computer vision, and machine learning methodologies, which provide a remarkable performance.

The BC datasets of the University of California Irvine (UCI) are a benchmark online resource utilised extensively in the literature. Nithya and Santhi obtained 97.8% efficiency using a multiboost ensemble technique (Frank and Asuncion 2010; Weli 2020). Another study proposed an IoT-based student healthcare monitoring model using innovative technologies to continuously assess student vital signs and identify biological and behavioural abnormalities (Souri et al. 2020). In this concept, crucial data was collected using IoT devices, and data analysis was performed using ML methods to predict potential dangers of student physiological and behavioural abnormalities. Student health information was saved in the cloud layer. Accordingly, the data analysis tasks were conducted to determine the students' health state. The results of this layer determined the subsequent actions in the monitoring layer (Ngiam and Khor 2019). The experimental findings showed that the suggested model efficiently and accurately detected the students' condition. The SVM attained an excellent accuracy of 99.1% after analysing the given model, which was a good outcome for our goal. The results outperformed decision trees, random forests, and multilayer perceptron neural networks. Table 2 reviews the various ML models used for detecting BC.

As can be seen from Table 2, the performance of SVM has been acceptable in the previous research as it reached about 97% accuracy. The comparison between the studies of Naji et al. and Khourdifi & Bahaj revealed that the classification accuracy of SVM is better. Also, many other studies presented remarkable findings in this regard. For instance, a novel technique termed BDR-CNN-GCN was proposed by Zhang et al. in 2021 (Zhang et al. 2021) which consisted of a graph convolutional network (GCN) and CNN to better identify malignant lesions in breast mammograms. The proposed method outperformed the other state-of-art BC detection methods and five suggested NN models. In another work (Zhang et al. 2018), the breast dataset was selected as the open-access mini MIAS dataset to balance the dataset, cost-sensitive learning. The training set's size was increased using data augmentation and a CNN with nine additional layers. The authors contrasted the rectified linear unit (ReLU), leaky ReLU, and parametric ReLU activation functions. Six pooling methods were also contrasted, including rank-based average pooling, rank-based weighted pooling, rank-based stochastic pooling, and average, max, and stochastic pooling. The findings emphasised the superiority of the DL method over the traditional artificial intelligence methods in terms of detection accuracy. Notably, ensemble algorithms combine the results of different base models to improve the overall predictive performance in a single model. By combining the results of multiple models,

Table 2. A brief review of the methods, purposes, and accuracy.

No	References	Method	Purpose	Accuracy
1	(Naji et al. 2021)	SVM, Random Forest, Logistic Regression, Decision tree and KNN	Anticipating and diagnosing BC based on ML models	The superiority of SVM with 97.2% accuracy
2	(Khourdifi and Bahaj 2018)	RF, NB, SVM, and K-NN	Classification and prediction of BC	SVM with 97.9%
3	(Ngiam and Khor 2019)	SVM, CART, NB and kNN	Detection and prediction	KNN, NB, and CART have better accuracy
4	(Ghosh et al. 0000)	Wisconsin Breast Cancer (Diagnostic) Dataset. Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU)	Detection and anticipation	99% accuracy
5	(Fatima et al. 2020)	Comprehensive analysis and review	Examining the various ML models	SVM had remarkable accuracy
6	(Shanbehzadeh et al. 2022)	Naïve Bayes (NB), Bayesian network (BNeT), random forest (RF), MLP, SVM, C4.5, eXtreme Gradient Boosting (XGBoost), decision tree and two ensemble algorithms, including Confidence weighted voting and Voting	Assessing BC based on the evaluation of the previous ML techniques	With AUC values of 0.799 and 0.798, the RF algorithm had the best performance both before and after executing FS.
7	(Nomani et al. 2022)	Particle swarm optimized wavelet neural network (PSOWNN)	BC prediction	Specificity of 98.8%, precision of 98.6%, accuracy of 95.2%
8	(Vrdoljak et al. 2023)	Neoadjuvant systemic therapy (NST) based on XGBoost	Detection of axillary lymph node status	AUC: 0.762 [0.726, 0.795]
9	(Dehdar et al. 2023)	XGBoost, RF, NNs, and LR	Anticipation of factors for delayed BC diagnosis in Iranian women	99% accuracy
10	(Mirza et al. 2023)	The diagnostic model's accuracy was improved by cutting-edge statistical techniques and cross-validation with different ML techniques, which also forecasted a fresh diagnostic nine-gene signature.	Reducing the number of genes in a large cohort of transcriptomics data in order to build a diagnostic model for cancer classification.	With accurate diagnosis and prognosis at a lower cost, the found gene signature biomarkers enhanced healthcare management.

the advantages and disadvantages of the various models emerge, correct predictions are reinforced, and incorrect predictions are cancelled out. Most ensemble algorithms are 'black boxes' because the underlying base models are randomly generated and are not led by exact predictions (Ho ; Ng and Soo 2017).

In addition, the performance of the ML techniques have been always various in the updated publications. The RF approach demonstrated superior performance both prior to and during feature selection (FS), achieving area under the curve (AUC) values of 0.799 and 0.798, respectively. Moreover, the utilisation of the Confidence Weighted Voting technique enhanced the classifier's performance, leading to the attainment of the optimal outcome, characterised by an 80% (Shanbehzadeh et al. 2022). This study employed various machine learning classification techniques, including NB, LR, SVM, KNN, DT, and ensemble techniques such as RF, Adaboost, and XGBoost. These techniques were applied to a breast cancer dataset, and their performance was assessed using various performance measures. The research findings obtained by Nemade and Fegade (Nemade and Fegade 2023) indicated that the decision tree and XGBoost classifier exhibited the best accuracy, reaching 97%, compared to other models. Additionally, the XGBoost classifier achieved the highest AUC value of 0.999. The research comprehensively analysed ML techniques employed for detecting breast cancer. The analysis unveiled a widespread use of SVM, KNN, and DT in the extant scholarly literature. Moreover, recent studies have noticed a notable increase in the utilisation of deep learning DL methodologies, including CNN and RNN.

ML models are predominantly influenced by the data they are trained on and the patterns that may be extracted. These individuals lack moral discernment, empathy, and ethical deliberation. In disciplines such as law, justice, social work, and philosophical decision-making, where intricate ethical deliberations, contextual factors, and moral tenets are important, exclusive dependence on ML may result in choices that might raise ethical concerns or lack human empathy. For example, ML models alone may not sufficiently handle identifying the suitable penalty in legal situations, making choices in child welfare services, or addressing intricate ethical challenges that need consideration of individual circumstances and moral reasoning. These scenarios frequently need human comprehension, compassion, and the capacity to decipher intricate, context-dependent elements, which machine learning models may not fully encompass. In instances of this nature, although ML can aid in examining data or providing insights, the ultimate decision-making procedure may necessitate the inclusion of human judgement and ethical deliberation to guarantee equitable, impartial, and ethically sound results. Recognising these inherent constraints, a concerted endeavour is underway to include ethical frameworks in advancing AI and ML. The primary objective of ethical AI frameworks is to establish a harmonious alignment between artificial intelligence technology and human values, as well as ethical standards. This alignment facilitates decision-making processes that are characterised by responsibility and thoughtfulness. Nevertheless, the complete incorporation of moral thinking and ethical judgement into AI systems continues to pose a significant obstacle.

6. Conclusion

ML teaches computers to understand patterns from data in a specific domain by developing mathematical models. In its most basic form, machine learning is a two-step process. First, a model is constructed utilising sample data as input, referred to as the 'training set', plus the model receives the correct outputs. Models can be created using a variety of ML algorithms, such as Logistic Regression (LR). Following training, the model is evaluated using previously unknown data, referred to as the 'test set'. The model is meant to foresee the output of the test set during the testing stage, which occurs with a certain level of accuracy. A model can be efficient if it performs well in training and testing and vice versa. The current systematic review reviewed the various ML models used for detecting BC based on accuracy, precision, and recall. It was found that despite such remarkable interest, many gaps and shortcomings still need to be tackled as soon as possible. The previous works mainly benefited from SVM, KNN, and DT in detecting BC. As stated in the literature, SVM performs better for classification as it can also be improved using the other solutions. In the future, more comparative studies are required to make better comparisons. Different deep learning-based techniques, such as CNN, DNN, RNN, DBN, and AE-based approaches, have recently been developed to diagnose breast cancer. The most prominent deep-learning method, CNN, has been used in multiple studies to detect breast cancer. Combining various risk factors in breast cancer prediction modelling could aid early illness detection and provide the appropriate therapy. The advent of DL methodologies, particularly CNN and RNN designs, offers a promising opportunity for the diagnosis of breast cancer. These methodologies have considerable potential in effectively managing unstructured data, namely medical photos, and indicate enhanced performance in the extraction of features and classification. The assessment underscored the critical necessity of tackling many issues, such as enhancing the interpretability of machine learning models, managing unbalanced datasets, and guaranteeing the resilience and applicability of the created algorithms. By incorporating approaches of explainable artificial intelligence AI and undertaking thorough validation on a wide range of datasets, it is possible to alleviate these issues.

Regarding the future works, the challenges of image processing techniques used in the medical field require more consideration. The integration of multimodal data in breast cancer diagnosis is still limited since several research primarily concentrate on analysing specific types of data such as mammograms, genetic markers, and histology. There is a need for comprehensive models that adeptly include several data modalities, including imaging, genomics, proteomics, and clinical data, aiming to augment accuracy and resilience in the domains of diagnosis and therapy prediction. Despite the encouraging outcomes observed in controlled research investigations, the application of machine learning models in clinical practice continues to face constraints in terms of validation. The research should thoroughly validate these models inside authentic clinical environments to evaluate their effectiveness, practicality, and influence on patient outcomes. Also, the primary focus of this study is on early detection and precision medicine, emphasising the implementation of early detection models that exhibit high levels of sensitivity and specificity. The integration of such models into clinical practice has the potential to enhance patient outcomes greatly. Furthermore, developing more sophisticated models for predicting therapy response and the identification of individualised therapeutic alternatives based on molecular profiles is a promising avenue of research.

Disclosure statement

"The author declares that (s)he has no relevant or material financial interests that relate to the research described in this paper".

Notes on contributor

Ali Jafari completed his Master's degree from K. N. Toosi University of Technology. He is a second-year social administration student at the University of Chicago Crown Family School of Social Work, Policy, and Practice. He received a bachelor's degree in industrial engineering from University of Tabriz and a master's degree in data science from K.N.Toosi University of Technology in Tehran, Iran. He is interested in machine learning, computer vision, and natural language processing.

References

Abou Tabl A, Alkhateeb A, ElMaraghy W, and Ngom A. 2017. Machine learning model for identifying gene biomarkers for breast cancer treatment 805 survival. Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; p. 607–607.

Ak MF. 2020. A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. 810 Healthcare. 8(2):111. MDPI.

Alafeef M and Srivastava I, Pan D. 2020. Machine learning for precision breast cancer diagnosis and prediction of the nanoparticle cellular internalization. ACS Sens. 5(6):1689–1698. doi: 10.1021/acssensors.0c00329.

Alarabeyyat A and Alhanahnah M. 2016. Breast cancer detection using k-nearest neighbor machine learning algorithm. 2016 9th International Conference on Developments in eSystems Engineering (DeSE); IEEE, p. 35–39.

AlFayez F, El-Soud MWA, Gaber T. 2020. Thermogram breast cancer detection: a comparative study of two machine learning techniques. Appl Sci. 10(2):551. https://www.mdpi.com/2076-3417/10/2/551.

Allegra A, Alonci, A., Campo, S., Penna, G., Petrungaro, A., Gerace, D., & Musolino, C. 2012. Circulating microRnas: new biomarkers in diagnosis, prognosis and treatment of cancer. Int J Oncol. 41(6):1897–1912. doi:10.3892/ijo.2012.1647.

Allugunti VR. 2022. Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. Int J Eng Comput Sci. 4(1):49–56. doi: 10.33545/26633582.2022.v4.i1a.68.

Amrane M, Oukid S, Gagaoua I, and Ensarİ T. 2018. Breast cancer classification using machine learning. 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT); Apr 18-19; p. 1–4, doi: 10.1109/EBBT.2018.8391453.

Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. 2016. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging. 35 (5):1207–1216. doi: 10.1109/TMI.2016.2535865.

Asri H, Mousannif H, Al Moatassime H, Noel T. 2016. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Comput Sci. 83:1064–1069. doi: 10.1016/j.procs.2016.04.224.

Bayrak EA, Kırcı P, and Ensari T. 2019. Comparison of machine learning methods for breast cancer diagnosis. 2019 Scientific Meeting on Electrical- Electronics & Biomedical Engineering and Computer Science (EBBT); April 24-26; p. 1–3. doi: 10.1109/EBBT.2019.8741990.

Bazazeh D and Shubair R. 2016. Comparative study of machine learning algorithms for breast cancer detection and diagnosis. In 2016 5th international conference on electronic devices, systems and applications (ICEDSA); IEEE, p. 1–4.

Bhardwaj R, Nambiar AR, and Dutta D. 2017. A study of machine learning inhealthcare. 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); July 4-8; 2, p. 236–241, doi: 10.1109/COMPSAC.2017.164.

Callahan A, Shah NH. 2017. Chapter 19 - machine learning in healthcare. In: Sheikh A, Cresswell KM, Wright A Bates DW, editors Key advances in clinical informatics. Academic Press; pp. 279–291.

Cardoso F, Van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga J-Y, Brain E, Causeret S, DeLorenzi M, et al. 2016. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med. 375 (8):717–729. doi:10.1056/NEJMoa1602253.

Chen Y-W, Jain LC. 2020. Deep learning in healthcare. *Cham*: Springer.

Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. 2021. Ethical machine learning in healthcare. Annual Review Of Biomedi Data Sci. 4 (1):123–144. doi: 10.1146/annurev-biodatasci-092820-114757.

Dehdar S, Salimifard, K., Mohammadi, R., Marzban, M., Saadatmand, S., Fararouei, M., & Dianati-Nasab, M. 2023. Applications of different machine learning approaches in prediction of breast cancer diagnosis delay. Front Oncol. 13:1103369. doi: 10.3389/fonc.2023.1103369.

Denise Jozwik R, Teresa Chapple C, Shannon Makinen R, Tessa Morrison C, Patti Rose R, Eck R. 2023. PEDIATRIC CANCER IN IDAHO, 2011–2020.

Dhillon A, Singh A. 2019. Machine learning in healthcare data analysis: a survey. J Biol Today's World. 8(6):1–10.

Dye TD, Bogale S, Hobden C, Tilahun Y, Deressa T, Reeler A. 2012. Experience of initial symptoms of breast cancer and triggers for action in Ethiopia. Int J Breast Cancer. 2012:1–5. doi: 10.1155/2012/908547.

Fatima N, Liu L, Hong S, Ahmed H. 2020. Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. IEEE Acces. 8:150360–150376. doi: 10.1109/ACCESS.2020.3016715.

Frank A, Asuncion A. 2010. Multitude Classifier Using Rough Set Jelinek Mercer Naïve Bayes for Disease Diagnosis. Circuits Syst.

Gallicchio L, Devasia TP, Tonorezos E, Mollica MA, Mariotto A. 2022. Estimation of the number of individuals living with metastatic cancer in the United States. JNCI J Nat Cancer Inst. 114(11):1476–1483. doi: 10.1093/jnci/djac158.

Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK. 2019. Predicting factors for survival of breast cancer patients using machine learning techniques. Bmc Med Inform Decis. 19(1):48. 2019/03/22. doi: 10.1186/s12911-019-0801-4.

Gayathri B, Sumathi C, Santhanam T. 2013. Breast cancer diagnosis using machine learning algorithms–a survey. Int J Parallel Emergent Distrib Syst. 105–112.

Ghosh P, Azam S, Hasib KM, Karim A, Jonkman M, and Anwar A. 2021. A performance based study on deep learning algorithms in the effective prediction of breast cancer. 2021 International Joint Conference on Neural Networks (IJCNN); IEEE, p. 1–8.

Graydon J, 1997. Information needs of women during early treatment for breast cancer. J Adv Nurs. 26(1):59–64. doi:10.1046/j.1365-2648.1997.1997026059.x.

Havaei M, Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y. & Larochelle, H. 2017. Brain tumor segmentation with deep neural networks. Med Image Anal. 35:18–31. doi: 10.1016/j.media.2016.05.004.

Ho TK. Random decision forests. 1995. Proceedings of 3rd international conference on document analysis and recognition; 1. IEEE, p. 278–282.

Horsley SJ, Horsley GW. 1962. Twenty years' experience with prophylactic bilateral oophorectomy in the treatment of carcinoma of the breast. Ann Surg. 155(6):935–939. doi: 10.1097/00000658-196206000-00013.

Houssein EH, Emam MM, Ali AA, Suganthan PN. 2021. Deep and machine learning techniques for medical imaging-based breast cancer: a comprehensive review. Expert Syst Appl. [2021/04/1] 167:114161. doi:10.1016/j.eswa.2020.114161.

Islam MM, Haque MR, Iqbal H, Hasan MM, Hasan M, Kabir MN. 2020. Breast cancer prediction: a comparative study using machine learning techniques. SN Comput Sci. 1(5):290. 2020/09/1. doi; 10.1007/s42979-020-00305-w.

Jain V, Chatterjee JM. 2020. Machine learning with health care perspective. *Cham*: Springer; pp. 1–415.

Jasti VDP, Zamani, A. S., Arumugam, K., Naved, M., Pallathadka, H., Sammy, F. & Kaliyaperumal, K. 2022. Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis. Secur Commun Netw. 2022:1–7. doi: 10.1155/2022/1918379.

Kamboj A, Tanay P, Sinha A, and Kumar P. 2021. Breast cancer detection using supervised machine learning: a comparative analysis. Information Management and Machine Intelligence: Proceedings of ICIMMI 2019; Springer, p. 263–269.

https://trends.google.com/trends/explore?date=now%201-d&q=machine%20learning%20in%20medical%20field&hl=en-US

Kaul DTripathy B. 2022 . Deep learning in healthcare. Deep Learn Data Anal: Recent Tech Pract Appl. 97–115

Khourdifi Y and Bahaj M. 2018. Applying best machine learning algorithms for breast cancer prediction and classification. 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS); Dec 5-6; p. 1–5, doi: 10.1109/ICECOCS.2018.8610632.

Latif S, Qadir J, Farooq S, Imran MA. 2017. How 5G wireless (and concomitant technologies) will revolutionize healthcare? Future Internet. 9(4):93.

Ma C, Nepal M, Kim J-H, Fan P, Fei P. 2019. A new look at molecular biology of breast cancer. Cancer Biol Ther. 20(1):1–5. doi: 10.1080/15384047.2018.1507259.

McCoy LG, Banja JD, Ghassemi M, Celi LA. 2020. Ensuring machine learning for healthcare works for all. BMJ Health & Care Inf. 27(3):e100237. doi: 10.1136/bmjhci-2020-100237.

Meenalochini G and Ramkumar S. 2021. Survey of machine learning algorithms for breast cancer detection using mammogram images. Materials Today: Proceedings; 37, p. 2738–2743, 01 Jan; doi: 10.1016/j.matpr.2020.08.543.

https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.

https://cdas.cancer.gov/datasets/plco/19/#:~:text=The%20Breast%20dataset%20is%20a,cancer%20diagnosed%20in%20the%20trial

Michael E, Ma H, Li H, Qi S. 2022. An optimized framework for breast cancer classification using machine learning. Biomed Res Int. 2022:1–18. doi: 10.1155/2022/8482022.

Mirza Z, Ansari, M. S., Iqbal, M. S., Ahmad, N., Alganmi, N., Banjar, H. & Karim, S. 2023. Identification of novel diagnostic and prognostic gene signature biomarkers for breast cancer using artificial intelligence and machine learning assisted transcriptomics analysis. Cancers. [Online]. Available. 15(12):3237. https://www.mdpi.com/2072-6694/15/12/3237

Mittal S, Hasija Y. 2020. Applications of deep learning in healthcare and biomedicine. Deep Learn Tech Biomed Health Inf. 57–77.

Mohammed SA, Darrab S, Noaman SA, and Saake G. 2020. Analysis of breast cancer detection using different machine learning techniques, Data Mining and Big Data: 5th International Conference, DMBD 2020 Proceedings; Belgrade, Serbia, Jul 14–20; 5, Springer, p. 108–117.

Mojrian S et al. 2020. Hybrid machine learning Model of extreme learning 960 machine radial basis function for breast cancer detection and Diagnosis; a multilayer fuzzy expert system. 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Oct 14-15; p. 1–7. doi: 10.1109/RIVF48685.2020.9140744.

Naji MA, El Filali S, Aarika K, Benlahmar EH, Abdelouhahid RA, Debauche O. 2021. Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Comput Sci. 191:487–492. doi: 10.1016/j.procs.2021.07.062.

Nemade V, Fegade V. 2023. Machine Learning Techniques for Breast Cancer Prediction. Procedia Comput Sci. [2023/01/1] 218:1314–1320. doi:10.1016/j.procs.2023.01.110.

Ngiam KY, Khor IW. 2019. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 05/1/2019. 20(5): e262–e273. doi. 10.1016/S1470-2045(19)30149-4

Ngiam KY, Khor W. 2019. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 20(5):e262–e273. doi: 10.1016/S1470-2045(19)30149-4.

Ng A, Soo K. 2017. Numsense! Data science for the layman: no math added. (No Title).

Nomani A, Ansari Y, Nasirpour MH, Masoumian A, Pour ES, Valizadeh A. 2022. Psowns-CNN: a computational radiology for breast cancer diagnosis improvement based on image processing using machine learning methods. Comput Intell Neurosci. 2022:5667264. 2022/05/11. doi:10.1155/2022/5667264.

O H C D Panel NI. 1979. Special report. Treatment of primary breast cancer. N Engl J Med. 301:340.

Osareh A and Shadgar B. 2010. Machine learning techniques to diagnose breast cancer. 2010 5th International Symposium on Health Informatics and Bioinformatics, April 20-22; p. 114–120. doi: 10.1109/HIBIT.2010.5478895.

Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H. 2021. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Acces. 9:71194–71209. doi: 10.1109/ACCESS.2021.3079204.

Sarwal R, Prasad, U., Gopal, K. M., Kalal, S., Kaur, D., Kumar, A. & Sharma, J. 2021. Investment opportunities in India's healthcare sector.

Sha Z, Hu L, Rouyendegh BD. 2020. Deep learning and optimization algorithms for automatic breast cancer detection. Int J Imaging Syst Technol. 30(2):495–506. doi: 10.1002/ima.22400.

Shailaja K, Seetharamulu B, and Jabbar M. 2018. Machine learning in healthcare: a review. 2018 Second international conference on electronics, communication and aerospace technology (ICECA); IEEE. p. 910–914.

Shanbehzadeh M, Kazemi-Arpanahi H, Bolbolian Ghalibaf M, Orooji A. 01/1/2022 2022. Performance evaluation of machine learning for breast cancer diagnosis: a case study. Inf Med Unlocked. 31:101009. doi:10.1016/j.imu.2022.101009.

Shanbehzadeh M, Kazemi-Arpanahi H, Ghalibaf MB, Orooji A. 2022. Performance evaluation of machine learning for breast cancer diagnosis: a case study. Inf Med Unlocked. 31:101009. doi: 10.1016/j.imu.2022.101009 .

Sharma S, Aggarwal A, and Choudhury T. 2018. Breast cancer detection using machine learning algorithms. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS); Dec 21-22; p. 114–118. doi: 10.1109/CTEMS.2018.8769187.

Sharma GN, Dave R, Sanadya J, Sharma P, Sharma K. 2010. Various types and management of breast cancer: an overview. J Adv Pharm Technol Res. 1 (2):109.

Shickel B, Tighe PJ, Bihorac A, Rashidi P. 2017. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inf. 22(5):1589–1604.

Siddique S, Chow JC. 2021. Machine learning in healthcare communication. Encyclopedia. 1(1):220–239. doi: 10.3390/encyclopedia1010021.

Siddiqui SY, Haider A, Ghazal TM, Khan MA, Naseer I, Abbas S, Rahman M, Khan JA, Ahmad M, Hasan MK, et al. 2021. IoMT cloud-based intelligent prediction of breast cancer stages empowered with deep learning. IEEE Acces. 9:146478–146491. doi: 10.1109/ACCESS.2021.3123472.

Souri A, Ghafour MY, Ahmed AM, Safara F, Yamini A, Hoseyninezhad M. A new machine learning-based healthcare monitoring model for student's condition diagnosis in internet of things environment. Soft Comput. 2020. 2020/11/1. 24 22 :17111–17121. 10.1007/s00500-020-05003-6.

Swain S, Bhushan B, Dhiman G, Viriyasitavat W. 2022. Appositeness of optimized and reliable machine learning for healthcare: a survey. Arch Comput Methods Eng. 29(6):3981–4003. doi: 10.1007/s11831-022-09733-8.

Tabl AA, Alkhateeb A, ElMaraghy W, Rueda L, Ngom A. 2019. A machine learning approach for identifying gene biomarkers guiding the treatment of breast cancer. Front Genet. 10:256. doi: 10.3389/fgene.2019.00256.

Thornes R. 1967. Endogenous factors influencing host-tumor balance. Univ Chicago Press, Chicago. 255.

Usmani UA and Jaafar J. 2022. Machine learning in healthcare: Current trends and the future. International Conference on Artificial Intelligence for Smart Community: AISC 2020; Dec 17–18; Petronas, Malaysia: Universiti Teknologi, Springer. p. 659–675.

Vaka AR, Soni B, Reddy S. 2020. Breast cancer detection by leveraging machine learning. ICT Express. 6(4):320–324. doi: 10.1016/j.icte.2020.04.009.

Vrdoljak J, Boban Z, Barić D, Šegvić D, Kumrić M, Avirović M, Perić Balja M, Periša MM, Tomasović Č, Tomić S, et al. 2023. Applying explainable machine learning models for detection of breast cancer lymph node metastasis in patients eligible for neoadjuvant treatment. Cancers. 15 (3):634. doi:10.3390/cancers15030634.

Waks AG, Winer EP. 2019. Breast cancer treatment: a review. JAMA. 321 (3):288–300. doi: 10.1001/jama.2018.19323.

Wang Y, Acs, B., Robertson, S., Liu, B., Solorzano, L., Wählby, C. & Rantalainen, M. 2022. Improved breast cancer histological grading using deep learning. Ann Oncol. 33(1):89–98. [01/1/2022]. doi:10.1016/j.annonc.2021.09.007.

Wang X, Ahmad, I., Javeed, D., Zaidi, S. A., Alotaibi, F. M., Ghoneim, M. E. & Eldin, E. T. 2022. Intelligent hybrid deep learning Model for breast cancer detection. Electronics [Online]. Available. 11(17):2767. https://www.mdpi.com/2079-9292/11/17/2767

Weli ZNS. 2020. Data mining in cancer diagnosis and prediction: review about latest ten years. Curr J Appl Sci Technol. 39(6):11–32. doi: 10.9734/cjast/2020/v39i630555.

Wiens J, Shenoy ES. 2018. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. Clin Infect Dis. 66(1):149–153. doi: 10.1093/cid/cix731.

Wu J, Hicks C. 2021. Breast cancer type classification using machine learning. J Pers Med. 11(2):61. [Online]. Available. https://www.mdpi.com/2075-4426/11/2/61.

Yan Z, Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Zhang, S. & Zhou, X. S. 2016. Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition. IEEE Trans Med Imaging. 35(5):1332–1343. doi:10.1109/TMI.2016.2524985.

Yue W, Wang Z, Chen H, Payne A, Liu X. 2018. Machine learning with applications in breast cancer diagnosis and prognosis. Designs. 2(2):13. doi: 10.3390/designs2020013.

Zeebaree DQ, Haron H, Abdulazeez AM, and Zebari DA. 2019. Machine learning 1080 and region growing for breast cancer segmentation. 2019 International Conference on Advanced Science and Engineering (ICOASE); April 2-4; p. 88–93. doi: 10.1109/ICOASE.2019.8723832.

Zhang Y-D, Pan C, Chen X, Wang F. 2018. Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. J Comput Sci. [7/01/2018] 27:57–68. doi:10.1016/j.jocs.2018.05.005.

Zhang Y-D, Satapathy SC, Guttery DS, Górriz JM, Wang S-H. 2021. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. Inf Process Manag. 58 (2):102439. 3/01/2021. doi: 10.1016/j.ipm.2020.102439.

Zheng J, Lin D, Gao Z, Wang S, He M, Fan J. 2020. Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. IEEE Acces. 8:96946–96954. doi: 10.1109/ACCESS.2020.2993536.