BioMed Target Journal. 2024, 2(1):1-13
https://doi.org/10.59786/bmtj.211

Review Article

Efficient Machine Learning and Deep Learning Techniques for Detection of Breast Cancer Tumor

Dlshad Abdalrahman Mahmood [1,*], **Sadegh Abdullah Aminfar** [1]

[1] Department of Computer Science, Faculty of Science, Soran University, Soran, 44008, Iraq

*** Correspondence**
402238179007@soran.edu.iq

Article Info
Received: Feb 26, 2024
Revised: Apr 01, 2024
Accepted: Apr 07, 2024

Abstract

The detection of cancer tumors is an essential component that has important consequences for the speedy involvement of medical professionals and the enhancement of patient outcomes. This review paper presents a complete study of the current body of research and methodology, as well as an in-depth assessment of the use of machine learning (ML) and deep learning (DL) in the detection of cancer tumors. In addition, the article gives a full analysis of the approaches involved. Machine learning and deep learning, which effectively handle ambiguity in the identification of malignant tumors, provide an alternative method for dealing with the complexity of brain tissue. This method is offered by a combination of machine learning and deep learning. The first part of the review draws attention to the significance of making an accurate diagnosis of breast cancer, highlights the limits of traditional diagnostic methods, and investigates the cutting-edge area of medical imaging technology. After that, it investigates the fundamentals of ML and DL and how they might be used to deal with the challenges that are inherent in the interpretation of complicated imaging data. In addition, the paper explores the ways in which models enhance the processes of feature extraction, picture segmentation, and classification in breast tumor detection systems.

Keywords: Breast tumor, Machine learning, Deep learning, Diagnosis.

INTRODUCTION

Cancer is the second most common reason people die. In particular, the rate of breast cancer in women is much higher in poor countries than in rich ones. For instance, 1.38 million people with breast cancer are diagnosed in Pakistan each year, with one-third of them dying from the disease. 9.6 million people die from cancer worldwide [1,2]. An aberrant cell proliferation that infiltrates the surrounding tissues in the human body is linked to a breast cancer tumor. Tumors may be classified as benign or malignant. Non-cancerous cells that develop locally and do not spread across the body make up a benign tumor. On the other hand, a malignant tumor is made up of cancerous cells that have the ability to proliferate uncontrolled, spread to other areas of the body, and invade tissues [3]. Robust trials are necessary in medical image analysis studies to demonstrate the practicality of the suggested techniques. Nonetheless, studies are often conducted using data that the researchers have chosen, which may originate from various organizations, demographics, and scanners [4]. Over the last five years, the area of automated breast cancer diagnosis in digital mammography and digital breast tomosynthesis has been affected by the artificial intelligence (AI) revolution in computing, which is mostly due to deep learning and convolutional neural networks [5]. In a diagnostic context, deep learning algorithms are able to identify metastases in tissue slices of lymph nodes from women with breast cancer that have been stained with hematoxylin and eosin and compare them with the diagnosis of pathologists [6]. The community has a high frequency of breast cancer, which puts a

significant strain on resources. In addition to detecting breast cancer in its early stages and producing high-quality images with a lower radiation dosage, digital mammography also improves patient survival [7]. CNN models are favored in the medical domain over classical learning models because of their benefits in terms of speed and dependability [8]. The rapid diagnosis of breast cancer made possible by the availability of histopathological images qualifies them for use in computer-based image analysis and learning techniques. Reduced mortality rates may be achieved with early identification of cancer. By giving a thorough explanation of the most effective models and picture kinds, I want to enhance our capacity to aid in the identification of breast cancer in this review. It is clear from carefully reviewing a number of studies on computer-aided breast cancer detection that a number of techniques have been tried and tested and have shown promise.

MATERIAL and METHODS

Various medical imaging techniques, such as magnetic resonance imaging (MRI), mammography, ultrasound, and tomography, can be used to investigate the symptom [9]. This research employs three mammographic datasets—the digital database for screening mammography (DDSM), the curated breast imaging subset of DDSM, and the Mammographic Image Analysis Society—to train and evaluate specific approaches [10]. The subtypes of breast cancer were classified utilizing this method and 7,909 images of 82 individuals from the BreakHis database [11]. The findings were obtained from the publicly accessible Digital Database for Screening Mammography (DDSM) and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM). Acquiring a high degree of accuracy requires instruction on a vast quantity of data [12]. By utilizing Haralick patterns, color histograms, and Hu moments, it is possible to retrieve manually created features. The second approach utilizes transfer learning in conjunction with the pre-existing VGG16, VGG19, and ResNet50 networks to extract features and establish a baseline model. Following this, the standard classifiers are trained using the eliminated features [13]. An approach proposed utilizes convolutional neural networks (CNNs) for the purpose of categorizing images extracted from hematoxylin and eosin-stained breast biopsy samples. Four categories are applied to the images: benign lesion, in situ carcinoma, aggressive cancer, and normal tissue; the two categories are carcinoma and non-carcinoma [14].

MAIN BODY

Dataset

In order to locate breast cancer as quickly as possible and choose the appropriate model for dataset photos, several researchers debate and employ a variety of datasets to identify breast cancer in women. In reality, all researchers agree that early detection One of the factors that prevents more women from dying from cancer is early detection. The MIAS dataset is formatted in portable grey map (PGM) format and has a size of 1024 * 1024. The MIAS dataset comprises a total of 322 images, distributed across three groups. Specifically, there are 61 images representing the benign case, 52 images representing the malignant case, and 209 images representing the normal condition. The ground-truth information for the mammography images includes details such as the background tissue, the class of abnormality present, the type of tumor, the coordinates of the abnormality center, and the estimated radius for enclosing the abnormality circle[3]. The DDSM is a repository containing 2,620 digitized film mammography studies. The dataset includes cases with validated pathology information, encompassing normal, benign, and malignant instances. The DDSM is a valuable resource for developing and testing decision support systems due to its large database and reliable validation process. The CBIS-DDSM collection comprises a portion of the DDSM data that has been carefully chosen and organized by a skilled mammographer. The images have undergone decompression and conversion into DICOM format [10,15]. The BreakHis dataset, sometimes referred to as the Breast Cancer Histopathological Imaging Database (BreakHis), is acquired by completing an online form found on the vision laboratory's website. There are a total of 7909 samples, which have been obtained from 82 distinct patients [2]. A total of 544 whole slide images (WSI) were gathered from 80 breast cancer patients at the pathology

department of Colsanitas Colombia University in Bogotá, Colombia. The tumor tissue fragments were preserved in formalin and then encased in paraffin [16]. A total of 110 histopathology images of breast cancer were obtained from JUMC. These images are histological images of breast cancer that have been stained using the H&E staining procedure. The specimens were acquired via the Optika vision camera, which was affixed to a light microscope. The images were captured at various magnification levels (40X, 200X, and 400X) with a resolution of 2592 × 1936 [17]. The initial dataset comprises 162 WSI of breast cancer specimens. Each WSI was scanned at a resolution of 40x. From these WSI, a total of 277,524 patches, each measuring 50x50, were retrieved [18]. The BUSIS dataset comprises 562 images, with 306 images containing benign masses and 256 images containing malignant tumors (Xian et al., 2018). In the BUSI dataset, we utilized a subset of 630 images that specifically include instances of mass finds. Out of these, 421 images exhibit benign masses, while the remaining 209 images showcase malignant tumors [19]. Table 1 displays all of the datasets that were reviewed in this research, along with the name, quantity, and category of each picture. Figures 1, 2, and 3 show a few examples from several datasets related to breast cancer tumours.

Table 1. Summarize of exist datasets utilized in breast cancer detection.

Datasets	Number of images	Benign	Malignant	Normal
MIAS [3,15]	322	61	52	209
DDSM [10]	564	262	302	
CBIS-DDSM [10,15]	330			
BreakHis [2,11,13]	7909	2480	5429	
private dataset [16]	544			
Zenodo [17]	300			
JUMC [17]	110			
WSI [18]	277,524			
BUSIS (Xian et al., 2018)	562	306	256	
BUSI [19]	630	421	209	

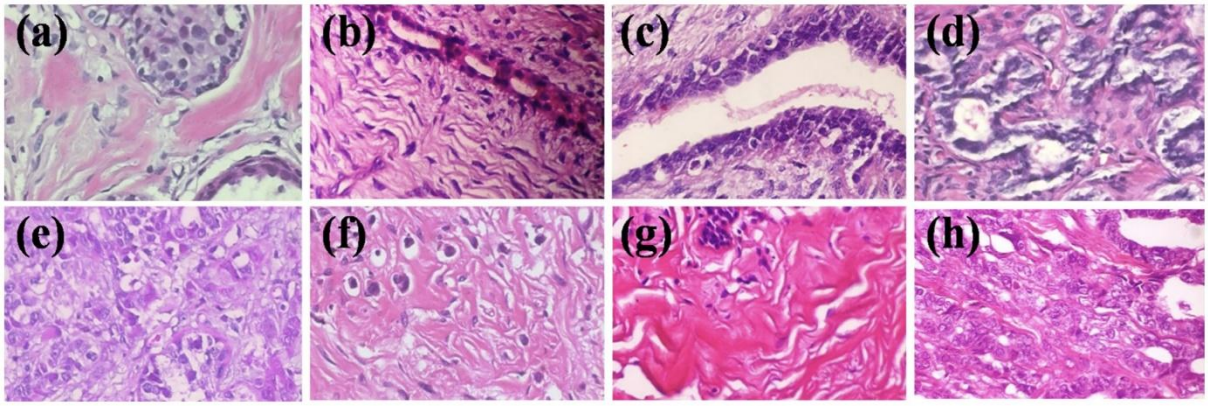

Figure 1. Samples of histopathological images from the BreakHis dataset for eight different types of breast cancer at ×200 magnification (a) adenosis, (b) fibroadenoma, (c) phyllods tumor, (d) tubular adenoma, (e) ductal carcinoma, (f) lobular carcinoma, (g) mucinous carcinoma, and (h) papillary carcinoma[13]

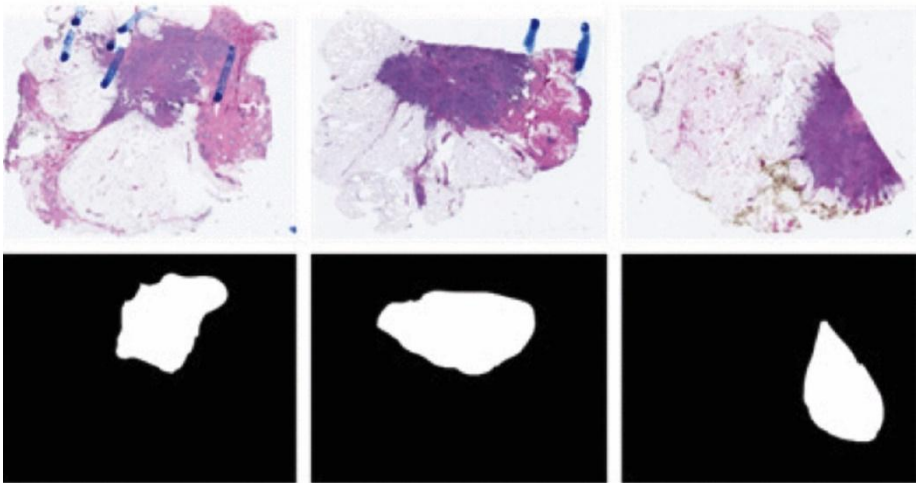

Figure 2. Sample images from dataset and corresponding masks [20].

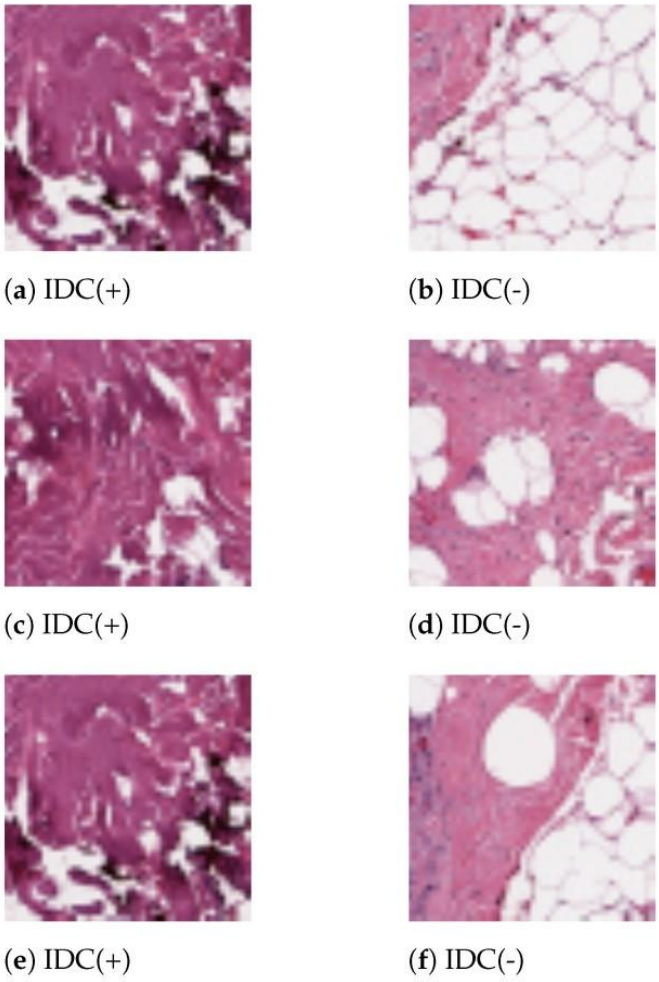

Figure 3. Some examples of the pictures in this dataset (a,c,e) IDC(+) patches. (b,d,f) IDC(-) patches[18].

Methods

CNN-based models

Every piece of research that has been updated for this study has made use of both transfer and deep learning. Some studies used a CNN structure that was already built in, while others used models that had already been trained to find breast cancer. Sharma and Mehra came up with a new way to find breast cancer by comparing and studying two machine learning algorithms. The objective is to classify a balanced BreakHis dataset using automated multi-classification dependent on magnification. To get handmade qualities out of the first way, Hu moments, color histograms, and Haralick patterns are used. In the second method, which uses transfer learning, networks that are already in place (VGG16, VGG19, and ResNet50) are used to pull out features and set a baseline. After being pulled, the features are used to teach the standard algorithms what to do. At all magnifications, the outcomes indicate that using pre-trained networks as feature extractors was more effective than the baseline method and the hand-made method. It has also been said that the enhancement is necessary to make the rating even more accurate [13]. In the same study, Saber et al., the features are taken from the MIAS dataset using a CNN design that has already been trained. Examples of CNN architectures are Inception V3, ResNet50, VGG-19, VGG-16, and Inception-V2 ResNet. For various datasets, Saber et al. suggested creating a new DL model based on the TL method. This would help in the automatic detection and diagnosis of the breast cancer (BC) suspected area using two methods: 80-20 and cross-validation. The way DL systems are designed makes them problem-specific. TL applies what they learned from one problem to another problem that is related [3].

Hameed et al., and Motlagh et al. used pretrained models which for the final sorting of pictures from the histology of breast cancer that show carcinoma and non-carcinoma, Hameed et al. used our collected dataset to show an ensemble deep learning method. We trained four different models using VGG16 and VGG19 designs that had already been trained. All of the models underwent initial five-fold cross-validation tests: completely trained VGG16, fine-tuned VGG16, fully trained VGG19, and fine-tuned VGG19. Then, employing an ensemble method that calculated the mean of the expected probabilities, we discovered that the combination of refined VGG16 and refined VGG19 produced classification results that were competitive, particularly in the cancer class [16]. Motlagh et al. also made use of the BreakHis dataset compiled by Shakra and Mehra. Implementing machine learning techniques may accelerate this procedure and prove more dependable and economical than conventional approaches. To show this idea, deep neural networks that have been learned and fine-tuned are used. Sort the different types of cancer using 6,402 training samples from tissue microarrays (TMAs) to test the method. After that, ResNetV152 told the difference between cancerous and noncancerous breast lumps. Also, ResNetV150 and ResNetV152 sorted tumors into two groups: those that were cancerous (ductal

carcinoma, lobular carcinoma, mucinous carcinoma, and cystic carcinoma) and those that were normal [11].

Aljuaid et al. used computer-aided identification and medical picture analysis in the same study. These methods may help teach and support less experienced medical workers, which may help to automate and speed up the process of finding and classifying cancer. CNNs are crucial for rapidly locating and categorizing cancer images within massive files of medical images. A novel computer-assisted screening method for classifying breast cancer into two or more categories is presented in this study. It uses DNNs (ResNet18, ShuffleNet, and InceptionV3) along with transfer learning on the BreakHis dataset that is open to the public [2].

Hirra et al. created a new patch-based deep learning method called Pa-DBN-BC. It uses the Deep Belief Network (DBN) to recognize and classify pictures of histopathology breast cancer. To get features, there is a pre-training phase that is not supervised and a fine-tuning phase that is monitored. The network instantly pulls out features from picture fixes. Using pictures of histopathology, logistic regression is used to put the spots into groups. The traits that were taken out of the patches are fed into the model, which then shows the result as a chance matrix as either a positive sample (cancer) or a negative sample (background). The whole slide histopathology image dataset is used to train and test the suggested model. It has pictures from four different groups of data [20]. Zewdie et al. suggest a multi-class classification method for breast cancer type, group, and grade that is built on deep learning. The system was trained and tested using histopathological images from the online databases "BreakHis" and "zendo," as well as from Jimma University Medical Center (JUMC). These images were taken with an Optikam PRO5 camera attached to an Optika microscope and set to four different magnification levels: 40x, 100x, 200x, and 400x. All of the pictures were fixed up and handled before they were sent to the ResNet50 model that had already been taught. The method that was made could divide breast cancer into a number of subtypes and two main types: normal and aggressive [17].

Khan et al. showed a new deep learning method for finding and classifying breast cancer in breast biopsy pictures. They used the idea of transfer learning to do this. Deep learning systems are usually built one at a time and are designed to be problem-specific. In contrast to traditional learning methods that grow and yield in isolation, the goal of transfer learning is to use the knowledge gained while solving one problem to solve another connected problem. To get information from pictures, the suggested system uses CNN designs that have already been trained, like GoogLeNet, ResNet, and VGGNet for Visual Geometry Group. Then, these traits are sent to a fully linked layer for average pooling classification. This is how cancerous and healthy cells are separated [21]. Zhang et al. proposed a novel machine-learning approach known as BI-RADS-Net-V2 for the automated detection of breast cancer in ultrasound images. The BI-RADS-Net-V2 can reliably tell the difference between normal and dangerous tumors and gives both numeric and semantic reasons. The reasons are given in terms of the Breast Imaging Reporting and Data System (BI-RADS), which is a physical feature that doctors use to diagnose and report mass findings (Zhang et al., 2023) that has been proven to work in the real world.

Zaalouk et al. suggested making a computer-aided diagnosis (CAD) system based on deep learning to help doctors do their jobs better. Five pre-trained convolutional neural network (CNN) models are looked at and judged in order to reach this goal. They are Xception, DenseNet201, InceptionResNetV2, VGG19, and ResNet152. It also talks about a new way to learn through sharing. Pictures of tissues from the BreakHis collection are used to teach and test these methods. Several tests are done to see how well these models work using binary and eight-class classes that depend on and don't depend on magnification [22]. CNN was used by Zhang et al. to make a lot of forecast models based on InceptionV3, VGG16, ResNet50, and VGG19. A total of 1007 pictures of the test group were used to test the prediction models. Of these, 788 were normal and 219 were cancerous. Following the construction of receiver operating characteristic curves, the critical area AUCs were determined.

Once the model with the highest AUC was chosen, its ability to make a diagnosis was compared to that of sonographers who did and analyzed ultrasonographic tests on 683 pictures from the reference group, with 493 being normal and 190 being cancerous [23]. The EDLCDS-BCDC approach, proposed by Ragab et al., aims to detect the presence of breast cancer using USIs. This technique involves two basic rounds of pre-processing for USIs: wiener filtering and contrast enhancement. In addition, the Chaotic Krill Herd Algorithm (CKHA) is utilized in conjunction with Kapur's entropy (KE) for the purpose of image segmentation. Furthermore, a combination of three sophisticated deep learning models, namely VGG-16, VGG-19, and SqueezeNet, is employed to extract features. The Cat Swarm Optimization (CSO) algorithm, in conjunction with the Multilayer Perceptron (MLP) model, is employed to accurately categorize images as either indicative of the presence or absence of breast cancer [24]. Jabeen et al. introduced a novel framework for classifying breast cancer based on ultrasound images. This framework utilizes deep learning techniques and combines the most optimal features. The process begins with the utilization of a pre-trained DarkNet-53 model. The output layer of this model is then adjusted to accommodate the augmented dataset classes. Subsequently, the modified model undergoes training through transfer learning. Features are extracted from the global average pooling layer of the model. These features are then subjected to two enhanced optimization algorithms, namely reformed differential evaluation (RDE) and reformed gray wolf (RGW), to select the most optimal ones. Finally, the selected features are combined using a novel probability-based serial approach and classified using machine learning algorithms [25]. Papandrianos et al. study the use of artificial intelligence approaches, namely deep-learning algorithms for medical image analysis, to tackle the significant issue of diagnosing bone metastasis in breast cancer patients [26]. Zahoor et al. want to explore strategies for disease prevention and develop novel categorization methods to mitigate the risk of breast cancer in women. Optimal feature optimization is conducted to accurately classify the results. The accuracy of the CAD system was enhanced by decreasing the occurrence of false-positive results.The Modified Entropy Whale Optimization Algorithm (MEWOA) is presented for the purpose of deep feature extraction and classification, utilizing fusion techniques [27]. Lin et al. introduced DeepMO, a model that utilizes deep neural networks and multi-omics data to categorize different subtypes of breast cancer. The Cancer Genome Atlas (TCGA) included three forms of omics data: mRNA data, DNA methylation data, and copy number variation (CNV) data [28]. Mobark et al. suggested a CNN model called CoroNet to automatically diagnose breast cancer using the CBIS-DDSM dataset. The Xception architecture, which has been pre-trained on the ImageNet dataset, is utilized. Additionally, it has undergone complete training on whole-image breast cancer using mammograms [29].

CNN and ML-based models

Several researchers have successfully used machine learning algorithms as classifiers to predict breast cancer diagnosis [30,31]. In this article, we will examine the latest advancements in the detection and diagnosis of breast cancer. Salma et al. and Ragab et al. utilized CNN with SVM techniques. Salma et al. described a new method based on VGG16. ResNet50 and VGG-16 are employed and retrained to identify two groups as opposed to one thousand, with minimal processing requirements and high precision. To fix the problem of not having enough tagged data, transfer learning and data addition are also used. The support vector machine (SVM) predictor, which works better, is used instead of the last fully linked layer. K-fold cross-validation is used to find out how well the model works. Three mammographic datasets—the edited breast imaging subset of DDSM, the digital database for screening mammography (DDSM), and the mammographic image analysis society—are utilized for method instruction and evaluation. The study talks about convolutional neural networks from beginning to end, without any preparation or postprocessing [10]. Ragab et al. developed the concept of a DCNN. AlexNet, GoogleNet, and ResNet are various forms of DCNN. The process: Utilizing DL techniques for feature extraction and classification, this research develops a novel CAD system to assist physicians in

the categorization of breast cancer patches observed on imaging. Four distinct evaluations are conducted in order to determine the most effective approach. The initial component comprises fully trained and optimized DCNN networks from beginning to conclusion. The deep features extracted by the DCNNs are transmitted to an SVM predictor with distinct kernel functions in the second phase. Combining deep features as demonstrated in the third experiment would increase the accuracy of the SVM models. PCA is ultimately implemented in the fourth trial to minimize the size and computation cost of the enormous feature vector generated during feature fusion. The digital mammogram database (MIAS) from the Mammographic Image Analysis Society and the edited breast image subset of the Digital Database for Screening Mammography (CBIS-DDSM) were utilized in the investigations [15].

An IDC Breast Cancer dataset containing 78,786 IDC positive images and 198,738 IDC negative images was utilized by Roy et al. to classify 277,524 images into IDC (+) and IDC (-) groups. This was accomplished using partially statistical and partially textural features, including Haralick texture features, SIFT, SURF, and ORB. These characteristics are subsequently compiled into a collection of 782 features. A collection of machine learning algorithms, including MultiLayer Perceptron, Random Forest, Extra Trees, XGBoost, AdaBoost, and CatBoost, are employed to classify a dataset comprising four characteristics. The Pearson association value is subsequently employed in feature selection [18]. Dou and Meng present a better optimization method (GSP_SVM) that combines GA, PSO, simulated annealing, and an SVM method. Based on the MCC, AUC, and other metrics, the findings show that a very high level of classification accuracy has been reached. When compared to current optimization methods, this approach may help doctors make better decisions about breast cancer secondary diagnoses, which would improve the diagnosing efficiency of hospitals [32]. Evaluating microscopic images stained with Human Epidermal Growth Factor Receptor 2 (HER2) is difficult, time-consuming, and prone to errors when done manually. The variation in staining, overlapping regions, and large, nonhomogeneous slides are the reasons behind this. Furthermore, in order to accurately analyze HER2 images, it is necessary to employ a method that involves identifying key characteristics that can effectively capture the challenging aspects of the images, such as the irregular cellular structure and the coloration of the tissue. Rashid and et al., proposed a trainable metaheuristic approach that utilizes a transfer learning model to select the optimal features. Furthermore, the proposed model is effective in minimizing model intricacy and computational expenses while also preventing overfitting. The proposed cascaded design consists of four main components: (1) converting whole slide images (WSIs) into tiled images and improving contrast using fast local Laplacian filtering (FlLpF); (2) extracting features using a ResNet50 CNN technique based on transfer learning; (3) selecting the most informative features using a non-dominated sorting genetic algorithm (NSGA-II) optimizer; and (4) classifying HER2 scores using a support vector machine (SVM) [33]. Ensemble-based Channel and Spatial Attention Network (ECS-A-Net) for the automatic classification of infected regions inside breast cancer (BC) images is proposed. The proposed framework comprises two phases. In the initial phase, various augmentation techniques are applied to expand the size of the input data. The second phase involves an ensemble technique that simultaneously utilizes modified SE-ResNet50 and InceptionV3 as a backbone for feature extraction. This is followed by the sequential application of Channel Attention (CA) and Spatial Attention (SA) modules to select more prominent features [34]. Humayun et al. propose a deep learning algorithm to largely predict breast cancer risk based on this foundation. The methodology relies on transfer learning, utilizing the InceptionResNetV2 deep learning model [35].

Mirimoghaddam et al. came up with a new way to solve the problem using guided deep learning. It is recommended to use a GAN-based model to find and sort HER2 levels and make high-quality shots of them. Transfer learning methods were used to judge the source and created pictures [36].

Shi et al. offered a useful and light mixed learning design that could identify and split breast

cancers at the same time. It adds a segmentation job to a cancer categorization network so that the backbone network can learn models that focus on tumor sites. They also come up with a new numerically stable loss function that easily controls the balance between how specific and sensitive the cancer detection is. A set of 1511 breast ultrasound pictures is used to test the suggested method [37].

Magnuska et al. proposed using a dataset of breast US images—252 instances of malignancy and 253 cases of benignity—to actualize and contrast several CAD assistance systems for lesion identification and classification. After preparing eight distinct datasets, which included pre-processed and spatially enhanced pictures, machine learning algorithms were taught to identify lesions. An analysis was conducted to compare the radionic signature (RS) derived from manually produced segments and the RS acquired from detection devices. YOLOv3, which underwent training using logarithmic derivatives of US images, demonstrates enhanced performance in the detection of breast lesions [38]. An automated data-driven model based on the YOLO algorithm is utilized to diagnose breast cancer in mammography. This model is designed to assist clinicians in making decisions during breast cancer screening or detection programs. The CBIS-DDSM and INbreast datasets, which are publicly accessible, were utilized as sources to apply the transfer learning technique to a private dataset of full-field digital mammography [39].

Table 2. A summarize of the results of published studies

Study	Method	Objective	Dataset	Accuracy	sensitivity	specificity	AUC
[11]	ResNet	Classified benign and malignant	BreakHis	99.8%	98.33%	98.3%	97.6%
[3]	VGG16	Classified benign and malignant	MIAS	97.83%	97.83%	99.13%	99.5%
[2]	ResNet18	Binary classification of benign or malignant	BreakHis	99.7%	97.53%	97.8%	-
	InceptionV3			97.66%	97.64%	97.59%	
	Net			96.94%	96.7%	96.85%	
	ShuffleNet	Multi-class classification of benign or malignant		97.81%	97.65%	97.31%	
				96.07%	96.03%	96%	-
				95.79%	95.7%	95.5%	
[22]	Xception	Computer-aided diagnosis (CAD)	BreakHis	98.99%	98.71%	-	-
[36]	InceptionResNetV2	A GAN-based model is proposed for generating high-quality HER2 images	HER2	94.2%	94.2%	-	95.26%
[25].	DarkNet53	Classification using DE and BGWO	BUSi	99.1%	99.1%	-	-

DISCUSSION

The results derived from all the studies that were reviewed and employed ML and DL techniques are discussed and analyzed in this section.

CNN-based models

It is required that the findings of articles published on the subject be the subject of discussion and analysis. Motlagh et al. identified that employing the deep learning ResNet technique with specific parameters for cancer diagnosis is a dependable and effective approach, as opposed to conventional methods. The main objective of this study was to examine how the suggested frameworks could be applied to the identification of cancer subtypes. Various classifications by deep learning implemented in ResNet and Inception are demonstrated. The ResNet frameworks exhibit accuracy rates of 99.8%, 98.7%, 94.8%, and 96.4%, respectively, when applied to four distinct categories of cancer, two primary types of breast cancer, sub-types linked to malignant and benign cancer, and average false positive values that are negligible. The limited quantity of histological cancer images in relation to the number of model parameters could potentially impact the effectiveness and dependability of the deep learning models [11]. Using the idea of transfer learning, Khan et al. presented a unique deep learning framework for the identification and categorization of breast cancer. Three distinct CNN architectures are used to extract features from breast cytology pictures, and

the idea of transfer learning is then merged to increase classification accuracy. Lastly, the effectiveness of the suggested framework is evaluated both independently against various CNN designs and against other approaches already in use. The average classification accuracy of the (Google, VGGNet, and ResNet) architectures is 93.5 percent, 94.15%, and 94.35%, respectively [21]. The CNN-based breast cancer prediction algorithm created by Zhang et al. showed excellent accuracy in breast ultrasound imaging. The resulting InceptionV3, VGG16, ResNet50, and VGG19 models had accuracy scores of 0.905, 0.866, 0.851, and 0.847, in that order. The absence of standardized image acquisition protocols may have influenced the study findings, as the breast ultrasound images were obtained from a prior investigation. [23]. Hameed et al. demonstrated an ensemble deep-learning approach to classify images of breast cancer tissue into distinct categories by utilizing the dataset that we had gathered. An ensemble consisting of enhanced VGG16 and VGG19 models produces a more dependable model. The proposed group method demonstrates competitive performance when it comes to classifying complex histopathological images associated with breast cancer. The ensemble technique was employed to determine that the fine-tuned VGG16 and VGG19 ensemble exhibited competitive performance in classification, specifically in the carcinoma class, by calculating the mean of the projected probabilities. For the carcinoma class, the combined set of optimized VGG16 and VGG19 models had an overall accuracy of 95.29% and a sensitivity of 97.73%. The collection is constrained, comprising only 845 photos, which may not comprehensively depict the range and fluctuation of breast cancer histopathological scans [16].

Saber et al. suggested a unique deep learning model to improve the classification results on the MIAS dataset. This approach is intended to assist physicians in the identification and diagnosis of BC. Three classifications were created from the MIAS images: benign, malignant, and normal. The original MIAS dataset underwent pre-processing to identify the malignant area, remove non-breast regions, improve contrast in breast pictures, and remove noise. An idea of adding more data to the CNN system was also put forward to make it work better. To do this, the information needs to be made bigger. When the freeze and fine-tuning steps were used on the above dataset, the mass-lesion classification accuracy was improved. In comparison to four other models, the VGG16 model produced the best results in terms of accuracy, sensitivity, specificity, AUC, and F-score. In conclusion, compared to other current methods, a definite improvement may be obtained by incorporating CNN, which employs learning transfer in the screening process. 98.96% accuracy, 97.83% sensitivity, 99.13% specificity, 97.35% precision, 97.66% F-score, and 0.995 AUC were shown in the findings. These outcomes surpass those of the previously discussed techniques. The study solely assesses the effectiveness of the proposed model on the MIAS dataset, thereby restricting the applicability of the findings to different datasets or real-life situations [3]. A Deep Belief Network (DBN) model is introduced by Hirra et al. as a means of classifying images of breast cancer according to histopathology. The proposed method utilizes picture fragments of uniform dimensions to facilitate comprehension and identification. Using the entire slide histopathology image dataset, the proposed model was trained and evaluated, and it achieved 86% accuracy. It contained images from four distinct data categories [20]. Zewdie et al. demonstrated a deep learning-based approach for multi-class breast cancer classification and grade-level detection of IDC from tissue images. Utilizing a learned ResNet50 model, the features were generated. The aforementioned data set was employed to train a soft-max classifier capable of classifying whole slide histopathological images of breast cancer into four distinct subtypes: fibroadenoma tumor, adenosis tumor, and phyllodes tumor. Diffusive carcinoma comprised four additional cancer varieties, namely cystic carcinoma, ductal carcinoma, lobular carcinoma, and mucinous carcinoma. The suggested method achieves 93.78% accuracy for grade recognition, 95.78% accuracy for benign sub-type classification, 96.75% accuracy for cancer sub-type classification, and 93.86% accuracy for binary classification, as determined by the test results. The research does not

address the potential constraints or difficulties linked to the application of the suggested approach in a clinical environment, such as the requirement for specific equipment, technical proficiency, and infrastructure [17].

Aljuaid et al. showed how deep neural networks with transfer learning can be used to help computers make diagnoses of breast cancer by sorting pictures into groups. BrakeHis made 7909 pictures of people with breast cancer public. These pictures came from the diagnoses of 82 different people. Not only were data enhancement methods used to make the sorting process better, but also a number of picture illumination factors were looked at. Three DNNs were used with the imaging-based approach to sort pictures of breast cancer into groups. Three fifths of the photos were used for testing, and sixty-five percent were used for training. A variety of metrics were employed to assess performance, including accuracy, precision, sensitivity, and specificity. ResNet is the most precise and efficient classifier due to its ability to attain mean accuracies ranging from 97.81% to 99.70% for both binary and multi-class classification. The research lacks a comprehensive examination of the computational resources necessary for executing the suggested approach, which may hinder its practical implementation in real-world scenarios [2]. Zaalouk et al. used the BreakHis dataset to test how well the Xception, DenseNet201, Inception ResNet V2, VGG19, and ResNet152 models worked after being trained. The main goal of this study was to create a reliable deep learning model that would help pathologists diagnose breast cancers in histopathological images of any type of tumor and at any level of magnification. To reach this goal, many tests were done on each of these models to look at them from every possible breast cancer classification point. So, four types of classifications were done on all the pre-trained models at two different learning rates: binary classification that doesn't depend on magnification, eight-class classification that doesn't depend on magnification, and binary classification that does depend on magnification. Xception did very well in these tests, which were done at two different learning rates for all models that had already been taught. The Xception model has had the highest rating accuracy across all tests, which shows that it has a lot of promise. For trials that don't need magnification, its accuracy ranges from 93.32% to 98.99%. For trials that do need magnification, it ranges from 90.22% to 100% [22]. Shi et al. identified the EMT-Net as a portable and effective method for diagnosing breast cancer. When the network's model format was converted to Tensorflow Lite, the file size increased to 20MB. This has the potential to enable rapid breast cancer diagnosis via mobile devices. The proposed network exhibits a significantly higher accuracy in tumor classification compared to a single-task network due to its dual-task architecture comprising tumor segmentation and tumor classification as its secondary function. A novel concept, numerically stable weighted cross-entropy loss could be implemented to achieve a balance between specificity and sensitivity in breast cancer detection. One can readily increase the sensitivity of the model by increasing the positive weight [37]. Jabeen et al. conducted an experiment on an upgraded Breast Ultrasound Images (BUSI) dataset, using transfer learning and an optimizer. The experiment resulted in a best accuracy of 99.1% [25]. The Humayun et al. conducted experimental research on a breast cancer dataset, which showed a high level of model performance, achieving an accuracy rate of 91%. The model incorporates risk markers to enhance breast cancer risk assessment scores and demonstrates encouraging outcomes in comparison to current methodologies. Deep learning algorithms incorporate risk markers to enhance accuracy ratings [35].

CNN and ML-based models

A range of visual characteristics were noted for their influence and contribution by Roy et al. In order to determine the most crucial collection of characteristics that would provide the best accuracy, researchers looked at statistical features obtained from the GLCM as well as picture texture features, such as SURF, SIFT, and ORB. After examining the relationships between each classifier model's unique predicting abilities, an ensemble of several classifier models was created. Ultimately, the CB classifier that was trained using the updated feature set had a 92.55% accuracy rate. The research primarily

examines the utilization of conventional machine learning classifiers and ensembling approaches, but does not investigate the possibilities of deep learning architectures, which have demonstrated encouraging outcomes in the identification of breast cancer [18]. The results obtained from the HER2SC and HER2GAN datasets demonstrate that the NSGA-II-DL model outperforms existing techniques, achieving an accuracy of 94.4%, precision of 93.71%, specificity of 98.07%, sensitivity of 93.83%, and an F1-score of 93.71% for the HER2SC dataset [33]. The ECS-A-Net model conducted thorough trials comparing several state-of-the-art methodologies on two benchmarks, DDSM and MIAS. The model obtained an accuracy of 96.50% on the DDSM dataset and 95.33% on the MIAS dataset [34].

A study by Magnuska et al. carefully looked at the many steps a CAD system should take to find and label normal or cancerous breast tumors in US pictures. First, it was shown that when making programs to find breast lesions, it's best to use spatial changes and picture pre-processing to add to the data. Second, suggest putting LE and IoU together to make it easier to find breast lesions in US pictures. Third, the YOLOv3 technique shows breast tumors more reliably and consistently than the Viola–Jones-based method. Fourth, the job of segmentation can be skipped, and the true RS for classifying breast cancer can be found just by looking at the detecting bounding boxes [38].

Zhang et al. made BI-RADS-Net-V2, a deep network-based breast ultrasound detection system that can be interpreted. With the addition of medical information to BI-RADS, this method gives people reliable and useful analysis. The automatic monitoring system may spread more easily because it is more likely to earn the trust of end users and be very accurate. It helps the idea of early, uniform screening for breast cancer grow. The experiment results show that adding BI-RADS may improve the diagnostic model's ability to generalize and make it more accurate in a learning framework with multiple tasks. Systematic BI-RADS descriptor forecast is another useful tool for showing how accurate the discriminative model is. The information distillation-based numeric answer, on the other hand, might look into why discriminative models go wrong and offer ways to make the models work better [40].

Despite the absence of HER2 data, Mirimoghaddam et al. successfully generated highly accurate fabricated images utilizing the CGAN model. In addition, by employing transfer learning models on both simulated and actual data, we improved the systems' ability to determine the HER2 score by more than 94%. The models are also capable of discerning tendencies in the counterfeit data with precision. Each image now undergoes model testing in less than one second, a significant improvement over the previous method [36].

The exclusive dataset is a genuine and diverse case study, comprising 190 masses, 46 asymmetries, and 71 distortions. A comparison was conducted between various Yolo architectures, including YoloV3, YoloV5, and YoloV5-Transformer. Furthermore, Eigen-CAM was utilized to provide model introspection and generate explanations by emphasizing all the questionable locations of interest inside the mammography. The compact YoloV5 model achieved the highest level of performance, with a mAP of 0.621, when applied to our exclusive dataset [39].

CONCLUSION

In conclusion, this research provides an in-depth analysis of the utilization of machine learning (ML) and deep learning (DL) techniques in the diagnosis of breast cancer, highlighting their considerable potential in advancing the field of breast cancer detection. Based on the analysis of the study data, it is evident that machine learning (ML) and deep learning (DL) possess the capacity to be valuable in the treatment of cancer tumors. However, there are some obstacles that need to be addressed. A range of models, including ShuffleNet, ResNet, VGG16, ResNet18, InceptionV3Net, and Xception, were utilized by researchers to identify photographs extracted from databases. The objective of the study was to identify breast cancer in women during its initial stages. Evaluates the efficacy of their techniques across many domains, encompassing pre-processing, breast tumor segmentation, feature extraction, and enhancing image legibility. Integrating models with techniques

for identifying breast tumors is a notable advancement in the creation of diagnostic systems that are more precise, adaptable, and responsive to specific circumstances. The progress of breast cancer research relies on the continuous clinical application of CAD, which will improve patient outcomes and facilitate the development of state-of-the-art methods for breast tumor detection.

Conflict of Interest

The authors declare they have no conflicting interests.

REFERENCES

1. Ahmad S, Ur Rehman S, Iqbal A, Farooq RK, Shahid A, Ullah MI. Breast Cancer Research in Pakistan: A Bibliometric Analysis. *SAGE Open*. 2021;11(3)doi:10.1177/21582440211046934
2. Aljuaid H, Alturki N, Alsubaie N, Cavallaro L, Liotta A. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Comput Methods Programs Biomed*. Aug 2022;223:106951. doi:10.1016/j.cmpb.2022.106951
3. Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H. A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. *IEEE Access*. 2021;9:71194-71209. doi:10.1109/access.2021.3079204
4. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans Biomed Eng*. Jul 2016;63(7):1455-62. doi:10.1109/TBME.2015.2496264
5. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin Cancer Biol*. Jul 2021;72:214-225. doi:10.1016/j.semcancer.2020.06.002
6. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. Dec 12 2017;318(22):2199-2210. doi:10.1001/jama.2017.14585
7. Zeeshan M, Salam B, Khalid QSB, Alam S, Sayani R. Diagnostic Accuracy of Digital Mammography in the Detection of Breast Cancer. *Cureus*. Apr 8 2018;10(4):e2448. doi:10.7759/cureus.2448
8. Shahidi F, Mohd Daud S, Abas H, Ahmad NA, Maarop N. Breast Cancer Classification Using Deep Learning Approaches and Histopathology Image: A Comparison Study. *IEEE Access*. 2020;8:187531-187552. doi:10.1109/access.2020.3029881
9. Aristokli N, Polycarpou I, Themistocleous SC, Sophocleous D, Mamais I. Comparison of the diagnostic performance of Magnetic Resonance Imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: A review. *Radiography (Lond)*. Aug 2022;28(3):848-856. doi:10.1016/j.radi.2022.01.006
10. Salama WM, Elbagoury AM, Aly MH. Novel breast cancer classification framework based on deep learning. *IET Image Processing*. 2020;14(13):3254-3259. doi:10.1049/iet-ipr.2020.0122
11. Motlagh MH, Jannesari M, Aboulkheyr H, et al. 2018;doi:10.1101/242818
12. Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*. 2019;7:e6201. doi:10.7717/peerj.6201
13. Sharma S, Mehra R. Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images-a Comparative Insight. *J Digit Imaging*. Jun 2020;33(3):632-654. doi:10.1007/s10278-019-00307-y
14. Araujo T, Aresta G, Castro E, et al. Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS One*. 2017;12(6):e0177544. doi:10.1371/journal.pone.0177544
15. Ragab DA, Attallah O, Sharkas M, Ren J, Marshall S. A framework for breast cancer classification using Multi-DCNNs. *Comput Biol Med*. Apr 2021;131:104245. doi:10.1016/j.compbiomed.2021.104245
16. Hameed Z, Zahia S, Garcia-Zapirain B, Javier Aguirre J, Maria Vanegas A. Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models. *Sensors (Basel)*. Aug 5 2020;20(16)doi:10.3390/s20164373
17. Zewdie ET, Tessema AW, Simegn GL. Classification of breast cancer types, sub-types and grade from histopathological images using deep learning technique. *Health and Technology*. 2021;11(6):1277-1290. doi:10.1007/s12553-021-00592-0
18. Roy SD, Das S, Kar D, Schwenker F, Sarkar R. Computer Aided Breast Cancer Detection Using Ensembling of Texture and Statistical Image Features. *Sensors (Basel)*. May 23 2021;21(11)doi:10.3390/s21113628
19. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. *Data Brief*. Feb 2020;28:104863. doi:10.1016/j.dib.2019.104863
20. Hirra I, Ahmad M, Hussain A, et al. Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling. *IEEE Access*. 2021;9:24273-24287. doi:10.1109/access.2021.3056516
21. Khan S, Islam N, Jan Z, Ud Din I, Rodrigues JJPC. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*. 2019;125:1-6. doi:10.1016/j.patrec.2019.03.022
22. Zaalouk AM, Ebrahim GA, Mohamed HK, Hassan HM, Zaalouk MMA. A Deep Learning Computer-Aided Diagnosis Approach for Breast Cancer. *Bioengineering (Basel)*. Aug 15 2022;9(8)doi:10.3390/bioengineering9080391
23. Zhang H, Han L, Chen K, Peng Y, Lin J. Diagnostic Efficiency of the Breast Ultrasound Computer-Aided Prediction Model Based on Convolutional Neural Network in Breast Cancer. *J Digit Imaging*. Oct 2020;33(5):1218-1223. doi:10.1007/s10278-020-00357-7
24. Ragab M, Albukhari A, Alyami J, Mansour RF. Ensemble Deep-Learning-Enabled Clinical Decision Support System for Breast Cancer Diagnosis and Classification on Ultrasound Images. *Biology (Basel)*. Mar 14 2022;11(3)doi:10.3390/biology11030439
25. Jabeen K, Khan MA, Alhaisoni M, et al. Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors (Basel)*. Jan 21 2022;22(3)doi:10.3390/s22030807

26. Nikolaos Papandrianos EP, Athanasios Anagnostis, andAnna Feleki <Papandrianos 2020.pdf>. 2020;doi:10.3390/app10030997
27. Zahoor S, Shoaib U, Lali IU. Breast Cancer Mammograms Classification Using Deep Neural Network and Entropy-Controlled Whale Optimization Algorithm. *Diagnostics (Basel)*. Feb 21 2022;12(2)doi:10.3390/diagnostics12020557
28. Lin Y, Zhang W, Cao H, Li G, Du W. Classifying Breast Cancer Subtypes Using Deep Neural Networks Based on Multi-Omics Data. *Genes (Basel)*. Aug 4 2020;11(8)doi:10.3390/genes11080888
29. Mobark N, Hamad S, Rida SZ. CoroNet: Deep Neural Network-Based End-to-End Training for Breast Cancer Diagnosis. *Applied Sciences*. 2022;12(14)doi:10.3390/app12147080
30. Jader R, Aminifar S, Ejbali R. Predictive Model for Diagnosis of Gestational Diabetes in the Kurdistan Region by a Combination of Clustering and Classification Algorithms: An Ensemble Approach. *Applied Computational Intelligence and Soft Computing*. 2022;2022:1-11. doi:10.1155/2022/9749579
31. Rasool J, Sadegh A. An Intelligent Gestational Diabetes Mellitus Recognition System Using Machine Learning Algorithms. *Tikrit Journal of Pure Science*. 2023;28(1):82-88. doi:10.25130/tjps.v28i1.1269
32. Dou Y, Meng W. An Optimization Algorithm for Computer-Aided Diagnosis of Breast Cancer Based on Support Vector Machine. *Front Bioeng Biotechnol*. 2021;9:698390. doi:10.3389/fbioe.2021.698390
33. Rashid TA, Majidpour J, Thinakaran R, et al. NSGA-II-DL: Metaheuristic Optimal Feature Selection With Deep Learning Framework for HER2 Classification in Breast Cancer. *IEEE Access*. 2024;12:38885-38898. doi:10.1109/access.2024.3374890
34. Thwin SM, Malebary SJ, Abulfaraj AW, Park H-S. Attention-Based Ensemble Network for Effective Breast Cancer Classification over Benchmarks. *Technologies*. 2024;12(2)doi:10.3390/technologies12020016
35. Humayun M, Khalil MI, Almuayqil SN, Jhanjhi NZ. Framework for Detecting Breast Cancer Risk Presence Using Deep Learning. *Electronics*. 2023;12(2)doi:10.3390/electronics12020403
36. Mirimoghaddam MM, Majidpour J, Pashaei F, et al. HER2GAN: Overcome the Scarcity of HER2 Breast Cancer Dataset Based on Transfer Learning and GAN Model. *Clin Breast Cancer*. Jan 2024;24(1):53-64. doi:10.1016/j.clbc.2023.09.014
37. Shi J, Vakanski A, Xian M, Ding J, Ning C. Emt-Net: Efficient Multitask Network for Computer-Aided Diagnosis of Breast Cancer. *Proc IEEE Int Symp Biomed Imaging*. Mar 2022;2022doi:10.1109/isbi52829.2022.9761438
38. Magnuska ZA, Theek B, Darguzyte M, et al. Influence of the Computer-Aided Decision Support System Design on Ultrasound-Based Breast Cancer Classification. *Cancers (Basel)*. Jan 6 2022;14(2)doi:10.3390/cancers14020277
39. Prinzi F, Insalaco M, Orlando A, Gaglio S, Vitabile S. A Yolo-Based Model for Breast Cancer Detection in Mammograms. *Cognitive Computation*. 2023;16(1):107-120. doi:10.1007/s12559-023-10189-6
40. Zhang B, Vakanski A, Xian M. BI-RADS-NET-V2: A Composite Multi-Task Neural Network for Computer-Aided Diagnosis of Breast Cancer in Ultrasound Images With Semantic and Quantitative Explanations. *IEEE Access*. 2023;11:79480-79494. doi:10.1109/access.2023.3298569