NTU Journal of Engineering and Technology (2023) 2 (1) : 46-54
DOI: https://doi.org/10.56286/ntujet.v2i1

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

A survey: Breast Cancer Classification by Using Machine Learning Techniques

1st Ruaa Hassan Mohammed Ameen [1], 2nd Nasseer Moyasser Basheer[2], 3rd Ahmed Khazal Younis[3]
1. Master student at – Northern Technical University, Technical Engineering College of Mosul, Department of Medical Instrumentation Technology Engineering, Mosul, Iraq, ruaa.hassan1@ntu.edu.iq
2. Northern Technical University, Technical Engineering College of Mosul, Department of Medical Instrumentation Technology Engineering ,Mosul, Iraq, nmbasheer@ntu.edu.iq 3.Northern Technical University, Technical Engineering College of Mosul, Department of the Computer Techniques Engineering, Mosul, Iraq, ahmedkhazal@ntu.edu.iq .

Article Informations

Received: 26-09- 2022,
Accepted: 14-04- 2023,
Published online: 09-05-2023

Corresponding author:
Name: Ruaa H. Mohammed Ameen
Affiliation : Technical Engineering College of Mosul
Email: ruaa.hassan1@ntu.edu.iq

Key Words:
Breast cancer,
Artificial intelligence,
Machine learning,
Classification,
CAD systems.

ABSTRACT

Breast cancer in general is a common and a deadly disease. Early detection can significantly reduce the chances of death. Using automated feature extraction and classification algorithms, physician's experience in diagnosing and detecting breast cancer can be aided. This paper focuses on various statistical and machine learning(ML) studies of mammography datasets for enhancing the accuracy of breast cancer diagnosis and classification based on various variables. The Naïve Bayes,the K-nearest neighbors (KNN),the Support Vector Machine (SVM),the Random Forest (RF),the Logistic Regression(LR), Multilayer Perceptron (MLP), fuzzy classifier, and Convolutional Neural Network (CNN) classifiers, are the most widely used technologies in this field. This study provides an overview of the existing Computer-aided diagnosis/detection (CAD) systems based on artificial intelligence(AI) classification techniques and many types of medical image modalities. Potential research initiatives to build more efficient and accurate CAD systems have been investigated.

Introduction

This Word document can be used as a template for papers to be published in NTU-JET. Follow the text and explore the post notes for more formatting information.

Breast cancer is one of the leading causes of death for women all over the world. The dense population and patients' misunderstanding of illness signs and late getting medical counsel lead to greater death-rate. In addition, a lack of medical professionals and experts in rural regions exacerbates the difficulty of accurate early diagnosis. One answer to breast cancer early detection is to use information technology and medical data to construct medical support systems capable of assisting the doctor's reasoning and recognizing the cases. This increases the chances of healing and decreases the mortality rate [1], [2].

The most efficient way of diagnosing breast cancer is by imaging inspection. For diagnosis, many medical imaging methods like digital mammography (DM), the magnetic resonance imaging(MRI), ultrasound(US), and histopathology images. These are used to assist radiologists and clinicians in diagnosing the illness. Because image interpretation is operator-dependent and requires knowledge, the use of information technology is required, which can expedite and offer another opinion to the experts to improve the accuracy of the diagnosis [3]. Using techniques for feature extraction and classification: Computer-aided diagnosis/detection (CAD) is a useful tool for detecting abnormalities by physicians and professionals [4].

Many efforts have been exploited to build CAD systems that are dependent on pattern recognition, image processing, and artificial intelligence developments. Utilizing the CAD systems leads to reduce operator dependence, boosts diagnostic rates, and lowers the cost of medical auxiliary modalities [5], [6].

As a result, it might assist in reducing false-positive reactions; this may result in ineffective therapy as well as physical, psychological, and financial costs. It might help also to prevent false negative results, which could lead to treatment omissions [7].

The sensitivity with using CAD is increased about 10% rather than without CAD [8].

In general, the entire CAD system featured segmented constructions, irregularities detection, and feature extraction to further classify problems. As a result, CAD systems can be divided into four initial stages.

• Preprocessing is the first stage, which prepares the images for further phases. One of the most important tasks is to remove the noise that may be encountered during the mammogram [9].

• Segmentation is the second stage which is related to the region of interest (ROI) in the image, this is a technique for splitting the image to different sections based on visual qualities [10].

• Extraction and selection of features is the third stage. After retrieving features from preprocessed images, the most distinguishing characteristics are chosen. To minimize classification errors, the chosen characteristics must be enough for distinguishing between benign tumors (not cancer) and malignant tumors (cancer) areas. Despite all the efforts, there's no consensus about the traits that are most suited for this purpose. Many types of characteristics, like dynamic features, morphological features, and textural features, have historically been used to classify tumors [11].

• Classification is the last stage of the CAD system and it is described as the core of the system.

The CAD system is a data mining method that categorizes data. Many organizations goal is to uncover and extract features from great datasets using various Machine Learning Techniques (MLTs) [5].

The models that result are used to forecast unknown cases. Lots of Machine learning algorithms (MLA), like K-nearest neighbors (K-NN) [11], Artificial Neural Network (ANN) [12], Decision Tree (DT) [13],[14], have been applied in the medical area. As well as the Support Vector Machine (SVM) [15], [16]. The choice of suitable MLA to design a classifier responsible for differentiating different types of breast lesions is a critical component of CAD system development [17].

2. METHODOLOGY:

This section of the survey includes five parts: the first part is search criteria that clarify the goal of this review and cover the questions related to CAD systems. The second part presents Medical imaging modalities associated with breast cancer diagnosis. The third part shows the types of the dataset for digital mammogram images. The fourth part clarifies the process of image preprocessing. The fifth part presents the Feature extraction and its importance. The sixth part introduces the different types of classification techniques used in CAD systems.

2.1. Search criteria:

This study attempts to discuss many researches based on medical images and MLAs associated with breast tumor classification systems. Several aspects of importance in the development of CAD systems for breast cancer are addressed in this survey: medical imaging methods used in the CAD systems for breast tumors, Data sets used in CAD systems development, Preprocessing methods used in CAD systems, Segmentation methods used in CAD systems, Feature extraction methods used in CAD systems and Machine learning classifiers are being used in breast tumor medical imaging-based CAD systems.

2.2. Medical imaging modalities:

Various imaging modalities are currently applied for breast cancer detection, Fig.1 represent a Pie chart depicting the many modalities utilized in various CAD systems.

The imaging modalities studied in this research are as follows:

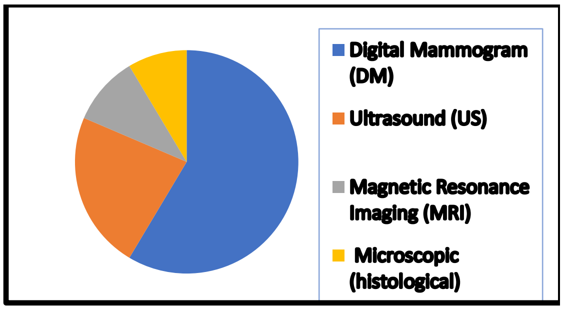

Figure1: Pie chart depicting the many modalities utilized in various CAD systems.

• Digital Mammogram (DM) is a unique kind of X-ray imaging used to make breast image with details of the breast. A mammogram is a transmission planar X-ray image made by a diverging X-ray beam. Thus, the breast volume attenuation is represented by light and dark shadows captured in a film combination process. At the result the produced image is a planar projection of the three-dimensional breast [18].

In clinical practice, (DM) is the most widely and extensively implemented screening tool. It is capable of identifying malignancies before they grow to the point where they may be easily identified and detected by a doctor. Although the DM has limitations, such as being an ineffective screening tool for women with thick breasts due to the use of ionizing radiation, X-ray mammography is still the gold standard for breast cancer screening that is widely used and has a high 2D resolution. As a result, DM may detect extremely minute differences in tissue composition as micro-calcifications [19], [20]. Fig. 2 presents examples of DM images.

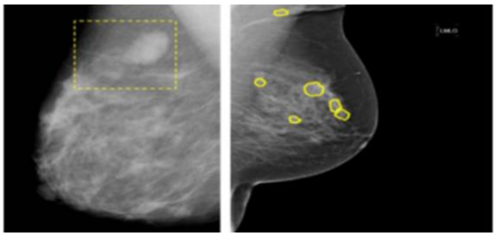

Figure 2**:** examples of DM images [19].

•Medical Ultrasound (US) is a real-time imaging technique that uses ultrasound to visualize muscles, tendons, and many internal organs in order to capture their size, structure, and any pathological lesions. For at least 50 years, radiologists and sonographers have used (US) to image the human body, and it has become a common diagnostic tool. For women with thick breasts, (US) is a handy mode of cancer detection. Furthermore, it can be used to detect tumors when mammography results are negative. US assess tumor size and can define anomalies found by DM [20], [21]. Figure.3 presents example of US image.

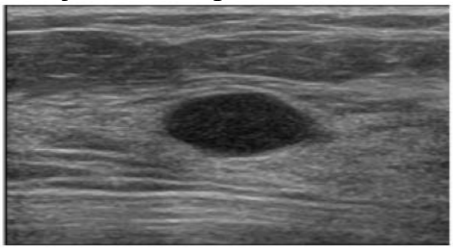

Figure3**:** example of US image[20].

• Magnetic Resonance Imaging (MRI) is a special technique for creating medical images based on a magnetic field and computer-generated radio waves to create detailed images of the organs and tissues; MRI images the entire breast and delivers it as thin slices covering the total breast volume. It has a lot of potential for screening high-risk cases and assessing the efficacy of treatment [22]. Figure4 shows examples of MRI images.

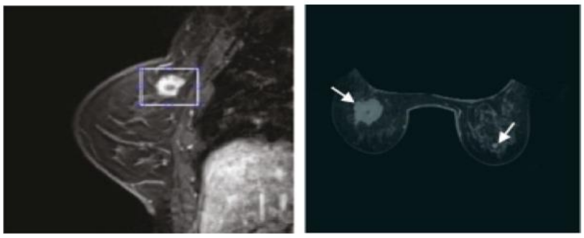

Figure 4: examples of MRI images [23].

•Histopathology is the microscopic examination of disease symptoms obtained from a biopsy or surgical specimen and fixed on glass slides. The pieces are stained to help detect tumors by allowing researchers to see various tissue elements under a microscope [23]. Counterstains are utilized to provide contrast, color, and visibility. The goal of staining is to reveal the components at the cellular level. Hematoxylin-eosin (H&E) is a staining component that pathologists have used for decades. Eosin stains cytoplasm and connective tissues, which are pink, while Hematoxylin stains cell nuclei, which are blue. Figure 5 presents examples of histopathology images [24].

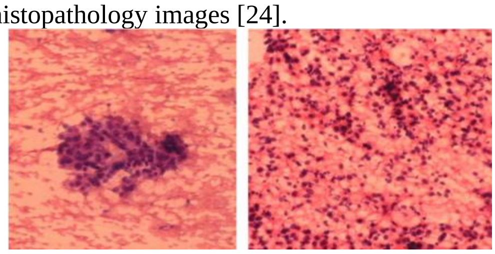

Figure 5**:** examples of histopathology images [23].

2.3. Datasets:

Using breast cancer classification algorithms, certain public image databases are commonly used. The DM images represent the majority of the images in these datasets, such as the Digital Database for Screening Mammography (DDSM) [25], Mammographic Image Analysis Society (MIAS) [26], Image Retrieval in Medical Applications (IRMA) [27], Wisconsin Diagnostic Breast Cancer (WDBC) [28], IN breast database [29] and Breast Cancer Digital Repository (BCDR). Table-1 shows the Available data set for digital mammography (DM) [30].

Furthermore, studies that employ different types of image modalities frequently rely on special or private datasets to implement their methodologies. Individual groups collect private databases independently of one another. Private datasets provide images of patient case studies gathered in research facilities or hospitals. It's a good idea to have common datasets with images from a variety of sources for various image modalities.

Table 1.Available datasets for Digital Mammography (DM)

Dataset	Number of images	Dataset size	Total patients
MIAS	322	2.3GB	161
DDSM	10.239	163.6GB	2620
BCDR	3703	100GB	1734
WBC	569	12.5KB	150
IN breast	419	9.01GB	115

2.4.Preprocessing:

Current preprocessing techniques for mammographic images are reviewed. The purpose of preprocessing is the enhancement of visual data by minimizing noise and undesired elements that are in the backdrop of the mammography this also helps to improve some aspects of the image that can be utilized for further procedure [31].

Preprocessing a mammography image can be done in a variety of ways, such as Filtering, Converting it into a binary image, Image transformation, Image histogram adjustment, Contrast Limited Adaptive Histogram Equalization (CLAHE), unsharp masking, Image border neighborhoods removal and Median filtering [32].

2.5. Segmentation stage:

It is a significant stage of the typical CAD system, this stage is used to split the suspicious regions that may contain masses from the background parenchyma (normal tissue), leads to partition the mammogram into many non-overlapping regions, next extract the (ROI), and find the suspicious mass candidates from ROIs [18]. The most important segmentation methods used in medical imaging are:

•Thresholding: Is a procedure aims to find an intensity value, or threshold, that distinguishes the desired classes. After that, segmentation is accomplished by grouping all pixels with intensity greater than the threshold into one class and all other pixels into another [18].

•Clustering: Without use of training data, clustering essentially performs the same task as classification methods. Three widely utilized clustering algorithms are K-means algorithm, fuzzy algorithm, and Expectation-Maximization (EM) algorithm [33].

• Region Growing (RG): is a tool for obtaining a region of interest in an image using predefined criteria. This criterion can be based on image edges or intensity information [34]. In its most basic form, RG requires a seed point that is manually chosen by an operator and extracts all pixel values with the same intensity value that is connected to the initial seed.

• Artificial Neural Network (ANN): ANNs are machine learning models that can be applied to image segmentation in a variety of ways. In medical imaging, the most common application is as a classifier, in which the weights are determined using training data and the ANN is then used to segment the new data [12].

2.6.Features extraction:

In order to reduce classification errors, the chosen characteristics must be capable of distinguishing between normal and malignant areas [31]. Distinct researchers utilized a different collection of attributes that they believe are most suited to this goal. This evaluation includes a large number of features, but only some of the most commonly used and selected features are shown in this area [32].

The most common feature categories utilized in this review articles for breast tumor classification are [35] [36]:

• Histogram: Such as Standard Deviation, Mean, Entropy and Energy.
• Morphology: Such as Standard Deviation of distance ratio, Area Overlap Ratio, Smoothness, Variance of Distance Ratio.
• Textural: Such as Contrast, Variance Homogeneity, Correlation Information Measurement, and Sum of the Sums, Inertia Variance, and Inverse Difference.
• Speculation: Such as Full Width Half Maximum (FWHM) border, FWHM Growing, variation in margin sharpness, Index of Radial Gradient,

Growing a Gradient, FWHM ROI and FWHM margin, and Irregularity are all geometric properties.

• Kinetic: Such as Maximum Enhancement, Uptake Rate, Time to Peak, Washout Rate, Signal Enhancement Ratio, Enhancement at the first post-contrast time point Curve shape index.

• Binary features: Such as Area, centroid, and Orientation (Axis of rotation).

2.7. Classification stage:

Several (MLTs) were used to detect, predict, and diagnose breast cancer. The dataset is separated into training and test sets for the classification algorithms. The training dataset is used to develop the model, and then the test dataset is used to validate the trained model extraction of data.

Several (MLTs) were employed in the breast cells classification during the procedure based on images' derived characteristics such as:

• Support Vector Machine (SVM) classifier: It is the most commonly employed (MLT) among the publications examined in the review. A supervised classifier it creates a model using a hyper plane as a border that separates multiple places in 2D space into several classes and divides them; the test sample is classified using this plan [37].

• Artificial Neural Network (ANN) classifier: It links all nodes in the same way as neurons in the human brain do. The input to one of the nodes is the total of the outputs of all the nodes to which it is linked multiplied by a weight. The output value of a definite node is processed by the "transfer function.

Neural Network (NN) is made up of successive layers. After assigning weights to nodes in the first hidden layer, an input layer takes data and transmits it to them. The result is translated to the following layer's nodes, and so on. The network's output is provided by the last layer. The name ANN has been integrated with a variety of different classification approaches [38] including: when it is combined with an associative classification methodology, it known as (ACNN), when fuzzy is added, it is known as (ACFNN) [39].

• KNN classifier: It classifies an unknown sample by computing its distance from all of the training samples. It calculates the k shortest distances. The most represented class among these k classes assigns the unknown sample's output class label [40].

• Decision Tree (DT): In a tree structure, (DT) places a series of well-crafted queries concerning the test record's properties. When a response is obtained, a follow-up question is asked until a resolution is reached regarding identifying the record's class. One root node, multiple internal nodes, and many terminal nodes make up a decision tree classifier. The goal of decision tree building is to identify a characteristic that provides the most information gain [41] [42].For prediction, RF integrates numerous decision trees built by pulling several classification trees together [43].

• Fuzzy classifier: It is describes an object's partial membership in distinct classes to varying degrees. Fuzzy IF-THEN rules describe a classifier. The most widely used unsupervised classification technique based on fuzzy logic is the fuzzy c-mean (FCM). The data points have membership values with the cluster centers, that will be iteratively updated [44].

•Naive Bayes (NB): It is based on the Bayes theorem and a probabilistic method. It indicates the likelihood that a particular pattern belongs to a specified class. The probability of a random class variable is calculated and assessed using data about the value of another set of random variables [45] [46].

•Deep Learning (DL): It is the current way in machine learning that resulted in novel strategies for training deep neural networks, which have produced extremely effective applications in a variety of pattern recognition tasks [47].

ANN is commonly used in DL, where DL employs the back propagation technique to discover complex structures in large data sets to determine how a machine should adjust its internal parameters that are required to calculate the representation of each layer from the preceding layer's representation [48].

•Multiple Instances learning (MIL): It had solved the learning issues using partial data label information. Each instance in MIL is characterized by a feature vector, and the class label is connected with it [49].

3. Evaluation metrics

Test data are entered into the classifier so as to be classified after it has been trained using samples of

training data in order to assess the performance of the classifier.

The samples are classified into two categories: positive and negative. Positive is the main category of interest, which represents malignant (cancerous) samples, and the other category is negative samples are benign (non-cancerous) samples.

The common used assessment measures for CAD systems are Accuracy (Acc), Sensitivity (Sn), Specificity (Sp), and Area Under the Curve (AUC). These are the most often cited performance measures in the selected articles.

The previous measures are defined as:

(1) Accuracy (Acc) indicates just how close the real class is to the anticipated one. It means that is, it represents the proportion of correctly categorized samples (non-cancer and cancer) with the total number of samples.

(2) Sensitivity (Sn) is the true positive rate, which reflects the percentage of (cancer) samples accurately diagnosed.
(3) Specificity (Sp) is the true negative rate, which determines the percentage of (non-cancer) samples that are accurately identified.
(4) Area Under the Curve (AUC) is a standard statistic that indicates a method of selecting optimum models and ignoring suboptimal ones. This depicts the genuine positive rate with the false positive rate. When the AUC is near one, the test is considered to be good. AUC higher than or equal to 0.5 and less than 1 are considered reasonable tests [50]. Equations of Acc, Sn, and Sp are given as follows [49],[50]:

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \qquad (3.1)$$

$$\mathrm{Sn} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad (3.2)$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \qquad (3.3)$$

Where TP: true positive, TN: true negative, FP: false positive, FN: false negative.”

The table 2 below summarizes the findings of previous researches in the field of breast cancer classification.

4. A survey of previous researches:

Table 2. A survey of the previous researches regarding ref. number, imaging modality, data set, machine learning technique and results:

Ref	Imaging modality	Used dataset	Machine learning technique	Evaluation results			
				Acc	AUC	Sp	Sn
[51]	DM	Wisconsin	CNN	97.20%		-	
[52]	DM	IN breast	SVM	71.0% to 83.1%	-	-	-
[53]	US	Private	KNN	-	0.93	89.5%	87.8%
[54]	MRI	Private 200 patients	KNN	74.7%	0.816	-	-
[55]	DM	Wisconsin	SVM	98%	0.9938	95.08%	98.24%
[56]	DM	WBC	SVM	97.14%	-	-	-
[57]	DM	Mini-MIAS	Fuzzy with SVM	98%	-	-	-
[58]	US	Ultrasound dataset	SVM	94.74%	-	-	-
			KNN	89.44%	-	-	-
		Elastography dataset	SVM	100%	-	-	-
			KNN		0.826	-	-
[59]	DM	Private 344 patients	DL	-	-	86.4%	100%
[60]	microscopic Images	Private 202 images	ANN	87.1%	-	77.27%	96.49%
			SVM	77.23%	0.9325	-	-
[61]	DM	Mini-MIAS	SVM	90%	-	-	-
			LR	92.5%	-	-	-
			KNN	89.3%	-	-	-
[62]	US	Private	KNN	67.7%	-	-	-
[63]	DM	MIAS	SVM	92.3%	-	-	-
[64]	DM	MIAS	SVM	92.5%	-	-	-
[65]	DM	MIAS	SVM	80.23%	-	98.86%	-
[66]	DM	WBC	SVM	-	-	-	-
[67]	DM	WBC	DL_ANN	98.24%	-	-	-
[68]	DM	WBC	NB	97.36%	-	-	-
[69]	DM	BCDG	KNN	77%			

5. Conclusions:

This survey aims to aid researchers in improving CAD systems to provide support to the medical society in breast cancer detection.

Based on the data obtained, and although there are a variety of circumstances, it can be said that certain facts can be assured for the researchers in this field. Some of these criteria include the datasets utilized for evaluation, the images used for evaluation, the number of images used, and the evaluation process (validation methodology, testing, and training set).

In general, it is concluded that the mostly used method of the Imaging modality is the Digital Mammogram (DM). The (SVM) classifier showed to be widely employed for breast tumor categorization and it is one of the best ways among the classifiers indicated in the survey table (Table 2).

Clinically, in countries where CAD is commonly used, there is a significant discussion regarding whether it can be utilized or not. This could be caused by several issues, such as ignoring worrisome pests due to lack of training which leads to wasted time and cost.

Otherwise, it is quite difficult in other countries to persuade physicians to admit CADs into clinics for their benefits as a supportive tool to develop physician performance. Noting that CAD systems can save radiologists and clinicians time and effort. “”CAD systems ought to be improved through their implementation and testing in clinics. Adequate and correct use will result in the clinical enforcement of CADs. This leads to improved performance, minimizing false positives that may result in psychological, economic, and physical losses. Also it may lead to lowering false-negative readings that may result in therapeutic neglect.

REFERENCES

[1] Doctor, G., & Varkey, M. (2022). Artificial Intelligence in Healthcare: Case Studies. In Handbook of Research on Lifestyle Sustainability and Management Solutions Using AI, Big Data Analytics, and Visualization (pp. 305-324). IGI Global.

[2] Censin, J. C., Peters, S. A., Bovijn, J., Ferreira, T., Pulit, S. L., Mägi, R., ... & Lindgren, C. M. (2019). Causal relationships between obesity and the leading causes of death in women and men. PLoS genetics, 15(10), e1008405..

[3] Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. International Journal of Environmental Research and Public Health, 18(1), 271..

[4] Halalli, B., & Makandar, A. (2018). Computer aided diagnosis-medical image analysis techniques. Breast Imaging, 85..

[5] Henriksen, E. L., Carlsen, J. F., Vejborg, I. M., Nielsen, M. B., & Lauridsen, C. A. (2019). The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review. Acta Radiologica, 60(1), 13-18.

[6] Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2022). Transformers in medical imaging: A survey. arXiv preprint arXiv:2201.09873.

[7] Kost, G. J. (2021). The impact of increasing disease prevalence, false omissions, and diagnostic uncertainty on coronavirus disease 2019 (COVID-19) test performance. Archives of Pathology & Laboratory Medicine, 145(7), 797-813.

[8] Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data, 6(1), 1-18.

[9] Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H. & Lungren, M. P. (2020). Preparing medical imaging data for machine learning. Radiology, 295(1), 4-15.

[10] Yan, Q., Gong, D., & Zhang, Y. (2018). Two-stream convolutional networks for blind image quality assessment. IEEE Transactions on Image Processing, 28(5), 2200-2211..

[11] Pathak, P., Jalal, A. S., & Rai, R. (2021). Breast Cancer Image Classification: A Review. Current Medical Imaging, 17(6), 720-740.

[12] Nasser, I. M., & Abu-Naser, S. S. (2019). Predicting tumor category using artificial neural networks.

[13] Chen, T., Shang, C., Su, P., Keravnou-Papailiou, E., Zhao, Y., Antoniou, G., & Shen, Q. (2021). A decision tree-initialised neuro-fuzzy approach for clinical decision support. Artificial Intelligence in Medicine, 111, 101986..

[14] Hazra, R., Banerjee, M., & Badia, L. (2020). Machine learning for breast cancer classification with ann and decision tree. In 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0522-0527). IEEE.

[15] Liu, Y., Ren, L., Cao, X., & Tong, Y. (2020). Breast tumors recognition based on edge feature extraction using support vector machine. Biomedical Signal Processing and Control, 58, 101825.

[16] Shia, W. C., & Chen, D. R. (2021). Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine. Computerized Medical Imaging and Graphics, 87, 101829.

[17] Gastounioti, A., McCarthy, A. M., Pantalone, L., Synnestvedt, M., Kontos, D., & Conant, E. F. (2019). Effect of mammographic screening modality on breast density assessment: digital mammography versus digital breast tomosynthesis. Radiology, 291(2), 320.

[18] Muramatsu, C., Hara, T., Endo, T., & Fujita, H. (2016). Breast mass classification on mammograms using radial local ternary patterns. Computers in biology and medicine, 72, 43-53.

[19] Yap, M. H., Pons, G., Marti, J., Ganau, S., Sentis, M., Zwiggelaar, R., ... & Marti, R. (2017). Automated breast ultrasound lesions detection using convolutional neural networks. IEEE journal of biomedical and health informatics, 22(4), 1218-1226.

[20] Moon, W. K., Chen, I. L., Chang, J. M., Shin, S. U., Lo, C. M., & Chang, R. F. (2017). The adaptive

computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound. Ultrasonics, 76, 70-77.
[21] Shan, J., Alam, S. K., Garra, B., Zhang, Y., & Ahmed, T. (2016). Computer-aided diagnosis for breast ultrasound using computerized BI-RADS features and machine learning methods. Ultrasound in medicine & biology, 42(4), 980-988..
[22] Jiao, H., Jiang, X., Pang, Z., Lin, X., Huang, Y., & Li, L. (2020). Deep convolutional neural networks-based automatic breast segmentation and mass detection in DCE-MRI. Computational and Mathematical Methods in Medicine, 2020.
[23] Tomaszewski, J. E. (2021). Overview of the role of artificial intelligence in pathology: the computer as a pathology digital assistant. In Artificial Intelligence and Deep Learning in Pathology (pp. 237-262). Elsevier.
[24] Singh, S., & Kumar, R. (2020). Histopathological image analysis for breast cancer detection using cubic SVM. In 2020 7th international conference on signal processing and integrated networks (SPIN) (pp. 498-503). IEEE.
[25] Kim, Y. J., Yoo, E. Y., & Kim, K. G. (2021). Deep learning based pectoral muscle segmentation on Mammographic Image Analysis Society (MIAS) mammograms. Precision and Future Medicine, 5(2), 77-82.
[26] Öztürk, Ş. (2020). Stacked auto-encoder based tagging with deep features for content-based medical image retrieval. Expert Systems with Applications, 161, 113693..
[27] Qiu, Y., Yan, S., Gundreddy, R. R., Wang, Y., Cheng, S., Liu, H., & Zheng, B. (2017). A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. Journal of X-ray Science and Technology, 25(5), 751-763.
[28] Abdulla, S. H., Sagheer, A. M., & Veisi, H. (2021). Breast Cancer Classification Using Machine Learning Techniques: A Review. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(14), 1970-1979.
[29] Alshammari, M. M., Almuhanna, A., & Alhiyafi, J. (2021). Mammography Image-Based Diagnosis of Breast Cancer Using Machine Learning: A Pilot Study. Sensors, 22(1), 203.
[30] https://www.kaggle.com
[31] William, W., Ware, A., Basaza-Ejiri, A. H., & Obungoloch, J. (2018). A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. Computer methods and programs in biomedicine, 164, 15-22.
[32] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. ACM computing surveys (CSUR), 50(6), 1-45.
[33] Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging, 32(4), 582-596.
[34] Biswas, H. (2021). Evaluation of Deep Learning Models for the Detection of Diabetic Retinopathy with Retinal Fundus Images and Development of a Prescreening Tool (Doctoral dissertation, Southern Illinois University at Edwardsville).
[35] Wang, Q., Mao, N., Liu, M., Shi, Y., Ma, H., Dong, J. & Xie, H. (2021). Radiomic analysis on magnetic resonance diffusion weighted image in distinguishing triple-negative breast cancer from other subtypes: a feasibility study. Clinical Imaging, 72, 136-141.
[36] Houssein, E. H., Emam, M. M., Ali, A. A., & Suganthan, P. N. (2021). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications, 167, 114161..
[37] Sharma, A., & Mishra, P. K. (2022). Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. International Journal of Information Technology, 14(4), 1949-1960..
[38] Xi, P., Shu, C., & Goubran, R. (2018). Abnormality detection in mammography using deep convolutional neural networks. In 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-6). IEEE..
[39] Boulehmi, H., Mahersia, H., & Hamrouni, K. (2018). Bone cancer diagnosis using GGD analysis. In 2018 15th International Multi-Conference on Systems, Signals & Devices (SSD) (pp. 246-251). IEEE.
[40] Huang, K., Li, S., Kang, X., & Fang, L. (2016). Spectral–spatial hyperspectral image classification based on KNN. Sensing and Imaging, 17(1), 1-13.
[41] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.
[42] Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. International Journal of Computer Sciences and Engineering, 6(10), 74-78..
[43] Naghibi, S. A., Pourghasemi, H. R., & Dixon, B. (2016). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environmental monitoring and assessment, 188(1), 1-27..
[44] OWSI, J., Kacprzyk, J., Shyrai, S., Szmidt, E., Viattchenin, D. A., & HORMAZABAL, J. H. (2018). A heuristic algorithm of possibilistic clustering with partial supervision for classification of the intuitionistic fuzzy Data. J. Multiple-Valued Logic Soft Comput., 31(4), 399-423.
[45] de Lima, S. M., da Silva-Filho, A. G., & Dos Santos, W. P. (2016). Detection and classification of masses in mammographic images in a multi-kernel approach. Computer methods and programs in biomedicine, 134, 11-29.
[46] Xiao, R., Cui, X., Qiao, H., Zheng, X., Zhang, Y., Zhang, C., & Liu, X. (2021). Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net. Biomedical Signal Processing and Control, 66, 102362.
[47] Al-Antari, M. A., Al-Masni, M. A., & Kim, T. S. (2020). Deep learning computer-aided diagnosis for breast lesion in digital mammogram. Deep Learning in Medical Image Analysis, 59-72.
[48] Jahangir, H., Golkar, M. A., Alhameli, F., Mazouz, A., Ahmadian, A., & Elkamel, A. (2020). Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough

ANN. Sustainable Energy Technologies and Assessments, 38, 100601..

[49] Mansour, R. F. (2020). A robust deep neural network based breast cancer detection and classification. International Journal of Computational Intelligence and Applications, 19(01), 2050007.

[50] Yassin, N. I., Omran, S., El Houby, E. M., & Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Computer methods and programs in biomedicine, 156, 25-45.

[51] Naji, M. A., El Filali, S., Aarika, K., Benlahmar, E. H., Abdelouhahid, R. A., & Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Computer Science, 191, 487-492.

[52] Diz, J., Marreiros, G., & Freitas, A. (2016). Applying data mining techniques to improve breast cancer diagnosis. Journal of medical systems, 40(9), 1-7.

[53] Phadke, A. C., & Rege, P. P. (2016). Fusion of local and global features for classification of abnormality in mammograms. Sādhanā, 41(4), 385-395.

[54] Waugh, S. A., Purdie, C. A., Jordan, L. B., Vinnicombe, S., Lerski, R. A., Martin, P., & Thompson, A. M. (2016). Magnetic resonance imaging texture analysis classification of primary breast cancer. European radiology, 26(2), 322-330.

[55] Gopal, V. N., Al-Turjman, F., Kumar, R., Anand, L., & Rajesh, M. (2021). Feature selection and classification in breast cancer prediction using IoT and machine learning. Measurement, 178, 109442

[56]. Daoud, M. I., Bdair, T. M., Al-Najar, M., & Alazrai, R. (2016). A fusion-based approach for breast ultrasound image classification using multiple-ROI texture and morphological analyses. Computational and mathematical methods in medicine, 2016

[57]. Hariraj, V., Khairunizam, W., Vikneswaran, V., Ibrahim, Z., Shahriman, A. B., Zuradzman, M. R., ... & Sathiyasheelan, R. (2018). Fuzzy multi-layer SVM classification of breast cancer mammogram images. Int J Mech Engg Tech, 9(8), 1281-99.

[58] Ramya, S., & Nanda, S. (2017). Breast cancer detection and classification using ultrasound and ultrasound Elastography images. IRJET, 4, 596-601.

[59] Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Lopez, M. A. G. (2016). Representation learning for mammography mass lesion classification with convolutional neural networks. Computer methods and programs in biomedicine, 127, 248-257.

[60] Jeleń, Ł., Krzyżak, A., Fevens, T., & Jeleń, M. (2016). Influence of feature set reduction on breast cancer malignancy classification of fine needle aspiration biopsies. Computers in biology and medicine, 79, 80-91.

[61] Farhan, A. H., & Kamil, M. Y. (2020). Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques. In IOP conference series: materials science and engineering (Vol. 928, No. 7, p. 072098). IOP Publishing.

[62] Pourasad, Y., Zarouri, E., Salemizadeh Parizi, M., & Salih Mohammed, A. (2021). Presentation of novel architecture for diagnosis and identifying breast cancer location based on ultrasound images using machine learning. Diagnostics, 11(10), 1870.

[63] Singh, S., & Kumar, R. (2020). Histopathological image analysis for breast cancer detection using cubic SVM. In 2020 7th international conference on signal processing and integrated networks (SPIN) (pp. 498-503). IEEE.

[64] Zhuang, Z., Yang, Z., Zhuang, S., Joseph Raj, A. N., Yuan, Y., & Nersisson, R. (2021). Multi-features-based automated breast tumor diagnosis using ultrasound image and support vector machine. Computational Intelligence and Neuroscience, 2021.

[65] Alfian, G., Syafrudin, M., Fahrurrozi, I., Fitriyani, N. L., Atmaji, F. T. D., Widodo, T., ... & Rhee, J. (2022). Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. Computers, 11(9), 136.

[66] Laghmati, S., Cherradi, B., Tmiri, A., Daanouni, O., & Hamida, S. (2020). Classification of patients with breast cancer using neighbourhood component analysis and supervised machine learning techniques. In 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet) (pp. 1-6). IEEE.

[67] Gupta, P., & Garg, S. (2020). Breast cancer prediction using varying parameters of machine learning models. Procedia Computer Science, 171, 593-601.

[68] Chaurasia, V., Pal, S., & Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 12(2), 119-126.

[69] Tafish, M. H., & El-Halees, A. M. (2018). Breast cancer severity degree predication using data mining techniques in the gaza strip. In 2018 International Conference on Promising Electronic Technologies (ICPET) (pp. 124-128). IEEE.