Applied Computer Science, vol. 20, no. 1, pp. 1–16
doi: 10.35784/acs-2024-01

Submitted: 2024-01-19 | Revised: 2024-02-27 | Accepted: 2024-02-29

Keywords: machine learning, breast cancer, CNN, image processing, healthcare

Malek M. AL-NAWASHI [0000-0002-5231-8155]*,
Obaida M. AL-HAZAIMEH [0000-0001-5641-4892]*,
Mutaz Kh. KHAZAALEH [0000-0002-2071-7020]*

A NEW APPROACH FOR BREAST CANCER DETECTION- BASED MACHINE LEARNING TECHNIQUE

Abstract

The leading cause of cancer-related mortality is breast cancer. Breast cancer detection at an early stage is crucial. Data on breast cancer can be diagnosed using a number of different Machine learning approaches. Automated breast cancer diagnosis using a Machine Learning model is introduced in this research. Features were selected using Convolutional Neural Networks (CNNs) as a classifier model, and noise was removed using Contrast Limited Adaptive Histogram Equalization (CLAHE). On top of that, the research compares five algorithms: Random Forest, SVM, KNN, Naïve Bayes classifier, and Logistic Regression. An extensive dataset of 3002 combined images was used to test the system. The dataset included information from 1400 individuals who underwent digital mammography between 2007 and 2015. Accuracy and precision are the metrics by which the system's performance is evaluated. Due to its low computing power requirements and excellent accuracy, our suggested model is shown to be quite efficient in the simulation results.

1. INTRODUCTION

Everyone, regardless of age or socioeconomic status, is vulnerable to the devastating effects of cancer. There are various varieties of cancer, but breast cancer is one of the most common tumors among women. This difficulty underscores the need for researchers to focus on improving cancer diagnosis and prognosis. Machine-learning strategies may have a major impact in cancer early detection and prognosis. Breast cancer, which begins in the ducts of the breast, affects a disproportionately high number of women globally. Among cancers that affect women, breast cancer is the second highest cause of mortality (Fatima et al., 2020; Houssein et al., 2021;Yassin et al., 2018). Breast cancer risk may be increased by the following factors: Age is a significant determinant in breast cancer development. Although females have a higher incidence of breast cancer, males are not impervious to the condition.

* Al-Balqa Applied University, Information Technology Department, Jordan, nawashi@bau.edu.jo, dr_obaida@bau.edu.jo, mutaz.khazaaleh@bau.edu.jo

An individual's risk is increased by having a mother, daughter or sibling who has received a breast cancer diagnosis. The susceptibility to breast cancer is elevated in individuals harboring specific gene mutations, including the genes BRCA1 and BRCA2. Prolonged utilization of HRT (i.e., "Hormone Replacement Therapy ") has the potential to elevate the associated risk. Delayed puberty or lack of children, premature menstruation, and postponed menopause are all variables that may influence a woman's risk (Nanda et al., 2002). Breast cancer symptoms, as shown in Figure 1, include unexplained lumps in either breast, changes in breast size, shape, or appearance, pain in either breast, nipple discharge other than breast milk, and changes in the texture or color of the breast skin (Svensson et al., 2020).

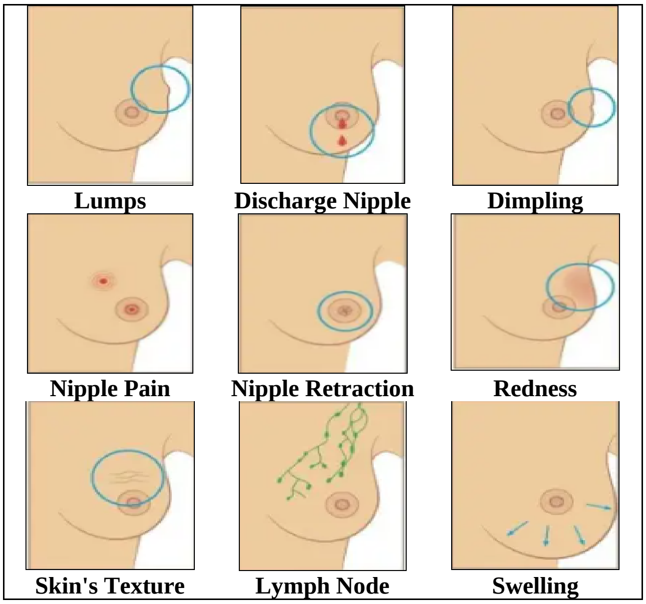


Fig. 1.Breast cancer symptoms

Figure 2 displays a single example of each distinct abnormality type, including mass, carcinoma, calcification, and asymmetry (Chang et al., 2021; Svensson et al., 2020).

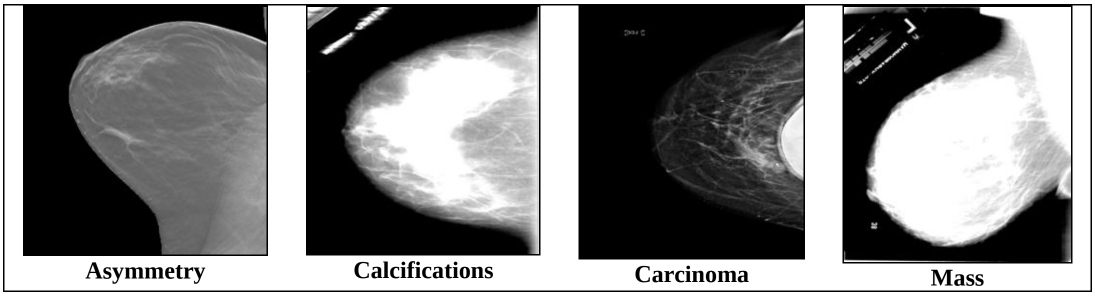


Fig. 2.Various forms of anomalies

The morphology of breast cancer cells is utilized to categorize the disease into various subgroups. DCIS (i.e., "Ductal Carcinoma In Situ") is a type of breast cancer that develops within the milk ducts and does not spread to surrounding tissues as shown in Figure 3 (Narod et al., 2015). If left untreated, it may occasionally advance to invasive breast cancer.

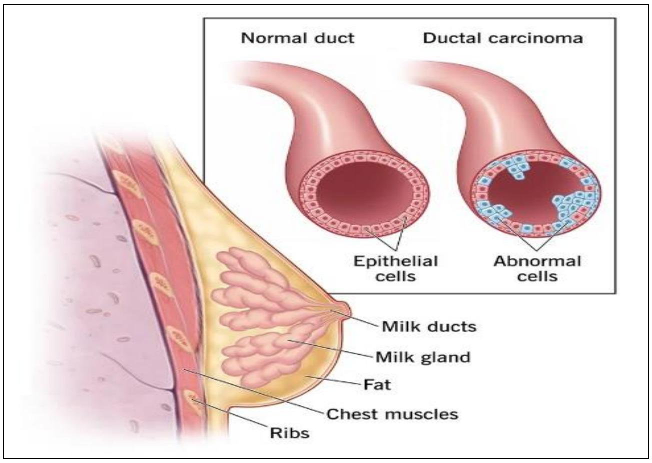


Fig. 3. DCIS

The predominant form is IDC (i.e., "Invasive Ductal Carcinoma"), comprising 80% of cases, and infiltrates adjacent breast tissues. In the classification of IDC, it is possible to categorize it according to the status of HER2andhormone receptors. ILC (i.e., "Invasive Lobular Carcinoma") arises from the lactiferous glands and constitutes 10-15% of breast cancer incidences. It is less prevalent than IDC and has greater challenges in terms of detection using mammography. Figure 4 displays the most common types of invasive breast cancer (Carlson et al., 2011).

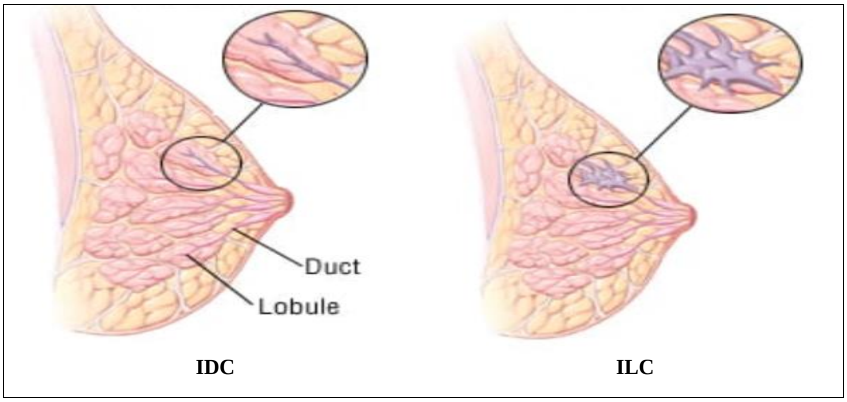


Fig. 4. Types of invasive breast cancer

TNBC (i.e., "Triple-Negative Breast Cancer") is a specific form of breast cancer characterized by the absence of hormone receptors and the HER2 protein. It exhibits heightened aggressiveness and reduced sensitivity to HER2-targeted and hormonal treatments. TNBC is commonly managed with the administration of chemotherapy. HER2-positive breast cancer as shown in Figure 5 is an uncommon and very aggressive kind of breast cancer that occurs due to an excessive production of the HER2 protein (i.e., "Human Epidermal growth factor Receptor 2"). The condition can manifest invasively or non-invasively, and it can be managed by targeted medicines like Herceptin (Loibl & Gianni, 2017).

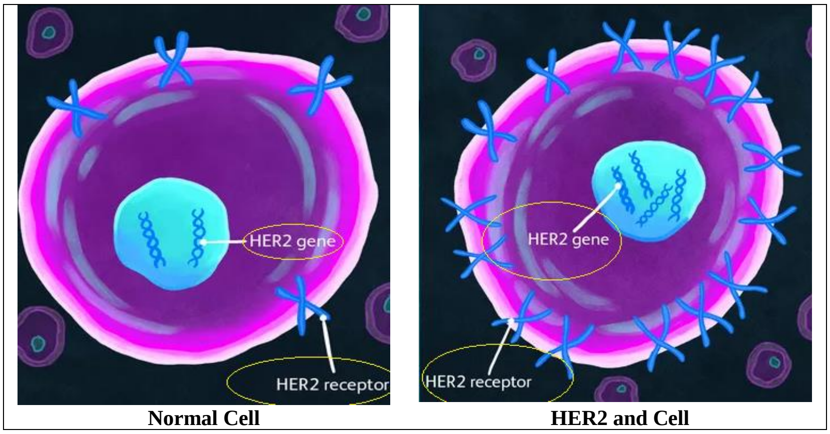


Fig. 5. HER2 Positive Cancer

IBC is an uncommon and aggressive subtype of breast cancer distinguished by breast erythema, elevated body temperature, and edema. Urgent and aggressive treatment is necessary. A rare and aggressive form, Paget's disease of the Nipple originates in the milk ducts and progresses to the areola and nipple epidermis. Metastatic breast cancer, referred to as stage IV, encompasses all breast cancer subtypes (HER2-positive,ILC, IDC) and is notoriously incurable (Al-hazaimeh et al., 2014; Loibl & Gianni, 2017; Nahar et al., 2018).

Breast cancer treatment requires early detection. Thus, breast cancer detection requires adequate screening technology. Ultrasonography, Mammography, and Thermograph are typical imaging modalities for screening this syndrome. Early detection of breast cancer is possible with mammography. Diagnostic ultrasound is typically utilized for women with dense breast tissue, as mammography is unsuccessful. These variables may make radiography and thermography better at spotting microscopic malignant tumors than ultrasonography. Artificial intelligence allows mammography to automatically identify diseases in medical pictures, which is vital to fighting breast cancer (Nallamala et al., 2019; Rivera-Franco & Leon-Rodriguez, 2018). The detection of early-stage breast cancer encounters a multitude of constraints and obstacles. Mammography and other breast cancer screening methods encounter a number of obstacles, including limitations in a false sense of security concerns, variability in screening interpretation, breast density, over diagnosis and over treatment, access and equity concerns, patient compliance issues, financial and resource constraints, sensitivity, risk prediction concerns, genetic and molecular influences, specificity, and invasive follow-up tests (Lu et al., 2009).Table 1 provides a succinct summary of various diagnostic methods, including their pros and cons. This paper provides a concise overview for readers to understand the current state of breast cancer diagnosis and the necessity for an improved diagnostic approach (Tagliafico et al., 2020).

Tab. 1. Current diagnostic approach- Pros and cons

Method	Pros	Cons
Ultrasound	Identifying cysts from other mass abnormalities.	Detection is poor in deep tissues.
Mammography	Changes in structure and calcifications are detected.	Reduced detection capability in breast tissue with high density.
Magnetic resonance imaging (MRI)	Thorough visibility of soft tissues.	Specialized skill is necessary to identify benign lesions.
Fine Needle Aspiration or Core Needle Biopsy (Biopsy)	The tissue samples are provided for a final diagnosis.	The sample size may not be adequate, and experienced medical professionals are needed.
Medical Genetics	Finds risky genetic mutations.	Useful for a narrow range of patients.

Convolutional neural networks (CNNs), a subset of machine learning, and artificial intelligence (AI) are now popular topics in the healthcare industry. Artificial intelligence (AI) and machine learning (ML) are utilized in the field of research that aims to create more advanced technological systems capable of performing complex tasks with reduced need on human intelligence (Al-hazaimeh et al., 2014; Al-hazaimeh et al., 2022; Al-Hazaimeh & Al-Smadi, 2019; Gharaibeh et al., 2023). An effective machine learning-based model is the goal of this paper. The proposed approach can detect malignancies in digital mammograms of varied densities, and the results may be compared to those obtained by utilizing state-of-the-art models. In order to detect breast cancer in digital mammograms of variable density, the authors presented a machine learning approach. As a result of this study, early diagnosis and tailored treatment can be improved. In addition, it can affect areas like research, cost-effectiveness, and global accessibility to healthcare, all of which have the potential to change breast cancer management and save lives. Because of the capabilities of machine learning in breast cancer diagnosis, healthcare expenses can be reduced and a positive impact can be made on breast cancer management around the world. In this paper, proceed with the subsequent sections in the following manner. The authors discuss prior research in the related work section of Section 2. Section 3 of the document outlined the proposed approach. The experimental outcomes are elaborated upon in Section 4. In Section 5, the conclusion to this investigation is presented.

2. RELATED WORK

Global breast cancer diagnoses in 2022 reached 2.73 million among women, resulting in 685,000 fatalities. By the conclusion of 2022, the number of women who had been diagnosed with breast cancer within the previous 5 years stood at 7.8 million, establishing it as the most widespread form of cancer globally. Incidence of breast cancer is observed in women of all countries worldwide, occurring at any age following puberty, with a higher prevalence in older age groups. Figure 6 provides a comprehensive summary of breast cancer statistics across various demographics in different countries (DeSantis et al., 2019; Wilkinson & Gathani, 2022).

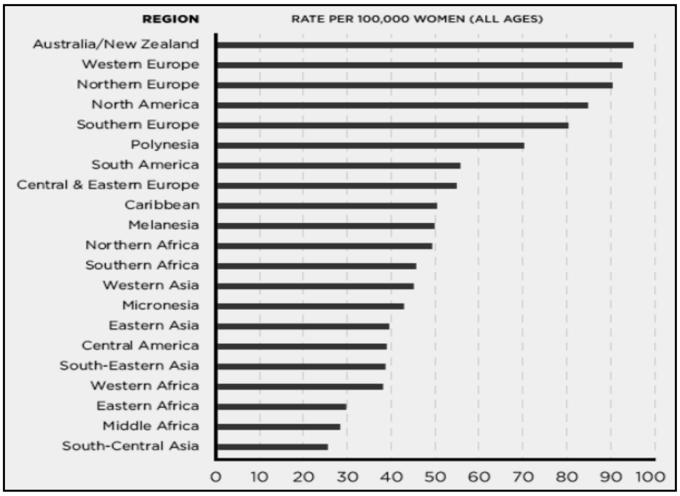


Fig. 6. Global breast cancer prevalence rates

Consequently, effective methods and tools are required for early breast cancer identification, since this is a prerequisite for effective breast cancer treatment (Alhindawi, et al., 2016; Barrios, 2022; Wilkinson & Gathani, 2022). Machine learning (ML) methods have been applied in various healthcare areas to identify BC (i.e., "Breast Cancer") in recent years. The algorithms' good performance has led other researchers to utilize them to detect the breast cancer (Hall et al., 2022). Various kinds of algorithms are used in machine learning. Table 2 summarizes the advantages and disadvantages of several machine learning techniques (Kharya et al., 2013).

Tab. 2. Machine learning methods- Pros and cons

Algorithm-Machine Learning	Pros	Cons
Neural Network	Capable of handling inputs that are noisy and representing Boolean functions.	Over-fitting occurs when there are too many attributes, and the structure of a network can be found through experimentation.
Naïve Bayes	The training algorithm is highly efficient, and the order in which instances are presented during training does not impact the results.	The presence of duplicate attributes can lead to misleading classification results, while the frequencies of attributes and classes have an impact on the accuracy of the classification.
Support Vector Machine	The decision rule's complexity can be easily controlled, and the models are capable of capturing non-linear class boundaries.	The training process is relatively sluggish in comparison to the efficiency of Bayes and Decision tree algorithms.
Decision Tree	The order of cases does not impact the training process and is easily comprehensible.	The accuracy of a decision tree is influenced by the order in which attributes are selected, and the presence of missing data for an attribute might lead to uncertainty.

The purpose of this part is to provide a comprehensive review of the literature on algorithms and procedures that are applicable to breast cancer detection methods, more especially to approaches that may identify breast cancer from digital images. Accurate detection requires the use of appropriate technologies to precisely identify the breast cancer features of the affected breast.

Using Machine Learning techniques as Bi-clustering Ada Boost techniques, Naïve Bayes classifier, RCNN classifier, HA-BiRNN (i.e., "Bidirectional Recurrent Neural Networks"), and SVM classifier, (Vaka et al., 2020) introduced a novel approach to identify breast cancer. The proposed methodology, a DNN (i.e., "Deep Neural Network") with Support Value, was compared to other machine learning techniques. The simulation results showed that the DNN algorithm performed better in terms of efficiency and image quality, which are important in modern healthcare systems. (Tiwari et al., 2020) developed a novel breast cancer detection method using Deep Learning (Recurrent Neural Network, Convolutional Neural Network, Artificial Neural Network) and Machine Learning (Naïve Bayes Classifier, Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbor, Logistic Regression). Comparing Machine Learning and Deep Learning methodologies revealed that CNN (97.3%) and ANN (99.3%) models have higher accuracy and efficiency. (Vasundhara et al., 2019) have presented an automated categorization system for mammography images, utilizing different machine learning algorithms, to classify them as Malignant, Normal, or Benign. A comparison analysis is conducted to evaluate the performance of Random Forests, Convolutional Neural Networks, and Support Vector Machines. The simulation results determined that CNN is the most effective classifier for instinctively classifying digital mammograms with the use of filtering and morphological processes. (Tanabe et al., 2020) devised a novel approach that employs a trained deep learning neural network system to classify different subtypes of breast cancer. Based on data collected from 221 real patients, the findings demonstrate a precision rate of 90.50 percent. This autonomous model has the capability to classify and identify breast cancer lesions without any requirement for human interaction. Assessing this model demonstrates its ability to analyze the condition of affected individuals during the detection phase, indicating that it is an advancement compared to previous methods.

This paper presents the first automated breast cancer diagnosis powered by a Machine Learning model. The classifier model used for feature selection was Convolutional Neural Networks (CNNs), and contrast limited adaptive histogram equalization was used to reduce noise (breast cancer). Moreover, five algorithms - Logistic Regression, Random Forest, Naïve Bayes classifier, KNN, and SVM classifier - are compared in the research. A large dataset consisting of 3002 merged photos was utilized to evaluate the system. Between 2007 and 2015, 1,400 women had digital mammography, and their data was compiled into this dataset.

3. PROPOSED APPROACH

In this paper, the authors offer a novel approach for automatically identifying breast cancer in digital images. The architecture of the suggested approach for detecting breast cancer disorders is depicted in Figure 7 as a flow diagram. Here, we dive deep into the proposed method's implementation steps.

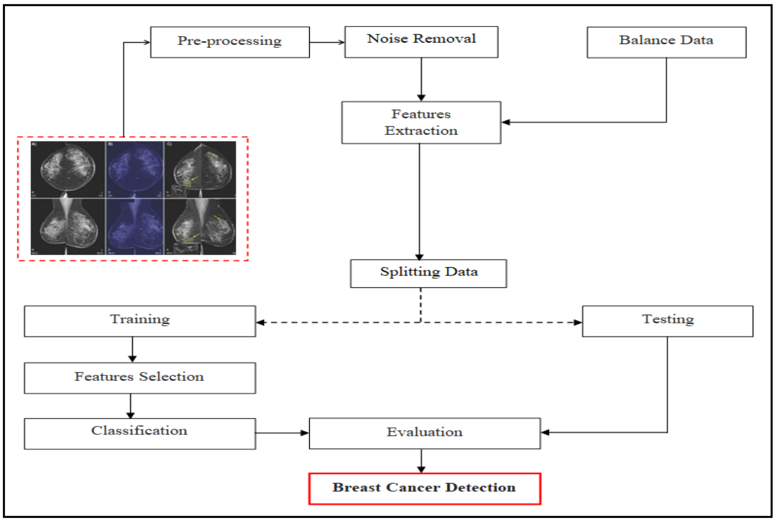


Fig. 7. Block diagram of proposed approach architecture

3.1. Pre-processing

In order to prepare images for use as an input to a model, pre-processing is necessary. One example is the requirement for image arrays of consistent size for the fully linked layers of convolutional neural networks. It's possible that preprocessing models can cut down on training and inference times. If the input images are quite large, reducing their size can drastically shorten the training time without severely degrading the model's performance. Though geometric image transformations like translation, rotation, and scaling are considered pre-processing techniques, the goal of pre-processing is to improve the picture data by enhancing certain elements necessary for subsequent processing or by decreasing unintended distortions (Gharaibeh et al., 2021; Gharaibeh et al., 2018; Ma'moun et al., 2014). In this paper, we will consider the pre-processing to be specified by Equation 1 and Equation 2:

$$d_1^i(m,n) = \begin{cases} d_1^i(m.n), if: e^i(m,n) \leq T^i \\ d_1^i(m.n), if: e^i(m,n) > T^i \end{cases} \tag{1}$$

$$e^i(m,n) = \begin{cases} \left|d_1^i(m.n)\right|, if \ \left|d_1^i(m.n)\right| > \left|d_1^i(m+1,n)\right| \\ 0\ , and \ \left|d_1^i(m.n)\right| > \ \left|d_1^i(m-1,n)\right| \end{cases} \tag{2}$$

3.2. Noise Removal

Generally, denoising (i.e., DE) is the act of eliminating undesirable distortions or noise from data. In other words, DE noising refers to the procedure of removing noise from a digital image. It is crucial to do DE noising prior to the image processing stage, as images obtained from a database often contain noise due to many factors such as image capture, coding, transmission, and processing. In order to eliminate unwanted sounds, it is necessary to employ a Non-Linear Wiener filter that utilizes a 4th order partial derivative with quad tree decomposition (Al-hazaimeh et al., 2014; Al-hazaimeh et al., 2022; Gharaibeh et al., 2018). The elimination of noises from input images is determined by:

$$f(g) = \frac{1}{\sigma^4 \sqrt{2\pi}} (g-m) e^{\frac{-(g-m)^2}{2\pi}} \, dg \qquad (3)$$

In this equation, g represents a pixel of the image, m represents the mean value of the Gaussian noise, and σ^2 is the standard deviation of the noise image. Following the completion of the DE noise reduction procedure, the CLAHE (i.e., "Contrast Limited Adaptive Histogram Equalization") algorithm must be utilized in order to rectify the uneven illumination. Enhancing the contrast of a breast image with the CLAHE approach can result in significant improvements to the image's critical data (Reza, 2004). In Figure 8, the overall flowchart of CLAHE algorithm. The outcome of the executed CLAHE algorithm is depicted in Figure 9.

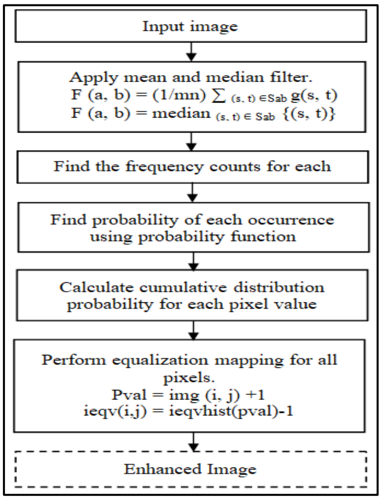
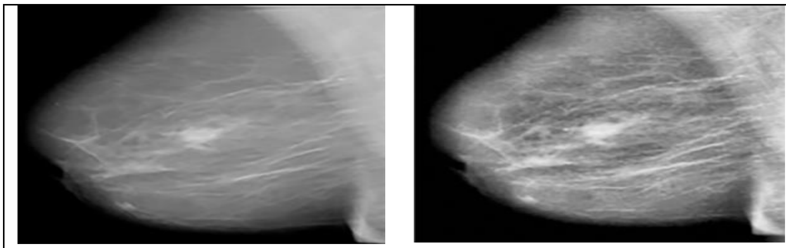


Fig. 8. The flowchart of CLAHE algorithm



Original Image **Image post-CLAHE processing**

Fig. 9. The CLAHE algorithm result

As shown in Figure 10, the histogram of the original image as well as the image after CLAHE transformation was applied.

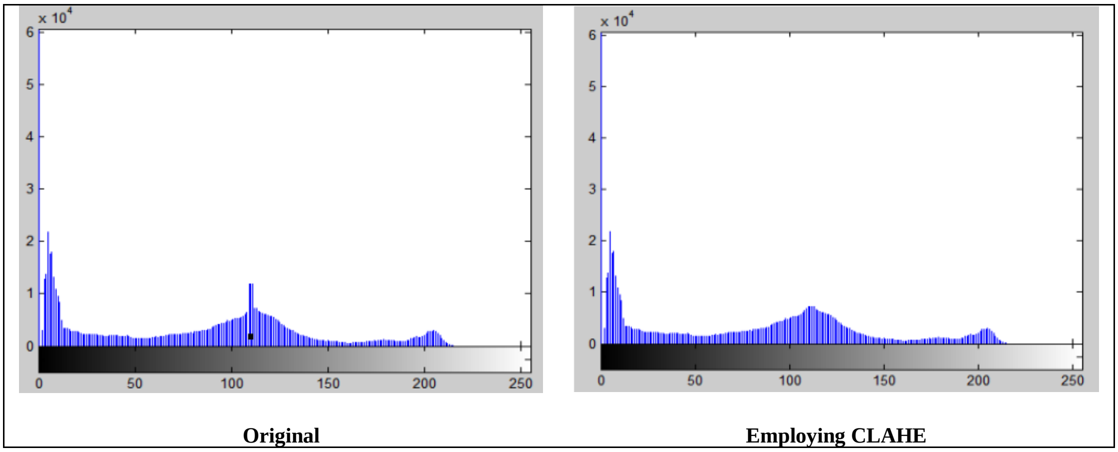


Fig. 10. Visual representation using histograms

3.3. Splitting Data

Divide the dataset into two parts - a training set and a testing set. This involves selecting rows from the dataset without replacement, often about 75 percent (although this percentage can be adjusted), and adding them to the training set. Twenty-five percent of the data is allocated to the test set. It should be noted that the colors in the "Features" and "Target" sections represent the destination of their respective data (i.e., "X-train", "X-test", "Y-train", "Y-test") for a specific train-test split as shown in Figure 11.

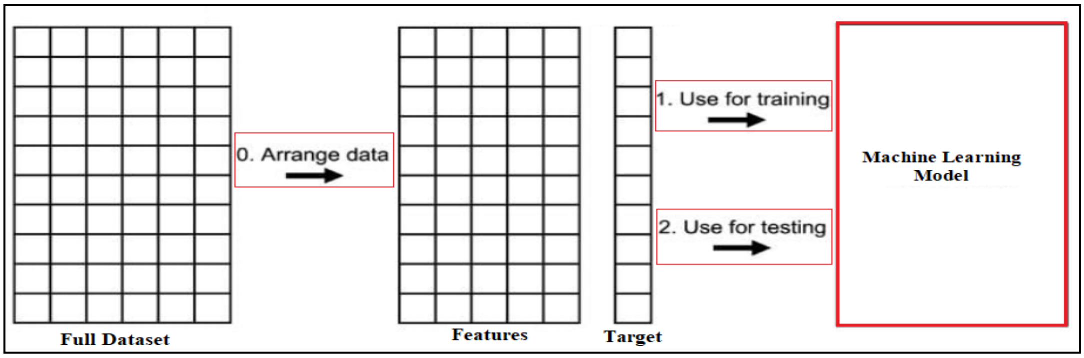


Fig. 11. Data splitting

3.4. Machine learning Model

In machine learning, features are learned for one task and then applied to another task so that no new features need to be learned. CNN models (i.e., "Convolutional Neural Network") that have been trained on the most popular datasets are typically used for this (Al-Hazaimeh et al., 2019; Al-Nawashi et al., 2017; Nahar et al., 2020; Sivapriya et al., 2019). The CNN model examines an input image and gives weights to features in an effort to distinguish between similar images. Several layers can be seen in Figure 12.

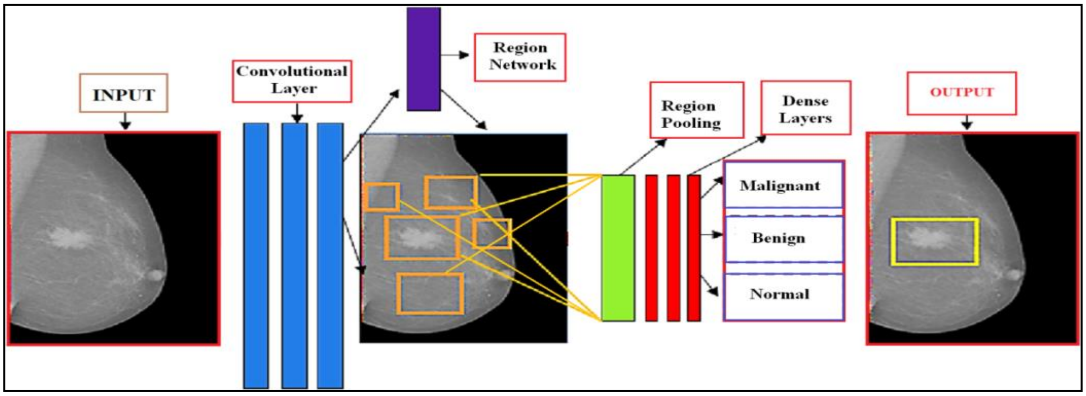


Fig. 12. CNN – detection and classification

The architecture described in this paper is founded on the sequential model, enabling the model to arrange sequential layers of the network in a sequential manner, starting from the input and progressing towards the output. The convolution layer employs filters that execute convolution operations while scanning the input image in relation to its dimensions. The hyper-parameters of this model consist of the filter size and stride, where the stride denotes the distance between successive receptive filters. The output that is obtained is referred to as a feature map or activation map. The initial breast photos are processed by including a 2-D convolutional layer. When calling the convolution layer function, the first argument is the number of output channels; for this case, we've used 16. In the proposed approach, we utilized a (3x3) filter kernel with a stride of 1 to accomplish spatial convolution. The kernel slides over the width and height of the picture. Throughout the studies, we also conducted analyses using various kernel sizes ranging from 1 to 7 in order to assess the variations. We have observed that the utilization of a tiny kernel size of (1×1) results in a constriction of the network's capacity. In this model, we have taken into account the use of padding to ensure that the input image is completely covered by the set filter and stride. The RELU (i.e., "Rectified Linear Unit") activation function, denoted as f(x) = max(0,x), has been utilized in our approach. Here, x represents the input. Max pooling is employed to decrease the spatial dimensions of the feature map. The Max-pooling function reduces the size of the input representation by selecting the highest value within a set window size for each dimension along the feature's axis. Furthermore, the process is replicated using an additional two convolutional layers, each producing 32 and 64 output channels respectively. Initially, we employed a 2x2 max pooling filter. Building upon the research undertaken in reference (Nguyen et al., 2013) to tackle the issue of deterioration, we have employed the identical concept of a deep residual learning framework. The system operates on recurrent units of convolutional layers with filter sizes of (1x1), (3x3). Every feature map's activation was computed using global average pooling.

4. RESULTS EVALUATION

Various evaluation tools can be used to analyze a classifier, including the accuracy, Recall, and precision. Our approach was constructed using MATLAB version 2017 and evaluated on many mammography images. The testing was conducted on a system equipped with an Intel Core i7 processor and 8GB of RAM. This research aims to assess the proposed method and compare it with state-of-the-art machine learning algorithms in terms of performance accuracy, recall, and precision. In this study, machine learning classifiers such as LR (Logistic Regression), SVM (Support Vector Machine), NB (Naive Bayes), RF (Random Forest), and KNN (K-Nearest Neighbors) are utilized for breast cancer prediction. These classifiers are chosen depending on the selected features. In addition, a set of metrics were employed to assess the recognition rate and compare the outcomes among machine learning classifiers. The summary of these metrics components is presented in Table 3. Figure 13 offers a comprehensive depiction of the developed ML model.

Tab. 3. Results of classifiers using several features

Method	Accuracy	Recall	Precision
Logistic Regression(LR)	0.9648123	0.9870125	0.9317851
k-nearest neighbors (KNN)	0.9466214	0.9325617	0.9198507
Support Vector Machine (SVM)	0.9589132	0.9649082	0.9210563
Naive Bayes(NB)	0.9248953	0.9130687	0.8970356
Random Forest(RF)	0.9599387	0.9301855	0.9766541
Proposed Approach	0.9871257	0.99401562	0.9731854

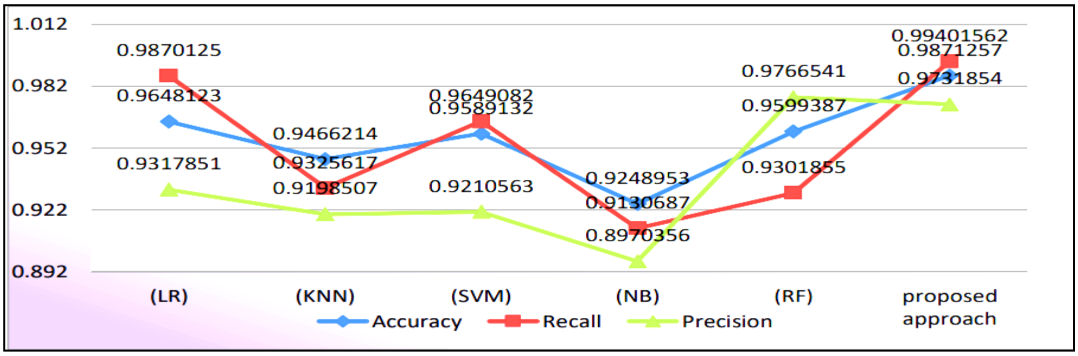


Fig. 13. Results of classifiers using several features

The technique we propose aids in the automatic detection of diseases and yields improved outcomes in the detection of breast cancer. Ultimately, the proposed method demonstrates a significant level of Recall, specifically 99.40%. The results demonstrate a notable precision of 97.31% and a high accuracy of 98.7%. To clarify, Table 4 presents a comparison between the proposed approaches and some of the existing methods (State of the art).

Tab. 4. Comparison to prior research

Reference	Dataset	Result – Accuracy	Pros	Cons
(Wang et al., 2019)	322 MIAS	93.36%	Comparisons are made between several ML classifiers	The number of features used is reduced
(Desai & Shah, 2021)	INBreast IRMA dataset	95.00%	Address the issue of over-fitting	High degree of complexity
(Rajakumari & Kalaivani, 2022)	322 MIAS images	96.56%	applicable to a dataset of photographs	Get lower value in confusion matrix
(Alanazi et al., 2021)	340 images (including 40 full-field digital images, 100 computed radiography images, and 200 MIAS images)	96.75%	Contrast is limited	This model was evaluated on a small dataset
(Mahmood et al., 2020)	322 MIAS images	95.71%	Big data sets and image data can also be utilized	Fewer features are used
(Melekoodappattu et al., 2023)	740 Digitalized Mammograms	97.00%	There is complexity, but much greater accuracy	It is not suitable for all phases of breast cancer
Proposed Approach	1400 digital mammography.	98.70%	Enhance the rate of cancer detection in regions containing hidden cancer masses and reduce error rates and enhance accuracy	It's hard to get training and testing datasets

For the reasons laid out in Table 4, the suggested approach demonstrates higher accuracy compared to some existing methods. This makes it a viable alternative for breast cancer diagnosis in practical scenarios. In addition, as can be shown in Table 5, our proposed solution has a shorter total computing time compared to the majority of previous methods.

Tab. 5. Comparison of Computational Time

Reference	Time – (ms)
(Wang et al., 2019)	1331.4
(Desai & Shah, 2021)	7000.0
(Rajakumari & Kalaivani, 2022)	4810.0
(Alanazi et al., 2021)	0368.4
(Mahmood et al., 2020)	0545.0
(Melekoodappattu et al., 2023)	0468.4
Proposed Approach	0340.0

5. CONCLUSION

Cancer mortality is highest for breast cancer. Early detection is essential for breast cancer. Different machine learning methods can diagnose breast cancer data. This study introduces Machine Learning-based breast cancer diagnosis. Using CNNs as a classifier model, features were selected and noise was reduced using CLAHE. This research also analyzes five algorithms: Random Forest, SVM, KNN, Naïve Bayes classifier, and Logistic Regression. To test the technique, 3002 merged photos were used. Data from 1400 digital mammography patients from 2007 to 2015 was included. Precision and accuracy measure system performance. Our model is efficient in simulations due to low-cost computer power and high accuracy. In the future, we can apply ensemble ML approaches to improve the performance of individual models and dimension reduction techniques to reduce the dimensions in datasets. To make substantial advancements in the detection and treatment of breast cancer, it is essential that researchers, data scientists, and medical professionals continue to work together.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing the original draft preparation, visualization, have been done by 1st and 2nd author. Review and editing have been done by 3rd.

Funding

The authors received no specific funding for this research.

Conflicts of Interest

Authors disclose no financial or non-financial interests in the subject matter, including honoraria, educational grants, speakers' bureaus, membership, employment, consultancies, stock ownership, expert testimony, patent-licensing arrangements, personal or professional relationships, affiliations, knowledge, or beliefs.

REFERENCES

Al-hazaimeh, O. M., Alomari, S. A., Alsakran, J., & Alhindawi, N. (2014). Cross correlation–new based technique for speaker recognition. *International Journal of Academic Research*, *6*(3), 232-239. https://doi.org/10.7813/2075-4124.2014/6-3/A.33

Al-hazaimeh, O. M., Abu-Ein, A. A., Tahat, N. M., Al-Smadi, M. M. A., & Al-Nawashi, M. M. (2022). Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images. *International Journal of Online & Biomedical Engineering*, *18*(13), 131-151. https://doi.org/10.3991/ijoe.v18i13.33985

Al-Hazaimeh, O. M., Al-Nawashi, M., & Saraee, M. (2019). Geometrical-based approach for robust human image detection. *Multimedia Tools and Applications*, *78*, 7029-7053. https://doi.org/10.1007/s11042-018-6401-y

Al-Hazaimeh, O. M., & Al-Smadi, M. (2019). Automated pedestrian recognition based on deep convolutional neural networks. *International Journal of Machine Learning and Computing*, *9*(5), 662-667. https://doi.org/10.18178/ijmlc.2019.9.5.855

Al-Nawashi, M., Al-Hazaimeh, O. M., & Saraee, M. (2017). A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. *Neural Computing and Applications*, *28*(1), 565-572. https://doi.org/10.1007/s00521-016-2363-z

Alanazi, S. A., Kamruzzaman, M., Islam Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021). Boosting breast cancer detection using convolutional neural network. *Journal of Healthcare Engineering*, *2021*, 5528622. https://doi.org/10.1155/2021/5528622

Alhindawi, N., Al-Hazaimeh, O. M., Malkawi, R., & Alsakran, J. (2016). A topic modeling based solution for confirming software documentation quality. *International Journal of Advanced Computer Science and Applications*, *7*(2), 200-206. https://doi.org/10.14569/IJACSA.2016.070227

Barrios, C. H. (2022). Global challenges in breast cancer detection and treatment. *The Breast*, *62*(1), S3-S6. https:/doi.org/10.1016/j.breast.2022.02.003

Carlson, R. W., Allred, D. C., Anderson, B. O., Burstein, H. J., Carter, W. B., Edge, S. B., Erban, J. K., Farrar, W. B., Forero, A., Giordano, S. H., Goldstein, L. J., Gradishar, W. J., Hayes, D. F., Hudis, C. A., Ljung, B. M., Mankoff, D. A., Marcom, P. K., Mayer, I. A., McCormick, B., … Zellars, R. (2011). Invasive breast cancer. *Journal of the National Comprehensive Cancer Network*, *9*(2), 136–222. https://doi.org/10.6004/jnccn.2011.0016

Chang, P. J., Asher, A., & Smith, S. R. (2021). A targeted approach to post-mastectomy pain and persistent pain following breast cancer treatment. *Cancers*, *13*(20), 5191. https://doi.org/10.3390/cancers13205191

Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN). *Clinical eHealth*, *4*, 1-11. https://doi.org/10.1016/j.ceh.2020.11.002

DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., Jemal, A., & Siegel, R. L. (2019). Breast cancer statistics, 2019. *CA: a cancer journal for clinicians*, *69*(6), 438-451. https://doi.org/10.3322/caac.21583

Fatima, N., Liu, L., Hong, S., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. *IEEE Access*, *8*, 150360-150376. https://doi.org/10.1109/ACCESS.2020.3016715

Gharaibeh, N., Abu-Ein, A. A., Al-hazaimeh, O. M., Nahar, K. M., Abu-Ain, W. A., & Al-Nawashi, M. M. (2023). Swin transformer-based segmentation and multi-scale feature pyramid fusion module for alzheimer's disease with Machine Learning. *International Journal of Online & Biomedical Engineering*, *19*(4), 22-50. https://doi.org/10.3991/ijoe.v19i04.37677

Gharaibeh, N., Al-hazaimeh, O. M., Abu-Ein, A., & Nahar, K. (2021). A hybrid svm naïve-bayes classifier for bright lesions recognition in eye fundus images. *International Journal on Electrical Engineering and Informatics*, *13*(3), 530-545. https://doi.org/10.15676/ijeei.2021.13.3.2

Gharaibeh, N., Al-Hazaimeh, O. M., Al-Naami, B., & Nahar, K. M. (2018). An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images. *International Journal of Signal and Imaging Systems Engineering*, *11*(4), 206-216. https://doi.org/10.1504/IJSISE.2018.10015063

Hall, K., Chang, V., & Mitchell, P. (2022). Machine learning techniques for breast cancer detection. *7th International Conference on Complexity, Future Information Systems and Risk COMPLEXIS* (pp. 116-122). SciTePress. https://doi.org/10.5220/0011123200003197

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. https://doi.org/10.1016/j.eswa.2020.114161

Kharya, S., Dubey, D., & Soni, S. (2013). Predictive machine learning techniques for breast cancer detection. *International journal of computer science and information Technologies*, *4*(6), 1023-1028.

Loibl, S., & Gianni, L. (2017). HER2-positive breast cancer. *The Lancet*, *389*(10087), 2415-2429. https://doi.org/10.1016/S0140-6736(16)32417-5

Lu, W., Jansen, L., Post, W., Bonnema, J., Van de Velde, J., & De Bock, G. (2009). Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis. *Breast Cancer Research and Treatment*, *114*, 403-412. https://doi.org/10.1007/s10549-008-0023-4

Ma'moun, A., Al-hazaimeh, O. M., Alhindawi, N., & Hayajneh, S. M. (2014). A dual curvature shell phased array simulation for delivery of high intensity focused ultrasound. *Computer and Information Science*, *7*(3), 49-57. https://doi.org/10.5539/cis.v7n3p49

Mahmood, T., Arsalan, M., Owais, M., Lee, M. B., & Park, K. R. (2020). Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *Journal of clinical medicine*, *9*(3), 749. https://doi.org/10.3390/jcm9030749

Melekoodappattu, J. G., Dhas, A. S., Kandathil, B. K., & Adarsh, K. (2023). Breast cancer detection in mammogram: Combining modified CNN and texture feature based approach. *Journal of Ambient Intelligence and Humanized Computing*, *14*, 11397-11406. https://doi.org/10.1007/s12652-022-03713-3

Nahar, K., Al-Hazaimeh, O., Abu-Ein, A., & Gharaibeh, N. (2020). Phonocardiogram classification based on machine learning with multiple sound features. *Journal of Computer Science*, *16*(11), 1648-1656. https://doi.org/10.3844/jcssp.2020.1648.1656

Nahar, K., Alhindawi, N., Al-Hazaimeh, O., Alkhatib, R., & Al-Akhras, A. (2018). NLP and IR based solution for confirming classification of research papers. *Journal of Theoretical and Applied Information Technology*, *96*(16), 5269-5279.

Nallamala, S. H., Mishra, P., & Koneru, S. V. (2019). Breast cancer detection using machine learning approaches. *International Journal of Recent Technology and Engineering*, *7*(5S4), 478-481.

Nanda, K., Bastian, L. A., & Schulz, K. (2002). Hormone replacement therapy and the risk of death from breast cancer: a systematic review. *American journal of obstetrics and gynecology*, *186*(2), 325-334. https://doi.org/10.1067/mob.2002.121077

Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V., & Sun, P. (2015). Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA oncology*, *1*(7), 888-896. https://doi.org/10.1001/jamaoncol.2015.2510

Nguyen, C., Wang, Y., & Nguyen, H. N. (2013). Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. *Journal of Biomedical Science and Engineering*, *6*, 551-560. https://doi.org/10.4236/jbise.2013.65070

Rajakumari, R., & Kalaivani, L. (2022). Breast cancer detection and classification using deep CNN techniques. *Intelligent Automation & Soft Computing*, *32*(2), 1089-1107. https://doi.org/10.32604/iasc.2022.020178

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, *38*, 35-44. https://doi.org/10.1023/B:VLSI.0000028532.53893.82

Rivera-Franco, M. M., & Leon-Rodriguez, E. (2018). Delays in breast cancer detection and treatment in developing countries. *Breast cancer: basic and clinical research*, *12*. https://doi.org/10.1177/1178223417752677

Sivapriya, J., Kumar, A., Sai, S. S., & Sriram, S. (2019). Breast cancer prediction using machine learning. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(4), 4879-4881. https://doi.org/10.35940/ijrte.D8292.118419

Svensson, B., Dylke, E., Ward, L., Black, D., & Kilbreath, S. L. (2020). Screening for breast cancer–related lymphoedema: self-assessment of symptoms and signs. *Supportive Care in Cancer*, *28*, 3073-3080. https://doi.org/10.1007/s00520-019-05083-7

Tagliafico, A. S., Piana, M., Schenone, D., Lai, R., Massone, A. M., & Houssami, N. (2020). Overview of radiomics in breast cancer diagnosis and prognostication. *The Breast*, *49*, 74-80. https://doi.org/10.1016/j.breast.2019.10.018

Tanabe, K., Ikeda, M., Hayashi, M., Matsuo, K., Yasaka, M., Machida, H., Shida, M., Katahira, T., Imanishi, T., Hirasawa, T., Sato, K., Yoshida, H., & Mikami, M. (2020). Comprehensive serum glycopeptide spectra analysis combined with artificial intelligence (CSGSA-AI) to diagnose early-stage ovarian cancer. *Cancers*, *12*(9), 2373. https://doi.org/10.3390/cancers12092373

Tiwari, M., Bharuka, R., Shah, P., & Lokare, R. (2020). Breast cancer prediction using deep learning and machine learning techniques. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3558786

Vaka, A. R., Soni, B., & Reddy, S. (2020). Breast cancer detection by leveraging Machine Learning. *Ict Express*, *6*(4), 320-324. https://doi.org/10.1016/j.icte.2020.04.009

Vasundhara, S., Kiranmayee, B., & Suresh, C. (2019). Machine learning approach for breast cancer prediction. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(1), 2619-2625.

Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., & Xin, J. (2019). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*, *7*, 105146-105158. https://doi.org/10.1109/ACCESS.2019.2892795

Wilkinson, L., & Gathani, T. (2022). Understanding breast cancer as a global health concern. *The British Journal of Radiology*, *95*(1130), 20211033. https://doi.org/10.1259/bjr.20211033

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. https://doi.org/10.1016/j.cmpb.2017.12.012