Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Classification of Breast cancer using Different Feature selection techniques with Ensemble Machine Learning Methods

Shruthi B. S.[1] and Ramesh S.[2]

[1] Research Scholar, Department of CSE,Jain University Bangalore
[2]*School of Computer Science and Engineering, JAIN (Deemed- to-be- University), Bangalore*

Abstract. Breast cancer is any malignant tumor that is found in the breast region. This cancer is one of the main causes of death among women. Hence, early identification of breast cancer using diagnostic Mam- myography and screening increase survival rates and the options available for breast cancer treatment. Manual screening and traditional laboratory techniques for detecting breast cancer are prone to error. Hence, classifier schemes based on image analysis are extensively used in medical diag- nosis. Consequently, multiple classifier algorithms have been smeared on medical datasets to predict breast cancer tumors. Advanced techniques like machine learning, which is a subset of artificial intelligence, are utilized to forecast breast cancer tumors. Also, the high number of image features in the breast mammograms decreases the speed of the classification process done with machine learning techniques. Therefore, feature selection procedures are implemented to reduce the computation cost and increase the performance of the classifier. When several machines learning approaches are integrated, then it is termed ensemble learning. How- ever, since there is no extensive literature that concentrates on ensemble machine learning techniques and various feature selection approaches utilized in the Classification of breast cancer. Hence, the current study attempts to review the steps involved in breast cancer diagnosis and Classification of Breast cancer using diverse Feature selection techniques with Ensemble Machine Learning methods. Finally, the study provides suggestions for improving the accuracy of breast cancer diagnosis using advanced machine-learning approaches

Keywords: Breast cancer**,** mammograms**,** Feature Selection, Feature Extraction, Machine Learning and Ensemble learning

1 Introduction

One of the main causes of female deaths worldwide is cancer. Breast cancer is one of the cancers that kills the most women out of all the ones there are. [1]. One in eight and one in twelve women in the modern world will develop breast cancer during their lifetime. [2]. Hence, the breast cancer diagnosis is very significant. The early detection of breast cancer will save millions of lives. Even though breast cancer diagnosis has been implemented for over forty years utilizing ultrasound, Magnetic Resonance Imaging, X-ray, and so on[3], biopsy procedures are having said that the main approaches used to detect breast cancer in a precise manner. Based on image analysis, a Breast cancer diagnosis has been investigated for the past few decades, and there have been various prominent investigation accomplishments in that region. Initially, the clinical images are gathered from a database which contains background noise and unessential information. The preprocessing procedure is usually used to segregate these from the mammogram and precisely create the image for further progression [4]. Women who are at increased risk of breast cancer can be examined through Mammography. This is the most utilized technique in hospitals and clinics [5]. To monitor high-risk breast cancer patients, the Magnetic Resonance Imaging(MRI) method can be executed along with the Mammography[6].

Conventional laboratory approaches such as MRI and Computed Axial Tomography have been considered to be beneficial. On the other hand, it delivers very scant information regarding the cancer's progression mechanism. Conversely, the improvements in Deoxyribonucleic acid microarray technology have delivered maximal throughput samples of gene expression. Machine Learning (ML) technique has been employed to identify breast cancer treatment or continued existence [7]. This ML approach has turned out to be a significant part of medical imaging investigation. These approaches have transformed over the decades from manual broadcasted inputs to programmed initializations. Improvements in the arena of machine learning have resulted in enhanced self-reliant and intelligent computer-aided diagnosis systems since the learning capability of machine learning approaches has been continually developing. Numerous automated procedures evolving with deep feature learning and representations are emerging in advanced times[8]. These machine-learning approaches will detect whether cancer is benign or malignant [9].

Various ML approaches are utilized to identify, organize, and detect breast cancer based on characteristics mined from medical imaging. The choice of proper image enhancement, segmentation, feature selection, feature extraction and forecasting or classification algorithm is regarded as the essential step in accurate cancer diagnosis upon mammograms, and it is one of the primary tasks in the field investigation. Classification techniques forecast the class label for unlabeled datasets based on their proximity to the learnt configuration [10]. Presently, the advanced approach of combining several ML algorithms, called ensemble learning, is being employed in detecting breast cancer. Various investigations examined the well-established Machine learning approaches embraced for every step involved in breast cancer diagnosis, such as Classification, identification and the segmentation process. Additionally, various scholarly works discuss and review the usage of several imaging modalities for cancer diagnosis. Also, several researchers deliberated and equated traditional ML algorithms employed for breast cancer detection. However, only a few studies attempt to review advanced machine learning techniques such as ensemble learning and the significance of feature selection and feature extraction procedures in the Classification of Breast cancer. To resolve these gaps, the present study intends to review the existing scholarly works regarding the Classification of breast cancer utilizing various feature extraction and feature selection approaches with advanced machine learning techniques like ensemble learning. Finally, this review provides suggestions for improving the accuracy of breast cancer diagnosis using advanced ML techniques.

2 Breast cancer- stages

The breast tumour is generally of two major kinds: malignant and benign. Among these 2 kinds, a benign lesion is not tumorous; it is some abnormality in the cell, and it is not the reason for breast cancer. Meanwhile, the malignant lesions are cancerous. The spreading of malignant cells takes place at a rapid level, leading to the progress of breast cancer. Both malignant and benign cells have irregular structures and appearances; hence, it is very difficult to differentiate between them using manual methods. It is challenging for radiologists to accurately categorize malignant and benign patterns in digital mammograms[11]. During its early phase, the pictorial clues are understated and vary in appearance, thus making the diagnosis of breast tumours difficult for specialists and doctors. Hence, an intelligent classifier is needed that can aid radiotherapists in categorizing apprehensive areas and detecting breast cancer.

Consequently, an academic work [12] examines the direct learning classifier for the soft clustered-based detection of malignant areas in mammograms. The presented approach has achieved 97% grouping exactness on the test set. This is generally higher when contrasted with the standard multi-facet Perceptron classifier and different winning methodologies. Another work [13] executed 3 expectation models for bosom disease survivability on 2 boundaries, like threatening and harmless malignant growth patients. This work utilized 3 popular information mining procedures: Radial Basis Function Network, Naive Bayes, and a decision tree (J48). The results of this examination uncovered that the Gullible Bayes shows a brilliant grouping precision of 97.36 percent when contrasted with different models, for example, outspread premise capability and choice tree.

Table 1. Classification of cancer stage as per the survival rate

(Source: [14])

Classification	5-year survival rate (%)	Stages of cancer
In situ	100^{*3}	0
Early invasive	$98.0\ (local)^{*3}$	I, IIa, IIb

Classification	5-year survival rate (%)	Stages of cancer
	83.6 $(regional)^{*3}$	
Locally advanced †	57^{4}	Ia, Ib, Ic IV
Metastatic †	23.4^{*3}	

Table 1 depicts the various stages of cancer, and it is explained as follows,

- **Stage 0**: Lobular carcinoma In situ is a coincidental minute finding of strange tissue development in the breast lobules. The findings of the study[15] reveal that the minimal risk of increasing Invasive Breast cancer followed by the occurrence of Lobular carcinoma In situ is seven per cent at ten years.
- **Stages I and II:** For early-stage Invasive breast cancer, surgery-modified radical mastectomy has been a widely utilized treatment. Also, breast-conserving surgery is been suggested[16].
- **Stage III and IV**: Five-year survival rates are better for the initial stage of barest cancer, such as Stage 0 and Stages I and II, when compared to the advanced stages like III and IV [17].

2.1 Diagnosis types

Various approaches for breast cancer diagnosis and evolving sophisticated techniques for improving identification accuracy have been varied for the past few years. Among those approaches, Mammography is the most widely adopted approach for screening breast cancer. It is utilized to identify the presence of micro-calcification, cysts and cancerous tumours in the breast [18]. Figure 1 denotes the usual specimen of the mammograms.

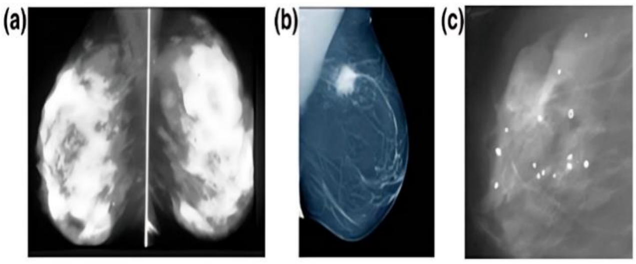

Figure 1. (a)Dense breast Mammogram (b) Mammogram representing a breast tumor (c) Micro calcifications present in breast mammogram (Source: [19])

Another investigation [20] reveals that Thermography, when utilized with well-established protocols, can recognize the initial symptoms of cancer for around eight to ten years before Mammography. Also, it is found that compared to Mammography or even the aggregated usage of high-frequency breast ultrasound and Mammography, surveillance with Magnetic Reasoning Imaging permits the initial diagnosis of hereditary breast cancer [21]. One more significant part of the enhancement of initial breast cancer identification is Thermography. This is not an alternative to Mammography; instead, it is a unique method. Therefore, a system which merges Thermography, Mammography and clinical examination can surge the possibility of early detection of Breast cancer. This thermal imaging can be utilized as the pre-screening diagnostic approach that will decrease the number of females who must go through the mammography screening[22]. In recent times, ML-based interpretable diagnosis systems have been capable of improving cancer diagnosis capabilities and reducing prediction inaccuracy [23].

3 Datasets of Breast Cancer

An interpreted database is significant for improving and endorsing ML systems for classifying breast cancer. These breast mammography images can aid in increasing the classification accuracy while integrated with ML methods. A Conventional study[24] employs the publicly available data of breast cancer tumours from the Wisconsin Hospital and applies feature selection and ML algorithms such as Naïve Bayes, C4.5 decision tree, and Support vector machine for Classification of Breast cancer. This study performs the breast cancer diagnosis by utilizing this Wisconsin Diagnostic Breast Cancer (WDBC)[25]. The results of the review uncover that the C4.5 choice tree achieves the ideal results in the finding of Bosom malignant growth when contrasted with the Help vector machine and Naïve Bayes. Similarly, another inquiry [26] utilizes the Wisconsin Breast Cancer dataset. This dataset is openly available at the University of California Irvine Machine Learning Repository[27]. This data set is used to distinguish the malignant tumour from the malignant.

Another investigation [28] introduces the dataset for the classification of histopathological images of breast cancer (BreakHis). This dataset is made out of 7,909 bosom malignant growth histopathology pictures from 82 patients with different amplification factors. Similarly, another study[29] performed extensive experiments on the diagnosis of breast cancer using datasets from the 2012 and 2014 challenges of the International Conference on Pattern Recognition. One more information base gained from the ICIAR 2018 Thousand Test is Bosom Malignant growth Histopathology Pictures [30]. There are 400 images in this database, and they fall into four distinct categories: invasive carcinoma, in situ carcinoma, benign carcinoma, and normal carcinoma. The high-goal pictures are digitized mechanized with an amplification variable of 200 X and indistinguishable circumstances. Subsequently, an inquiry[31] utilizes a small number of breast cancer histopathological images from the Bio-Image Semantic Query User Environment dataset. The photograph size is 896 ×768 pixels. It has images that are both harmful and good.

4 Preprocessing Methods

Preprocessing is needed to transform the image into a less complicated and effective form. It decreases the image's noise and aids in recognizing the focal regions. Data transformation and preprocessing are usually needed before implementing data mining over medical data, specifically the data compilation by utilizing information from the data itself [32].In this preprocessing method, the undesirable objects are segregated from the mammograms that involve background noises, labels and annotations, which is clearly illustrated in Figure 2 [33]. The noise is reduced, and the mammary-processing method enhances the mammogram image

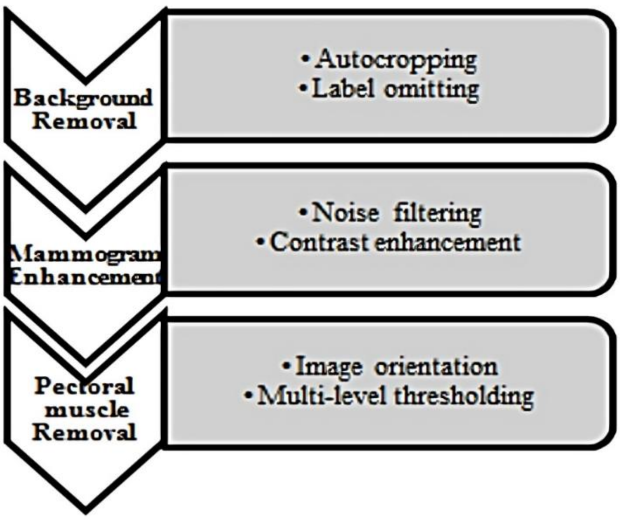

Figure 2. Steps involved in preprocessing of mammogram images

(Source: [33])

Likewise, another investigation [34] executes a four-step image preprocessing method that includes the bounding box identification, which consists of the tumour region. Through expanding and mining the bounding box, standardizing the intensity of the image and normalizing the image, the research acquired an accuracy of over ninety per cent. Furthermore, an inquiry [35] introduces efficient preprocessing techniques to produce datasets, which can reduce computational time for the neural network and enhance classification rates and accuracy. This study introduces the Huangs Fuzzy Thresholding and Rolling Ball algorithm for background removal. This introduced method accomplishes the background removal of the images. Also, for the removal of pectoral muscle, this investigation employs Hough's line transform and canny edge Detection method.

5 Feature Extraction and Selection Methods

The foremost procedure in medical image processing is the extraction of features. It entails the selection of specific parameters called features from a breast thermogram, which are then analyzed and compared to derive meaningful outcomes [36]. The research aims to propose an approach for extracting features that can aid in the Classification of breast abnormalities through Thermography. The recommended strategy involves utilizing an interval symbolic representation of image data, which takes into account the inherent variability in the data [37]. The conventional approach can significantly enhance the accuracy of Classification by utilizing interval data. The interval metaphorical feature extraction procedure contains a three-stage process. The suggested technique yields satisfactory results for the global misclassification rate and sensitivity for the malignant class, demonstrating a high level of performance in comparison with other relevant exams, including Mammography [38]

In a paper, they have created a diagnosing system based on support vector machines (SVM) that involves three stages. In the first stage, we use principal component analysis to remove redundant information and extract representative patterns from the original data [39]. By reducing the dimension of the feature space, the tried-and-true approach aims to make computations simpler. In the following stage, they apply the differential advancement calculation to find the ideal boundary values for SVM. They then, at that point, train a classifier to recognize approaching growths. They look at classification precision, acuity, explicitness, and the area under the receiver operating characteristic curves to estimate the classifier's effectiveness simultaneously. When contrasted with other assortment strategies like K-closest neighbor, irregular woods, packing, gullible bayes, and choice tree, their proposed method exhibits predominant execution [24, 40]. They conducted fivefold cross-validation tests on the Wisconsin Diagnostic Breast Cancer (WDBC) set of data from the University of California [41].

The significant procedures in the Classification of breast cancer are feature selection and feature extraction. It is a well-established fact that the malignant and benign lesions vary in terms of pixel values and shape[42]. It is essential to choose the features associated with the boundary and shape of the lesion. A significant approach to determining a subset of suitable features in constructing a model is called feature selection. Feature selection approaches are extensively utilized in model generalization for easier elucidation, preventing over-fitting issues, handling dimensionality and decreased training times [43]. Various feature selection approaches are employed and have been improved for the motive of the predictive model's optimization. Usually, there are 2 kinds of feature selection, such as wrapper and filtering approach. Due to its significance, an existing study[44] employs the crow search algorithm for feature selection in multifaceted problems. The results obtained using the crow search algorithm are compared with the outcomes of several algorithms. The outcomes of the simulation exhibit that this crow search algorithm will result in optimal results.

To accurately and quickly detect breast cancer, the introduced method employs a combination of supervised (Relief algorithm) and unsupervised (Autoencoder, PCA algorithms) techniques to select relevant features from a dataset[45, 46]. These selected features are then utilized to train and test a support vector machine classifier. The methodology additionally integrates k-overlay cross-approval for model approval and determination of ideal hyperparameters. The proposed method employs BC datasets for testing, and model performance evaluation metrics are used to evaluate the model's performance. Trial results show that highlights chosen by the Alleviation calculation are more successful in recognizing bosom malignant growth than those chose by the Autoencoder and PCA calculations. On certain features selected by the Relief algorithm, the proposed method achieves a remarkable accuracy of 99.91 percent. [45].

6 Classification using Machine learning methods

The research aims to anticipate breast cancer, which is the second most typical cause of mortality among women globally. Recognizing and preventing it in its early phases can enormously decrease the risk of fatality. To determine the most efficient algorithm, this investigation utilized four machine learning techniques - Random Forest, Naïve Bayes, Support Vector Machines (SVM), and K-Nearest Neighbors (K-NN)[47]. The findings displayed that SVM delivered the highest accuracy rate at 97.9%. This in turn lead to the discovery which will aid in choosing the most suitable machine-learning algorithm for predicting breast cancer [48].

Diverse medical applications have utilized machine learning approaches as valuable tools to support physicians in making informed decisions based on available data and in the development of medical expert systems. In this investigation, five of the most popular nonlinear machine-learning algorithms were utilized to notice breast cancer[49]. Each algorithm's methodology and key features were explained, and the Wisconsin Breast Cancer Diagnostic (WBCD) dataset was utilized as a benchmark to resemble the presented models' implementation, including MLP, KNN, CART, SVM, and NB. The MLP algorithm demonstrated the highest precision of 96.70% on the training data, outperforming the other four algorithms. The models' performance was assessed based on accurateness, precision, and recall of unseen data after estimation, and the k-fold cross-validation technique was used to test their capabilities. The consequences of introspection confirmed the MLP model's superior performance, achieving accuracy, precision, and recall scores of 99.12%, 99.00%, and 99.00%, accordingly [50].

The review meant to examine the adequacy of five AI calculations, in particular Help Vector Machine, Arbitrary Timberland, Strategic Relapse, Choice Tree (C4.5), and K-Closest Neighbors (KNN), on the Bosom Disease Wisconsin Demonstrative dataset [51]. Following this, we performed a comprehensive performance evaluation and comparison of these classifiers, with a primary focus on forecasting and diagnosing breast cancer using machine learning techniques. The study sought to identify the most effective algorithm based on the confusion matrix, accuracy, and precision. Notably, the Support Vector Machine outperformed all other classifiers, achieving the highest accuracy rate of 97.2\% [51].

Another study aims to assess the effectiveness of various Machine Learning classifiers in predicting breast cancer. Six commanded classification methods, i.e. SVM, NB, KNN, RF, DT, and LR, were operated to classify breast cancer disease. Sensitivity, specificity, f1 measure, and overall accuracy were employed to assess the breast cancer dataset[52, 53]. The effects indicate that SVM had the most elevated performance, with a classification precision of 97.07%. NB and RF also demonstrated high precision and came in second position. These consequences recommend that creating a predictive system based on machine learning could aid in the earlier detection of breast cancer, potentially diminishing its incidence [52].

7 Comparative analysis

This section performs the comparative analysis of various well-established and ensemble ML techniques employed in the Classification of breast cancer, and it is illustrated in Table 2.

Table 2. Various classification techniques and feature selection /extraction techniques are used in Breast cancer detection

Author	Dataset utilized	Feature selection/ Feature extraction technique	Classification method	Accuracy	Limitations Future Recommendations
[54]	Kaggle-Breast Cancer Dataset (KBCD)	Chi - Square	Tropical convolution Neural Networks (TCNNs)	99.46 % ,	In future, they need .to apply TCNNs to diverse deep neural networks
[55]	DDSM and IN breast	FDCT - WRP (FE)	MODPSO SVM	DDSM - 98.94 % MIAS - 98.76 %	In future, they will examine the efficiency of the existing design to enhance speculation with a few imaging modalities

					like X ray, biopsy .ultrasound , and so on .
[56]	Digital Database for Screening Mammog-ra-phy(DDSM)	---	Hybrid (ShuffleNet ResNET)	98 %	---
[57]	Mammog-raphy Im-age Analy-sis Da-taset(MIAS) INbreast , DDSM	ResNet - 18	Extreme Learning Machine(ELM)	INbreast 98.27 % MIAS - 98.14 % DDSM - 97.19 %	In future, they using suggest various preprocessing techniques with
[58]	(MIAS)	Texture, Shape, Statistic feature	SVM	92.50 %	Researchers aim to enhance the proposed system for diagnosing breast cancer by us-ing GMM on the entire Mini - MIAS da-taset .
[59]	Wiscon-sin's breast can-cer data reposi-tory .	Grid approach search	optimized KNN model	94.35 % .	---
[60]	Mammo-gram Image data set	relief algo-rithm	LS - SVM , KNN , Random For-est and Naive Bayes	Approx . 97 % by LS - SVM	---
[61]	Image Analysis Society Digital	A joined dark level grid and co-event (GLCM)di	Advanced Thermal Exchange Optimizer	93.79 %	---

	Mammo-gram Database	screte wavelet change (DWT) technique			
[62]	The informational index utilized in this study was gathered between Walk 2012 and Walk 2018at Jeonbuk National University Hospital (JNUH) .	Inception -v3 architecture modified	convolutional neural network (CNN)	88 %	In future , it 13 recommended to assess this work with more robust classification algorithms for detection of breast cancer
[25]	Wisconsin Diagnostic Breast Cancer (WDBC) Dataset .	Linear discriminant analysis	Radial kernel),ANN and Naïve Bayes " SVM (using basis	98.82 %	This work can be . extended by comparing several ML techniques for diagnosis of breast cancer .
[63]	Breast Ultra-soundImages(BUSI) dataset	reformed differential evaluation(RDE) and reformed gray wolf (RGW	machine learning algorithms	99.1 %	The database size can be increased in future for classification of cancer breast tumour .
[64]	METABRIC dataset	convolutional neural network	stacked – based ensemble machine learning model	90.2 %	The METABRIC dataset is very small and it is the main limitation of this study .
[65]	BreaKHis dataset	Dense-Net201	XGBoost	97 %	Later on , a few additional spaces and headings can be investigated to work on this technique .

					These include examining this method's memory and time efficiency . Plus , assessing this technique on one more comparable dataset can likewise be performed to affirm this strategy further .
[66]	Wisconsin Breast Cancer (WDBC)	" Correlation - based selection , Information Gain based selection and Sequential feature selection "	ensemble - based Max Classifier Voting	99.41 % .	In future, AI procedures can be coordinated with neuro - fluffy and profound learning approaches for productive finding. By machine learning combining CNN / Auto - encoder, neuro - fuzzy systems, and swarm optimizations to integrate optimal feature selection , feature extraction , and classification for medical imaging data , new hybrid models can be proposed .
[67]	" Breast Cancer Wisconsin (Diagnostic) Data Set "	" Particle Swarm Optimization , Genetic Search and Greedy Stepwise "	(C4.5 Decision Tree Algorithm) .Support Vector Machine , J48 Multilayer Perceptron (a feed - forward ANN) "	" J48 Decision Tree classifier with the Genetic Search algorithm for feature selection achieves accuracy 98.83 % "	" The future scope of work includes the prognosis of breast cancer using thermal images and IoT - based sensors " .

Accuracy metrics are the most widely embraced approach for performance assessment of breast cancer classification models. From the outcomes of table 2, it is revealed that the introduced Tropical convolutional Neural Networks for detection of breast cancer attains the highest accuracy of 99.46%. Also, it is revealed that the WDBC and MIAS and DDSM are the widely employed datasets for classification of breast cancer using various feature selection techniques and ensemble Machine learning technique.

8 Discussion

The study's extensive review revealed that the machine learning algorithm and an efficient method for image preprocessing, segmentation, feature selection, and extraction improved the accuracy of breast cancer predictions. In addition, it is shown that the accuracy of predictions for the classification of breast cancer is improved by combining a number of ML algorithms into a single unit. Similarly, an existing study [68] induces that the impediment of a solitary model exhibition can be diminished by the troupe AI calculation inside the setting of bosom malignant growth determination. The six principal steps associated with bosom malignant growth determination utilizing AI procedures are stacking pictures from the bosom disease dataset, preprocessing, division, include extraction, highlight choice lastly, the Order of bosom disease cancers. Also, conventional study [69] utilizes the WDBC dataset, and it is a broadly used dataset for bosom malignant growth determination with the guide of the ML approach. The effectiveness of the characterization strategies engaged with the expectation of bosom malignant growth can be improved by utilizing proper element choice methods and a reasonable dataset.

The ideal machine learning classifier is identified and improved by means of experimentation with the suitable medical dataset available for breast cancer diagnosis[70]. An academic work [71] performs the detection of breast cancer on various breast cancer datasets such as Wisconsin breast cancer, Wisconsin prognosis breast cancer, Mammographic Mass dataset, and WDBC. By choosing the proper feature selection technique and dataset, breast cancer prediction accuracy can be increased using the ML classification technique. ML techniques have been employed in categorizing medical data for the past few years. Even though various approaches have been improved to detect breast cancer tumours, various issues like accuracy exist. In order to resolve this gap, a conventional study[72] utilizes the ensemble approach to classify the breast cancer tumour. Likewise, another study [73] suggested an ensemble feature selection technique approach on the basis of mean criteria for selecting appropriate features for classifying breast cancer data.

Moreover, this review identifies the commonly used ML techniques in the Classification of breast cancer, such as SVM, convolutional neural network, decision tree, ensemble learning, etc. Correspondingly, another study [74] reveals that the quadratic SVM achieves the highest accuracy in the selection of breast cancer. The recent advanced ML approach is the Deep Learning (DL). Deep learning has greater potential within the context of image processing and analysis. Also, this advanced ML technique has a wider scope in the Classification of breast cancer [75]. In future, various large datasets can be utilized for the Classification of breast cancer using advanced ML approaches [76]. The outcomes of scholarly work reveal that the ensemble ML approach attained better outcomes with respect to the Classification of breast cancer when compared to other traditional machine learning techniques [70, 77-79]. Consequently, a recent study [80] strongly recommends integrating ensemble ML approaches and DL algorithms to increase the detection accuracy of Invasive breast cancer detection.

9 Conclusion

This review assesses and surveys the ML procedure, troupe draws near, and different element choice methods utilized for the arrangement of bosom disease growths. A robust classification system for breast cancer is built on this, in turn. Likewise, this audit recognizes the meaning of component choice and element extraction approach in the order cycle of bosom disease information utilizing the outfit ML approach. The arrangement precision of bosom disease can be upgraded by utilizing progressed ML approaches alongside the appropriate picture improvement procedures. Advanced machine learning methods like deep learning and ensemble learning have recently outperformed the other models that predict breast cancer tumors. The results of this audit can be helpful to examiners who are attempting to help radiologists by expanding the exactness of the discovery of bosom malignant growth. Different datasets are utilized for the distinguishing proof of bosom malignant growth, and a few existing investigations used little datasets for the expectation of bosom disease. The consequences of the audit unequivocally display that the group AI strategy, alongside the component determination approach, beats other customary methodologies in the forecast of bosom malignant growth information. Later on, this survey prescribes using huge datasets to analyze bosom malignant growth utilizing progressed ML strategies like profound learning.

References

1. C. Fitzmaurice *et al.*, "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study," *JAMA oncology,* vol. 4, no. 11, pp. 1553-1568, 2018.

2. D. G. R. Evans and A. Howell, "Breast cancer risk-assessment models," *Breast cancer research,* vol. 9, no. 5, pp. 1-8, 2007.

3. B. Stenkvist *et al.*, "Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations," *Cancer research,* vol. 38, no. 12, pp. 4688-4697, 1978.

4. R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement,* vol. 146, pp. 800-805, 2019.

5. G. A. Klevos *et al.*, "A breast cancer review: Through the eyes of the doctor, nurse, and patient," *Journal of radiology nursing,* vol. 36, no. 3, pp. 158-165, 2017.

6. K. S. Peairs, Y. Choi, R. W. Stewart, and H. F. Sateia, "Screening for breast cancer," in *Seminars in Oncology,* 2017, vol. 44, no. 1: Elsevier, pp. 60-72.

7. F. Cardoso *et al.*, "70-gene signature as an aid to treatment decisions in early-stage breast cancer," *New England Journal of Medicine,* vol. 375, no. 8, pp. 717-729, 2016.

8. S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," *Journal of medical Internet research,* vol. 21, no. 7, p. e14464, 2019.

9. I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in medicine,* vol. 23, no. 1, pp. 89-109, 2001.

10. J. G. Melekoodappattu and P. S. Subbian, "Automated breast cancer detection using hybrid extreme learning machine classifier," *Journal of Ambient Intelligence and Humanized Computing,* vol. 14, no. 5, pp. 5489-5498, 2023.

11. S. Ara, A. Das, and A. Dey, "Malignant and benign breast cancer classification using machine learning algorithms," in *2021 International Conference on Artificial Intelligence (ICAI)*, 2021: IEEE, pp. 97-101.

12. B. Verma, P. McLeod, and A. Klevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer," *Expert systems with applications,* vol. 37, no. 4, pp. 3344-3351, 2010.

13. Saravanakumar, S., & Thangaraj, P. (2019). A computer aided diagnosis system for identifying Alzheimer's from MRI scan using improved Adaboost. Journal of medical systems, 43(3), 76.

14. Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. Cluster Computing, 22(Suppl 1), 33-46.

15. Saravanakumar, S., & Saravanan, T. (2023). Secure personal authentication in fog devices via multimodal rank-level fusion. Concurrency and Computation: Practice and Experience, 35(10), e7673.

16. Thangavel, S., & Selvaraj, S. (2023). Machine Learning Model and Cuckoo Search in a modular system to identify Alzheimer's disease from MRI scan images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(5), 1753-1761.

17. Saravanakumar, S. (2020). Certain analysis of authentic user behavioral and opinion pattern mining using classification techniques. Solid State Technology, 63(6), 9220-9234.

18. V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *Journal of Algorithms & Computational Technology,* vol. 12, no. 2, pp. 119-126, 2018.

19. K. L. Maughan, M. A. Lutterbie, and P. S. Ham, "Treatment of breast cancer," *American family physician,* vol. 81, no. 11, pp. 1339-1346, 2010.

20. P. J. Chuba *et al.*, "Bilateral risk for subsequent breast cancer after lobular carcinoma-in-situ: analysis of surveillance, epidemiology, and end results data," *Journal of Clinical Oncology,* vol. 23, no. 24, pp. 5534-5541, 2005.

21. M. Clarke, "Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials," *Lancet,* vol. 365, pp. 1687-1717, 2005.

22. E. T. Warner *et al.*, "Time to diagnosis and breast cancer stage by race/ethnicity," *Breast cancer research and treatment,* vol. 136, pp. 813-821, 2012.

23. T. B. Borchartt *et al.*, "Thermal feature analysis to aid on breast disease diagnosis," in *Proceedings of 21st Brazilian Congress of Mechanical Engineering—COBEM*, 2011, pp. 24-28.

24. D. Singh and A. K. Singh, "Role of image thermography in early breast cancer detection-Past, present and future," *Computer methods and programs in biomedicine,* vol. 183, p. 105074, 2020.

25. L. U. YK Ng, FC Ng, LSJ Sim, E, "Statistical analysis of healthy and malignant breast thermography," *Journal of medical engineering & technology,* vol. 25, no. 6, pp. 253-263, 2001.

26. C. K. Kuhl *et al.*, "Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer," *Journal of clinical oncology,* vol. 23, no. 33, pp. 8469-8476, 2005.

27. M. Milosevic, D. Jankovic, A. Milenkovic, and D. Stojanov, "Early diagnosis and detection of breast cancer," *Technology and Health Care,* vol. 26, no. 4, pp. 729-759, 2018.

28. S. Wang, Y. Wang, D. Wang, Y. Yin, Y. Wang, and Y. Jin, "An improved random forest-based rule extraction method for breast cancer diagnosis," *Applied Soft Computing,* vol. 86, p. 105941, 2020.

29. A. T. Alhasani, H. Alkattan, A. A. Subhi, E.-S. M. El-Kenawy, and M. M. Eid, "A comparative analysis of methods for detecting and diagnosing breast cancer based on data mining," *Methods,* vol. 7, no. 9, 2023.

30. D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine learning classification techniques for breast cancer diagnosis," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 495: IOP Publishing, p. 012033.

31. E. A. Bayrak, P. Kırcı, and T. Ensari, "Comparison of machine learning methods for breast cancer diagnosis," in *2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT)*, 2019: Ieee, pp. 1-3.

32. N. Sharma, K. Sharma, M. Mangla, and R. Rani, "Breast cancer classification using snapshot ensemble deep learning model and t-distributed stochastic neighbor embedding," *Multimedia Tools and Applications,* vol. 82, no. 3, pp. 4011-4029, 2023.

33. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *Ieee transactions on biomedical engineering,* vol. 63, no. 7, pp. 1455-1462, 2015.

34. H. Chen, Q. Dou, X. Wang, J. Qin, and P. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Proceedings of the AAAI conference on artificial intelligence*, 2016, vol. 30, no. 1.

35. A. Sarmiento and I. Fondón, "Automatic breast cancer grading of histological images based on colour and texture descriptors," in *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15*, 2018: Springer, pp. 887-894.

36. I. Koné and L. Boulmane, "Hierarchical ResNeXt models for breast cancer histology image classification," in *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15*, 2018: Springer, pp. 796-803.

37. P. Lin, Z. Lei, L. Chen, J. Yang, and F. Liu, "Decision tree network traffic classifier via adaptive hierarchical clustering for imperfect training dataset," in *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, 2009: IEEE, pp. 1-6.

38. A. M. Omer and M. Elfadil, "Preprocessing of digital mammogram image based on otsu's threshold," *Am. Sci. Res. J. Eng. Technol. Sci*, vol. 37, no. 1, pp. 220-229, 2017.

39. F. Gao *et al.*, "SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 53-62, 2018.

40. A. R. Beeravolu, S. Azam, M. Jonkman, B. Shanmugam, K. Kannoorpatti, and A. Anwar, "Preprocessing of breast cancer images to create datasets for deep-CNN," *IEEE Access*, vol. 9, pp. 33438-33463, 2021.

41. A. A. Alhussan *et al.*, "Classification of breast cancer using transfer learning and advanced al-biruni earth radius optimization," *Biomimetics*, vol. 8, no. 3, p. 270, 2023.

42. P. Pramanik, S. Mukhopadhyay, S. Mirjalili, and R. Sarkar, "Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms," *Neural Computing and Applications*, vol. 35, no. 7, pp. 5479-5499, 2023.

43. M. C. Araújo, R. C. Lima, and R. M. De Souza, "Interval symbolic feature extraction for thermography breast cancer detection," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6728-6737, 2014.

44. A. Bhanushali, K. Sivagnanam, K. Singh, B. K. Mittapally, L. T. Reddi, and P. Bhanushali, "Analysis of Breast Cancer Prediction Using Multiple Machine Learning Methodologies," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 3, pp. 1077-1084, 2023.

45. R. R. Kadhim and M. Y. Kamil, "Comparison of machine learning models for breast cancer diagnosis," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 1, p. 415, 2023.

46. L. Yang and Z. Xu, "Feature extraction by PCA and diagnosis of breast tumors using SVM with DE-based parameter tuning," *International Journal of Machine Learning and Cybernetics*, vol. 10, pp. 591-601, 2019.

47. K. Jabeen *et al.*, "BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection," *Diagnostics*, vol. 13, no. 7, p. 1238, 2023.

48. M. L. Bermingham *et al.*, "Application of high-dimensional feature selection: evaluation for genomic prediction in man," *Scientific reports*, vol. 5, no. 1, p. 10312, 2015.

49. A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm," *Computers & structures*, vol. 169, pp. 1-12, 2016.

50. A. U. Haq *et al.*, "Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques," *IEEE Access*, vol. 9, pp. 22090-22105, 2021.

51. D. Houfani *et al.*, "An improved model for breast cancer diagnosis by combining pca and logistic regression techniques," 2023.

52. S. F. Khorshid and A. M. Abdulazeez, "Breast cancer diagnosis based on k-nearest neighbors: a review," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 4, pp. 1927-1951, 2021.

53. Y. Khourdifi and M. Bahaj, "Applying best machine learning algorithms for breast cancer prediction and classification," in *2018 International conference on electronics, control, optimization and computer science (ICECOCS)*, 2018: IEEE, pp. 1-5.

54. V. Kushwaha, "Breast cancer diagnostic using machine learning: applying supervised learning techniques to Coimbra and Wisconsin datasets," 2023.

55. A. Al Bataineh, "A comparative analysis of nonlinear machine learning algorithms for breast cancer detection," *International Journal of Machine Learning and Computing,* vol. 9, no. 3, pp. 248-254, 2019.

56. M. A. Naji, S. El Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Procedia Computer Science,* vol. 191, pp. 487-492, 2021.

57. F. J. M. Shamrat, M. A. Raihan, A. S. Rahman, I. Mahmud, and R. Akter, "An analysis on breast disease prediction using machine learning approaches," *International Journal of Scientific & Technology Research,* vol. 9, no. 02, pp. 2450-2455, 2020.

58. O. J. Egwom, M. Hassan, J. J. Tanimu, M. Hamada, and O. M. Ogar, "An LDA–SVM machine learning model for breast cancer classification," *BioMedInformatics,* vol. 2, no. 3, pp. 345-358, 2022.

59. A. Dwivedi, B. B. Dumka, N. Akhtar, M. F. Shan, and Y. Perwej, "Tropical Convolutional Neural Networks (TCNNs) Based Methods for Breast Cancer Diagnosis," *International Journal of Scientific Research in Science and Technology (IJSRST), Print ISSN,* pp. 2395-6011, 2023.

60. D. Muduli, R. Dash, and B. Majhi, "Fast discrete curvelet transform and modified PSO based improved evolutionary extreme learning machine for breast cancer detection," *Biomedical Signal Processing and Control,* vol. 70, p. 102919, 2021.

61. A. Sahu, P. K. Das, and S. Meher, "High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets," *Biomedical Signal Processing and Control,* vol. 80, p. 104292, 2023.

62. S. S. Chakravarthy and H. Rajaguru, "Automatic detection and classification of mammograms using improved extreme learning machine with deep learning," *Irbm,* vol. 43, no. 1, pp. 49-61, 2022.

63. A. A. A. Arafa, N. El-Sokary, A. Asad, and H. Hefny, "Computer-aided detection system for breast cancer based on GMM and SVM," *Arab Journal of Nuclear Sciences and Applications,* vol. 52, no. 2, pp. 142-150, 2019.

64. T. A. Assegie, "An optimized K-Nearest Neighbor based breast cancer detection," *Journal of Robotics and Control (JRC),* vol. 2, no. 3, pp. 115-118, 2021.

65. V. D. P. Jasti *et al.*, "Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis," *Security and communication networks,* vol. 2022, pp. 1-7, 2022.

66. X. Cai, X. Li, N. Razmjooy, and N. Ghadimi, "Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm," *Computational and Mathematical Methods in Medicine,* vol. 2021, 2021.

67. Y. Wang, E. J. Choi, Y. Choi, H. Zhang, G. Y. Jin, and S.-B. Ko, "Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning," *Ultrasound in medicine & biology,* vol. 46, no. 5, pp. 1119-1132, 2020.

68. K. Jabeen *et al.*, "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion," *Sensors,* vol. 22, no. 3, p. 807, 2022.

69. N. Arya and S. Saha, "Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model," *IEEE/ACM transactions on computational biology and bioinformatics,* vol. 19, no. 2, pp. 1032-1041, 2020.

70. X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Machine Learning with Applications,* vol. 6, p. 100154, 2021.

71. A. Sharma and P. K. Mishra, "Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis," *International Journal of Information Technology,* pp. 1-12, 2022.

72. Y. S. Solanki *et al.*, "A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches," *Electronics,* vol. 10, no. 6, p. 699, 2021.

73. [68] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, no. 2, pp. 687-699, 2018.

74. M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert systems with applications*, vol. 36, no. 2, pp. 3240-3247, 2009.

75. A. Rafid, S. Azam, S. Montaha, A. Karim, K. U. Fahim, and M. Z. Hasan, "An effective ensemble machine learning approach to classify breast cancer based on feature selection and lesion segmentation using preprocessed mammograms," *Biology*, vol. 11, no. 11, p. 1654, 2022.

76. S. A. El_Rahman, "Predicting breast cancer survivability based on machine learning and features selection algorithms: a comparative study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 8585-8623, 2021.

77. M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," *Engineering and Applied Science Research*, vol. 48, no. 1, pp. 65-72, 2021.

78. N. Pérez, A. Silva, and I. Ramos, "Ensemble features selection method as tool for breast cancer classification," *International Journal of Image Mining*, vol. 1, no. 2-3, pp. 224-244, 2015.

79. O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer," *International Journal of Engineering & Technology*, vol. 7, no. 4.36, pp. 160-166, 2018.

80. S. Alghunaim and H. H. Al-Baity, "On the scalability of machine-learning algorithms for breast cancer prediction in big data context," *Ieee Access*, vol. 7, pp. 91535-91546, 2019.

81. A. S. Assiri, S. Nazir, and S. A. Velastin, "Breast tumor classification using an ensemble machine learning method," *Journal of Imaging*, vol. 6, no. 6, p. 39, 2020.

82. U. Naseem *et al.*, "An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers," *IEEE Access*, vol. 10, pp. 78242-78252, 2022.

83. R. Kapila and S. Saleti, "An efficient ensemble-based Machine Learning for breast cancer detection," *Biomedical Signal Processing and Control*, vol. 86, p. 105269, 2023.

84. H. A. Alsayadi, A. A. Abdelhamid, E.-S. M. El-Kenawy, A. Ibrahim, and M. M. Eid, "Ensemble of Machine Learning Fusion Models for Breast Cancer Detection Based on the Regression Model," *Fusion: Practice & Applications*, vol. 9, no. 2, 2022.

85. S. Sharmin, T. Ahammad, M. A. Talukder, and P. Ghose, "A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection," *IEEE Access*, 2023.