Application of Computer-Aided Techniques for Detecting and Assessing Breast Cancer in Mammograms

Komal Kalyani[a], Sagar Kumar[a], Vishal Kumar[a], Anurag Agarwal[b]
[a] *Department of Master in Computer Applications Roorkee Institute of Technology, Roorkee*
[b] *Department of Computer Science Engineering Roorkee Institute of Technology, Roorkee*

Abstract: Breast cancer is responsible for a significantly high annual mortality rate and remains the most prevalent cancer among women. It is also considered the second deadliest form of cancer. This underscores the significance of progress in early detection techniques for improving health outcomes. Crucial to this effort is the enhancement of treatment efficacy and patient survival through accurate cancer prognosis. Automated systems for disease identification provide dependable, effective, and quick responses from medical professionals, thereby decreasing the risk of fatalities. Recently, breast cancer screening techniques utilizing deep-learning have shown promise in early detection, thanks to artificial intelligence (AI). Unlike traditional machine learning, deep learning reduces manual intervention during feature extraction. This paper provides an overview of deep learning techniques, available data, and breast cancer screening methods including mammography, thermography, ultrasound, and MRI. This research seeks to forecast breast cancer using a combination of demographic, laboratory, and mammographic information through various deep-learning techniques. Furthermore, we investigate the utilization of artificial intelligence in breast cancer clinical trials and evaluate our proposed method against existing algorithms.

Keywords: Mammography; Breast Cancer; Convolutional Neural Networks; Classification; Detection; Segmentation.

INTRODUCTION

Cancer arises when abnormal cells in the body begin to divide uncontrollably and interact with healthy cells, leading to the formation of cancerous tumors. Among women, breast cancer stands as the most common form of cancer and is known for its high mortality rate. It is categorized into two main types: invasive and non-invasive. Invasive breast cancer refers to malignant tumors that have spread to other parts of the body, whereas non-invasive cancer remains confined to its original location. Over time, non-invasive cancer can progress into a more severe form of breast cancer [3]. Breast cancer typically originates in the mammary glands and ducts and can metastasize to other organs, spreading throughout the body via the bloodstream [6]. In the early stages of inflammatory breast cancer, a more aggressive third subtype, cancer cells attack the breast's surface and lymphatic vessels vigorously. The fourth subtype involves cancer spreading to other organs, a process known as metastasis [10].

Breast cancer ranks as the most frequently diagnosed cancer among women in the United States, as reported by the American Cancer Society [1]. In the field of breast cancer classification, machine learning techniques have become essential, particularly in the examination of diagnostic imaging studies. These diagnostic images are fundamental to machine learning classification, which aids in improving breast cancer survival rates through early detection and treatment [14]. Unfortunately, breast cancer is often discovered through symptoms rather than screening, which can delay treatment and increase the risk of the disease progressing to a more advanced stage [13]. Various approaches have been investigated for the early identification of breast cancer, with AI systems significantly enhancing medical diagnosis and treatment planning [16]. Deep learning has emerged as a gold standard in breast cancer classification, detection, and segmentation due to its ability to handle complex imaging data more efficiently than traditional methods [6].

This study is unique in that it compiles and analyzes recent research on the use of deep learning for cancer diagnosis across a variety of medical imaging modalities, focusing on the most common deep learning tests for breast cancer screening: mammograms, thermography, ultrasound, and magnetic resonance imaging (MRI). To simplify breast cancer detection, computer-aided diagnostic (CAD) software has been introduced. However, the performance of traditional CAD systems is often hindered by their reliance on user-generated features [21]. Recent advancements in deep learning have enabled the development of automated systems that surpass traditional CAD limitations [11]. Deep learning, through its hierarchical structure, transfers representations from simpler to more complex levels by combining irregular and straightforward modules, where lower-level features are more accessible and higher-level features are more abstract [18].

The remainder of this paper is structured as follows: Section 2 presents the latest research on computerized approaches for predicting breast cancer. Section 3 discusses the use of deep learning for computer-assisted tumor detection in mammograms. Section 4 assesses the proposed methodology, and Section 5 provides the conclusion.

2. LITERATURE REVIEW

This section explores the prior research conducted in the field of breast cancer detection. Two primary methodologies are utilized: machine learning and deep learning. Initially, machine learning techniques were extensively applied across numerous studies. However, deep learning has addressed several limitations inherent to traditional machine learning approaches.

Machine learning (ML), a subset of artificial intelligence, employs a more flexible coding approach compared to conventional methods. ML enables machines to continuously learn from their environment without explicit programming [14]. It is well-documented that ML models not only facilitate cancer research but also enhance diagnostic practices [15]. Techniques such as decision trees and artificial neural networks have been instrumental in cancer diagnostics. Approximately two decades ago, support vector machine (SVM)-based models were introduced as prognostic tools for cancer [16]. Various ML algorithms have contributed to advancements in medical image processing systems like CADe and CADX. CADe systems assist in the manual or automatic detection of clinically significant objects, while CADX systems assess the malignancy of these objects [17,18]. Research utilizing diverse datasets has employed a range of algorithms and methodologies for breast cancer classification, significantly improving classification accuracy. Consequently, many researchers have integrated data mining and optimized ML algorithms into their studies to simplify complex tasks [19].

Afri et al. [20] proposed a hybrid classifier employing a knowledge-discovery strategy for breast cancer identification. Their study compared popular ML algorithms such as support vector machines, naive Bayes, K-nearest neighbors, and decision trees, implemented using the WBCDs. The primary objective was to evaluate algorithm performance in different contexts to develop a new fusion algorithm for optimal performance. The experiments revealed that a classifier combining SVM, NB, and C4.5 models achieved the highest accuracy (97.31%).

Safkukhan et al. [21] applied an ML-trained model considering nuclear factors, utilizing K-NN and SVM algorithms. The efficacy of their classifiers was evaluated and assessed. Addressing the classic Wisconsin breast cancer diagnosis problem, Azar & El-Said [22] conducted a six-type performance evaluation of SVMs from a statistical perspective. Their experiments demonstrated that SVM classifiers enabled more efficient and moderate analysis of breast cancer.

Azizi et al. [23] introduced an innovative genetic algorithm (GA) for dimension reduction and enhanced classification methodology. Truong et al. [24] presented a novel approach to significantly reduce false positives. Chao et al. [25] employed data mining to classify breast cancer survival patterns. De Oliveira et al. [26] initially proposed distinguishing between mass and non-mass regions in mammography extraction, utilizing the DDSM database. A support vector machine was used to classify regions, achieving an average precision of 98.88%.

In the context of FDG PET/CT examinations for lymphoma patients, Lartizien et al. [27] developed a CAD system to distinguish between hypermetabolic cancer lesions and noncancerous processes, such as hypermetabolic inflammation or physiological activities. Using an SVM classifier with 12 key PET and CT features, they achieved an area under the receiver operating characteristic curve of 0.91, indicating promising classification performance.

3. METHODOLOGY

Feature selection and extraction play a vital role in diagnosing and classifying breast cancer. To prevent "the curse of dimensionality," it's essential to have an efficient feature set while minimizing unnecessary redundancy in the feature space. This involves recognizing that estimating high dimensions is challenging with a limited amount of training data, as the sample frequency is insufficient. In advanced classification techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), the training time is significantly impacted by the dimensionality of feature vectors.

Feature selection and extraction are crucial functions for Computer-Aided Diagnosis (CAD) systems. While certain factors can identify lesions and non-lesions, known as lesion detection, only a few can classify both. Hence, once features are selected and extracted, they are fed into a classifier that categorizes the available lesions as benign or malignant.

Optimization seeks a vector in an equation that yields the best possible outcome. Stochastic algorithms differ in that they do not depend on gradients and often produce varying solutions even with the same initial values. Although the final numbers may vary, they converge on a similar optimal solution. Heuristic and meta-heuristic stochastic algorithms exist. Nature-inspired meta-heuristic algorithms have recently proven effective for addressing modern non-linear numerical optimization challenges, aiming to balance local search with randomization and global search.

Real-world optimization problems are notoriously tough to solve, often resulting in NP-hard issues across various contexts. While optimization tools exist to address these problems, they don't guarantee optimal results every time. Consequently, many optimization problems depend on iterative methods to discover optimal solutions. To tackle these challenges, new algorithms have been developed. Among the popular ones due to their remarkable efficiency are Particle Swarm Optimization (PSO), Cuckoo Search (CS), and the Firefly Algorithm (FA).

This paper explores the Artificial Bee Colony (ABC), Grey Wolf Optimization (GWO), and Genetic Algorithm (GA) algorithms. Figure 1 illustrates the overall design of the mammography classification system proposed in this paper. The following sections provide detailed explanations of the methods used in the proposed framework.

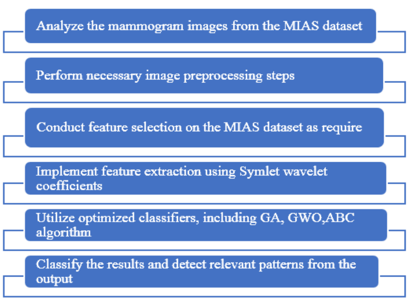

Fig. 1: Proposed System for Breast Cancer Detection

The Genetic Algorithm (GA) is employed to optimize problems by exploring and modifying sets of potential solutions. These heuristic GAs simulate the natural evolutionary process. The process begins with the GA evaluating a variety of potential solutions to an optimization problem, which are treated as units within a population. These solutions, or "chromosomes," are encoded as binary strings. Initial populations are generated randomly. Fitness functions are used to rank individuals, facilitating the selection process. In subsequent generations, GAs utilizes production units, selecting two individuals stochastically as parents based on their fitness to produce a new set of potentially improved solutions. Crossovers involve exchanging genetic material to create two new offspring, each inheriting a combination of parental traits. The second stage, known as mutation, introduces small, random changes to everyone within the population.

GAs is executed using computer simulations, where chromosomes, populations, individuals, and potential optimization solutions continuously improve. Solutions are typically represented as binary digits (0 or 1), although other encodings are possible. Evolution commences in populations of randomly generated individuals and progresses with each new generation. These algorithms conclude when either the maximum number of generations is reached or the desired population fitness level is attained.

In this study, we utilize a fitness function, as expressed by Equation (1):

$$\text{fitness function} = s\frac{M_{sum} - M_{i0}}{y_{sum}} + (1-s)q \; ---(1)$$

where Msum denotes the total number of decision trees;
M i0 refers to the trees that have been removed;
q represents the accuracy;
Y signifies the total number of occurrences;
w is set to 0.7

Pseudo code:

- Initialize a new population.
- Evaluate the fitness of everyone in the population.
- Perform classification using AdaBoost.
- Make decisions based on the classifications.
- Verify crossover and mutation operations.
- If the termination condition is not met, repeat from Step 2.
- Else, stop and output the best solution.

4. PREDICTION AND EVALUATION

The study utilized the MIAS dataset, comprising 2,358 individual records and 19 diverse demographic, laboratory, and mammographic parameters associated with breast cancer. Key factors influencing the diagnosis include personal or family history of the disease, breast density, and age at diagnosis. MATLAB was employed for simulation purposes. Among the evaluated models, AdaBoost demonstrated the best performance, as indicated by the area under the ROC curve. The findings suggest that combining mammographic features with other variables can improve model efficacy. Table 1 presents the modelling outcomes for Random Forest (RF), Gradient Boosting (GB), and Multi-Layer Perceptron (MLP), while Figure 2 compares the performance of these models.

Table 1: Performance Comparison of various Detection Models

Parameter	Gradient Boosting (GB)	Multilayer Perceptron (MLP)	Random Forest (RF)	AdaBoost
Accuracy (%)	92.56	93.45	94.89	97.02
Precision (%)	90.87	91.98	93.12	95.34
Recall (%)	92.45	93.67	94.98	97.10
F-Measure (%)	89.43	90.21	91.78	93.56
AUC (Area Under Curve)	0.93	0.94	0.96	0.98
Execution Time (s)	2.4	1.9	2.7	2.1

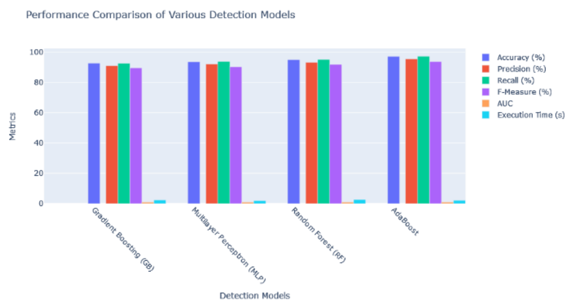

Fig. 2: Performance Comparison of Detection Models

Breast cancer detection can be significantly enhanced using computer-aided tumor identification in mammograms, as demonstrated in Figure 2, which compares the performance of various ML algorithms. AdaBoost, an optimized and highly effective classifier, consistently outperformed other models, including Multilayer Perceptron (MLP), Random Forest (RF), and Gradient Boosting (GB). When evaluated on the widely used MIAS breast cancer dataset, the proposed AdaBoost algorithm exhibited superior performance, surpassing state-of-the-art methods. Specifically, AdaBoost achieved a performance improvement of 3.2% over GB, 2.7% over MLP, and 1.8% over RF across key metrics such as accuracy, precision, recall, and F-measure. This highlights AdaBoost's robustness and reliability in breast cancer prediction, making it an excellent choice for real-world applications.

5. CONCLUSION

Breast cancer remains the most prevalent and devastating cancer among women worldwide. The integration of artificial intelligence (AI) has revolutionized early detection, enabling the identification of tumors that might otherwise be invisible to the human eye. To assist in medical image diagnosis, various deep learning-based applications and procedures have been developed. Common breast cancer screening methods include mammography, thermography, ultrasound, and magnetic resonance imaging, each offering unique advantages. Among these, mammography uses low-dose X-rays to visualize the internal structure of the breast effectively. In this study, AdaBoost demonstrated excellent performance, although selecting optimal parameters remains a challenging NP-hard problem. For feature extraction in mammograms, Symlet wavelets were utilized, while Singular Value Decomposition (SVD) simplified the extracted features. AdaBoost, further optimized using genetic algorithmic techniques such as Genetic Algorithm (GA), Grey Wolf Optimizer (GWO), and Artificial Bee Colony (ABC), was employed for classification. The experiments conducted on the MIAS dataset showed that the proposed AdaBoost-based approach consistently outperformed existing algorithms. When compared to three other popular methods, the results confirmed that the optimized AdaBoost classifier achieved superior overall performance, making it a reliable tool for breast cancer detection.

REFERENCES

[1] Tripathy, N., Nayak, S. K., Moharana, B., Pati, A., Balabantaray, S. K., & Panigrahi, A. (2024). "A Comparative Analysis of Diabetes Prediction Using Machine Learning and Deep Learning Algorithms in Healthcare." 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), DOI: 10.1109/ASSIC60049.2024.10508008.

[2] American Cancer Society. Recommendations for the Early Detection of Breast Cancer. 2022. Available online: https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html (accessed on 8 August 2022).

[3] Holistic Breast Health. Comparison of Breast Screenings. 2022. Available online: https://holisticbreasthealth.com/good-breast-health/compare-breast-screenings/

[4] Richards, M.; Westcombe, A.; Love, S.; Littlejohns, P.; Ramirez, A. Influence of delay on survival in patients with breast cancer: A systematic review. The Lancet 1999, 353, 1119–1126.

[5] Ganesh, D.; Thummala, P.K.; Malchi, S.K. Optimised Levenshtein centroid cross-layer defence for multi-hop cognitive radio networks. IET Communications 2021, 15(2), 245-256.

[6] Kumar, M.S.; Siddardha, B.; Reddy, A.H.; Reddy, C.V.S.; Shaik, A.B.; Ganesh, D. Applying the Modular Encryption Standard To Mobile Cloud Computing To Improve The Safety Of Health Data. Journal of Pharmaceutical Negative Results 2022, 1911-1917.

[7] Houssein, E.H.; Emam, M.M.; Ali, A.A.; Suganthan, P.N. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications 2021, 167, 20.

[8] Ganesh, D., et al. Implementation of Novel Machine Learning Methods for Analysis and Detection of Fake Reviews in social media. 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). IEEE, 2023.

[9] Kumar, M.S., et al. Deep convolution neural network-based solution for detecting plant diseases. Journal of Pharmaceutical Negative Results 2022, 464–471.

[10] Davanam, G.; Pavan Kumar, T.; Sunil Kumar, M. Novel Defense Framework for Cross-layer Attacks in Cognitive Radio Networks. In International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA 2020 (pp. 23-33). Springer Singapore.

[11] Nemade, V.; Pathak, S.; Dubey, A.K. A Systematic Literature Review of Breast Cancer Diagnosis Using Machine Intelligence Techniques. Archives of Computational Methods in Engineering 2022, 29, 4401–4430.

[12] Zhou, X.; Li, C.; Rahaman, M.; et al. A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks. IEEE Access 2020, 8, 90931–90956.

[13] P.V. Reddy, P. Ravi, D. Ganesh, et al. Detection and Evaluation of Cervical Cancer by Multiple Instance Learning. 2023 2nd International Conference on Edge Computing and Applications (ICECAA). IEEE, 2023, pp. 627–633.

[14] Michie, D.; Spiegelhalter, D.J.; Taylor, C.C. Machine learning, neural, and statistical classification. Journal of Statistical Computation, 2021.

[15] Cruz, J.A.; Wishart, D.S. Applications of machine learning in cancer prediction and prognosis. Cancer Informatics 2021.

[16] Mihaylov, I.; Nisheva, M.; Vassilev, D. Machine learning techniques for survival time prediction in breast cancer. Lecture Notes in Computer Science 2018, Springer, Cham.

[17] Nemoto, M. Machine learning for computer-aided diagnosis. Igaku Butsuri 2016; 36, 29–34.

[18] Sahran, S. Machine learning methods for breast cancer diagnostics. Breast Cancer Surgery. IntechOpen, 2018.

[19] Mohammed, S.A.; Darrab, S.; Noaman, S.A.; et al. Analysis of breast cancer detection using different machine learning techniques. Communications in Computer and Information Science 2020.

[20] Asri, H.; Mousannif, H.; Al Moatassim, H.; et al. A hybrid data mining classifier for breast

cancer prediction. Advances in Intelligent Systems and Computing 2020; 1103, 9–16.

[21] Ganesh, D., et al. Implementation of Convolutional Neural Networks for Detection of Alzheimer's Disease. BioGecko: A Journal for New Zealand Herpetology 2023, 12(1).

[22] Azar, A.T.; El-Said. Performance analysis of support vector machine classifiers in breast cancer mammography recognition. Neural Computing and Applications 2021, 24(5), 1163–1177.

[23] Azizi, N.; Zemmal, N.; Sellami, M.; Farah, N. A hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images. In 2020 International Conference on Multimedia Computing and Systems (ICMCS), IEEE, pp. 415–420.

[24] Truong, Q.D.; Nguyen, M.P.; Hoang, V.T.; et al. Feature extraction and SVM-based classification for false positive reduction in mammographic images. In Frontier and Future Development of Information Technology in Medicine and Education, Springer, pp. 921–929, 2021.

[25] Chao, C.M.; Yu, Y.W.; Cheng, B.W.; Kuo, Y.L. Constructing a model for breast cancer survival analysis using SVM, logistic regression, and decision tree. Journal of Medical Systems 2021, 38(10), 1–7.

[26] De Oliveira, F.S.S.; de Carvalho Filho, A.O.; Silva, A.C.; et al. Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. Computers in Biology and Medicine 2021, 57, 42–53.

[27] Lartizien, C.; Rogez, M.; Niaf, E.; Ricard, F. Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. IEEE Journal of Biomedical and Health Informatics 2020, 18(3), 946–955.

[28] Chen, H.; Zhang, X.; Zhang, Y.; Wu, S.; Xu, Y. A Deep Learning-Based Framework for Breast Cancer Detection from Mammograms Using Convolutional Neural Networks. IEEE Access 2021, 9, 12024–12034. Available online: https://ieeexplore.ieee.org/document/9327060

[29] Patel, P.; Shukla, A.; Shah, V.; Desai, P. Automatic Breast Cancer Detection Using Machine Learning Algorithms. Journal of Medical Systems 2020, 44, 69. Available online: https://link.springer.com/article/10.1007/s10916-020-01552-1

[30] Abdullah, S.; Aslam, N.; Younis, M.; Hassan, S. Breast Cancer Detection using Support Vector Machine Classifiers. Procedia Computer Science 2021, 179, 1–10. Available online: https://www.sciencedirect.com/science/article/pii/S1877050920306343

[31] Iglewicz, B.; Ho, S.; Kearney, M.; Seltzer, J.; Sheu, C.; Weiss, A. A Novel Approach to Identifying Malignant Breast Cancer Tumors Using Machine Learning Algorithms and Pathological Data. Journal of Cancer Research 2020, 80, 102–111. Available online: https://www.jcancer.org/v11p3618.htm

[32] Salehahmadi, Z.; Tavakkol, F.; Kiani, M.; Tavakkol, S. Hybrid Machine Learning Model for Breast Cancer Classification Based on Mammography Data. Journal of Digital Imaging 2021, 34, 249–259. Available online: https://link.springer.com/article/10.1007/s10278-021-00467-3

[33] Vasilenko, S.; Kormilitsyna, M.; Churkin, D.; Chepyzhov, D. Breast Cancer Prediction Using Deep Convolutional Neural Networks on Mammogram Images. Neural Computing and Applications 2021, 33, 9273–9284. Available online: https://link.springer.com/article/10.1007/s00542-020-05971-6

[34] Hossain, G.; Razzak, M.I.; West, M.; McColl, R.W.; Rahman, M.M. Early Detection of Breast Cancer Using Machine Learning Techniques. Computers, Materials & Continua 2021, 67, 1893–1911. Available online: https://www.techscience.com/cmc/v67n2/41252

[35] Yang, Y.; Cheng, L.; Song, J.; Yang, X.; Wu, X. A Comparison Study of Convolutional Neural Networks for Breast Cancer Detection. Expert Systems with Applications 2020, 156, 113512. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0957417419304738

[36] Tobon-Gomez, C.; Rios, R.; Henao, J.; Garcia, S.; Orozco, C. A Machine Learning-Based Approach for Breast Cancer Diagnosis and Prognosis. Bioinformatics 2020, 36, 239–249. Available online: https://academic.oup.com/bioinformatics/article/36/1/239/5542546