JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 784 - 789
https://publishoa.com
ISSN: 1309-3452

Review On Diagnosis for Early Stage of Breast Cancer

Mrs S. Maria Sylviaa[1], Dr. N. Sudha [2]

[1]Assistant Professor, Department of Computer Science, Nirmala College for Women
[2]Associate Professor, Department of Computer Science, Bishop Appasa my College of Arts and Science

ABSTRACT
Breast cancer is one of the most common types of cancer in women. The likelihood of rescuing a breast cancer patient is highly reliant on early identification and treatment commencement. The high death rate due to breast cancer is attributable to a lack of knowledge of the necessity of early detection of symptoms of various breast cancers and a lack of training in breast cancer symptom identification. A computer-aided diagnostic (CAD) expert system facilitates a pathologist for early breast cancer diagnosis and determines if the tumour is harmless or aggressive. The goal is to use breast records to examine the influence of CAD systems. This study included the evaluation of present state-of-the-art methodologies implemented for every phase, including traditional procedures, comparisons within approaches, and technical specifications with advantages and limitations. Finally, the research gaps in existing machine learning methodologies for implementation and recommendations for future researchers are described.

Keywords: Breast cancer, Computer-aided diagnosis, early-stage, symptom identification, peak mortality, Conventional Methods and Benign.

I. INTRODUCTION

There are trillions of cells in the human body. Cancer is described as when a cell divides improperly or uncontrollably in various places of the body. Its category is determined by the location of cancer on the body. This disease may cause death if it is left undetected. According to the World Health Organization (WHO), breast cancer is the most common cancer in women, impacting 2.1 million people each year. Also, it claimed the lives of over 627,000 women in 2018, accounting for almost 15% of all cancer deaths among women. Every year in the United Kingdom, roughly 55,200 new instances of breast cases were diagnosed, or about 150 every day between 2015 and 2017[1,2].

Early breast cancer identification considerably improves the chances of a successful treatment plan and ensures the patient's long-term survival. Nearly 98% of patients will live for five years or more if the disease is found and diagnosed early, compared to roughly 1 in 4 (26%) people who would live for five years or more if the condition is detected and diagnosed later. A 'two-week wait' is the most common technique for diagnosing breast cancer. Pathologists' criteria for diagnosing breast cancer involve a thorough microscopic examination typically. As a result, the CAD system facilitates the diagnostic process and lowers diagnostic uncertainty [3,4,5].

Many machine learning approaches have been applied to CAD systems. These approaches can exceed humans and adapt more efficiently over time. Therefore, incorporating machine learning into diagnosis can provide vital information to help pathologists evaluate and analyze massive amounts of medical data. In addition, it can quicken up the procedure because it can handle enormous amounts of data far faster than a pathologist. For example, a breast cancer diagnosis can be conceived as a classification task in machine learning, with the output being either malignant or benign stages of breast cancer [6,7].

Traditionally, categorization issues were solved using a variety of standard techniques. Nevertheless, these techniques may not account for concerns such as the unbalanced costs of misclassification within different groups, resulting in unintended outcomes. The goal is to use breast records to examine the influence of CAD systems. This study included evaluating traditional procedures, comparisons within approaches, and technical specifications with advantages and limitations. Finally, the research gaps in existing machine learning methodologies for implementation and recommendations for future researchers are described [8].

Section I examines the importance of early-stage breast cancer diagnosis, Section II includes an outline of the conventional techniques used for early-stage breast cancer diagnosis, Section III compares research strategies and their benefits and drawbacks, Section IV discusses inferences from existing work. Finally, section V discusses the conclusion and the future work.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 784 - 789
https://publishoa.com
ISSN: 1309-3452

II. LITERATURE REVIEW

Kumar et al [2017][9] established a concept for early-stage breast cancer diagnosis. When new cells proliferate uncontrollably, they form a mass of tissue known as a tumor classified as benign or malignant. Therefore, using classification algorithms, early detection necessitates an accurate diagnostic technique to determine whether the tumor is benign or malignant. The primary goal of this study is to compare the performance of supervised learning classification algorithms and their combinations using the voting classifier strategy. Voting is an ensemble strategy that combines different models for better categorization. The dataset was obtained from a database at Wisconsin University.

Amrane et al [2018][10] provided Naive Bayes (NB) classifier and the K Nearest Neighbour (KNN) for breast benign cancer/malign cancer classification. For evaluating the accuracy of these classifiers, cross validation is performed. The findings indicate that the KNN classifier outperforms the NB classifier in terms of accuracy (97.51 %) and error rate (96.19 %).

Kumar et al. [2020][11] concentrated on deploying several data mining classification approaches to identify normal and cancerous breasts. The Breast Cancer Wisconsin data collection of the UCI repository is an experimental dataset, with attribute clump thickness serving as an evaluation class. The performance of the following twelve algorithms was investigated on this data set: Ada Boost M1, Decision Table, J-Rip, J48, Lazy IBK, Lazy K-star, Logistics Regression, Multiclass Classifier, Multilayer–Perceptron, Nave Bayes, Random Forest, and Random Tree.

Rashmi, et al [2015][12] suggested Naïve Bayes classification algorithm and Naïve Bayes prediction algorithm. These algorithms are applied to determine if the tumor is normal or cancerous. The data from a database at Wisconsin University are utilized in this approach to calculate the success and failure rates.

Basunia et al. [2020][13] designed an ensemble procedure called a stacking classifier, which effectively diagnoses malignant or benign tumors by integrating different classification algorithms. The experiment uses the "Wisconsin Diagnosis Breast Cancer" dataset from the UC Irvine Machine Learning Repository. In general, the proposed Stacking classifier used a meta classifier to integrate the results of the top classifiers and delivered 97.20 % for breast cancer detection. In addition, the performance of various data mining technologies was thoroughly analyzed by different evaluation measures.

Laghmati et al. [2020][14]presented a comparison of various ML techniques, and four ML algorithms (KNN, decision tree, Binary SVM, and Adaboost) were employed to identify whether a patient has a normal and cancerous tumor. To minimize the number of features and the complexity, the dataset's features are fed into a feature selection model using Neighbourhood Components Analysis (NCA). As a result, the KNN model had the highest predictive accuracy of 99.12 %, the Binary SVM method had the best predictive specificity of 95.56 %, and the kNN and Adaboost approaches had the highest predictive sensitivity.

Gayathri and Sumathi [2015][15] applied fuzzy logic to determine the risk of developing breast cancer. The dataset for this study was obtained from the UCI machine learning library. This suggested research aims to shorten the diagnosing time of breast cancer by minimizing the contributing factors. The features were extracted using a Linear Discriminant Analysis (LDA), and the training was completed using the Mamdani Fuzzy inference framework. This approach has a 93% outcome.

Elouedi et al. [2014][16] proposed a hybrid breast cancer diagnosis technique based on decision trees and clustering. The suggested approach implies the distinction between normal and cancerous cases and refines the latter's therapy. An experimental study on the Wisconsin Breast Cancer Database gives a detailed analysis of the induced outcomes, which depicts that classification accuracy is improved by employing a clustering technique to classify distinct forms of Breast Cancer.

Tafish and El-Halees [2018][17] suggested a methodology to determine disease risk and get guidelines, reduce time and cost, and advance well-being according to data from Gaza Strip hospitals. To predict the severity of breast cancer, the model utilizes classification techniques such as SVM, ANN, and k-nearest neighbors. In addition, the association rules are applied to identify high severity breast cancer-related traits. Using the above classification approaches on the breast cancer dataset, we obtained a prediction accuracy of 77%, which is acceptable. Also, we might identify the most linked traits to severe breast cancer.

Ara et al. [2021] [18] utilized the Wisconsin Breast Cancer Dataset from the UCI library to assess the dataset and analyze the efficacy of several machine learning algorithms for detecting breast cancer. SVM, Logistic Regression, K-Nearest Neighbors, Decision Tree, Naive Bayes, and Random Forest classifiers were used to differentiate benign and malignant

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 784 - 789
https://publishoa.com
ISSN: 1309-3452

tumors. According to the findings, Random Forest and SVM have high accuracy than other classifiers with 96.5 %. These classifiers create an automatic detection approach for the early detection of breast cancer.

Alyami et al. [2017][19] used SVM and ANN to diagnose breast cancer in conjunction with feature selection. The correlation coefficient of each feature against the target class is utilized to choose features, and various feature subsets are being used. The algorithm is evaluated on the prevalent Wisconsin Diagnosis Breast Cancer (WDBC) dataset. 10- fold Cross-validation was performed for data partitioning, and the results show improved classification accuracy. In terms of comparison, empirical research revealed that SVM outperformed ANN, with a classification accuracy of 97.14% and 96.71%, respectively.

Bhardwaj et al. [2014][20] introduced a novel Genetically Optimized Neural Network (GONN) approach as a solution for classifying issues. To mitigate the harmful effect of these procedures, introduce unique crossover and mutation operations that differ from the conventional Genetic programming life-cycle. For example, use the GONN method to determine if a breast cancer tumor is malignant or benign. Our technique outperforms a Back Propagation neural network and a Support Vector Machine in classification accuracy by nearly 4% and 2%, respectively.

Addeh et al. [2017][21] introduced a technique for early breast cancer diagnosis relying on an adaptive neuro-fuzzy inference system (ANFIS) and feature extraction. ANFIS is an intelligent classifier, and the association rules (AR) mechanism is a feature selection algorithm. The cuckoo optimization algorithm (COA) is used to determine the best radius value. The method is tested on the Wisconsin Breast Cancer Database (WBCD), and the findings show that it has a high detection accuracy.

Yassi et al. [2014][22] developed a Fuzzy system (FS) to recognize the breast cancer category. To optimize an FS, a chaotic hierarchical cluster-based multispecies particle swarm optimization (CHCMSPSO) is applied. This study aims to learn high-accuracy Takagi-Sugeno-Kang (TSK) type fuzzy rules and improve the global search capability of CHCMSPO. Here 11 chaotic maps are employed so that the accuracy rate for differentiating benign and malignant cancer is greater than 90%. Furthermore, the Sinusoidal chaotic map attains a 99% accuracy rate because it aligns with the problem circumstances. This simulation is run on the UCI-Breast Cancer database.

Kamel et al. [2019][23] classified cancer using a Gaussian Naive Bayes method. The technique is evaluated using the Wisconsin Breast Cancer dataset (WBCD) and the lung cancer dataset. The suggested algorithm's analysis results achieved 98% accuracy in identifying breast cancer and 90% accuracy in detecting lung cancer.

Table:1. Comparison of existing methods for diagnosis for the early stage of breast cancer

Author name	Methods	Merits	Demerits
Bhardwaj et al. [2014]	Genetically Optimized Neural Network	High in categorization accuracy	The duration of the network is unknown
Elouedi, et al [2014] .	Hybrid diagnosis approach based on decision trees and clustering	Achieves higher accuracy	Increases the computational complexities
Gayathri and Sumathi [2015]	Mamdani Fuzzy inference model	Reduces the computation time	It does not test with the real-time dataset
Rashmi et al. [2015]	Naïve Bayes classification algorithm	Produces better results	Increases the error rate
Kumar et al. [2017]	Voting ensemble classifier	Produces best prediction results	The proposed model has the overfitting problem
Alyami et al. [2017]	SVM and ANN	Obtains 97.14% accuracy	It does not perform well when the dataset is large
Addeh et al. [2017]	An adaptive neuro-fuzzy inference system	Has high detection accuracy	Computational cost is high
Amrane et al. [2018]	Naive Bayes classifier and K Nearest Neighbour	Highest accuracy with the lowest error rate	The running time will increase for a larger dataset.

Tafish and El-Halees [2018]	Artificial neural networks, Support vector machines, and k-nearest neighbors	Flexible and robust	Time-consuming nature
Basunia et al. [2020]	Ensemble method named stacking classifier	Provides 97.20% accuracy	It does not test with a large volume dataset
Laghmati et al. [2020]	KNN, decision tree, Binary SVM, and Adaboost.	More reliable	Need to improve the accuracy
Kamel et al. [2019]	Gaussian Naive Bayes algorithm	High in classification performance.	All the features are independent
M Thilagaraj et al.[2020]	Deep CNN algorithm	Produces High accuracy in classification	Properties are modified
Lamiss Mohamed Abd El Aziz Sad et al[2020]	Dynamic contrast-enhanced (DCE) and diffusion-weighted imaging (DWI)	Obtains and analyzes both morphological and functional features	MRI exhibited a high negative predictive value with no false-negative cases.

III. INFERENCES FROM THE EXISTING WORK

Existing works did not detect outliers. Data mining uses outlier detection to increase diagnosis accuracy by pre-processing the dataset. Large databases include imperfect data objects. Outliers are data objects that deviate from the norm and are excluded from typical data mining processes. Thus, outlier identification is required to identify outliers, improve data object quality, and produce accurate data mining results. Traditionally, manual analysis and interpretation are used to turn data into knowledge. The issue with manual data analysis is that it is sluggish, expensive, time-consuming, and very subjective.

IV. SOLUTION

Breast cancer is the second largest cause of death in women globally. It begins as a single tissue. Usually, the inner lining of the lobules, the primary provider of milk in the ducts, and the beginning looks like lymph in the breast. Future research in this field will develop an automatic prediction technique for the early detection of cancer. The outlier identification (OD) algorithm was also utilized to find outliers in the cancer dataset. Following the detection of outliers, the pre-processed dataset is classified as harmless or aggressive cancer using an optimized classifier.

V. CONCLUSION AND FUTURE WORK

This work describes the utilization of breast cancer prediction approaches to detect the disease at an early stage. Each of the procedures presented has its strengths and shortcomings, and there is no yet other technique for diagnosing early-stage breast cancer. Moreover, the open research problems in the diagnosis of early-stage breast cancer are investigated by analysing the benefits and drawbacks of current approaches. As a result, future research in this field will focus on the idea of constructing an outlier detection (OD) algorithm to find outliers in the cancer dataset. Using an upgraded classifier, build an automatic model to predict whether the patient has harmless or aggressive cancer and different stages of cancer can be detected.

VI. REFERENCES

1. Ojha, U. and Goel, S., 2017, January. A study on prediction of breast cancer recurrence using data mining techniques. In 2017 *7th International Conference on Cloud Computing, Data Science & Engineering-Confluence (pp. 527-530). IEEE.*

2. Khuriwal, N. and Mishra, N., 2018, March. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In 2018 *IEEMA Engineer Infinite Conference (eTechNxT) (pp. 1-5). IEEE.*

3. Deepika, M. and Kalaiselvi, K., 2018, April. An empirical study on disease diagnosis using data mining techniques. In 2018 *Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 615-620). IEEE.*

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 784 - 789
https://publishoa.com
ISSN: 1309-3452

4. Jacob, D.S., Viswan, R., Manju, V., PadmaSuresh, L. and Raj, S., 2018, March. A survey on breast cancer prediction using data miningtechniques. *In 2018 Conference on Emerging Devices and Smart Systems (ICEDSS) (pp. 256-258). IEEE.*

5. Pritom, A.I., Munshi, M.A.R., Sabab, S.A. and Shihab, S., 2016, December. Predicting breast cancer recurrence using effective classification and feature selection technique. *In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 310-314). IEEE.*

6. Devi, R.D.H. and Deepika, P., 2015, December. Performance comparison of various clustering techniques for diagnosis of breast cancer. *In 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-5). IEEE.*

7. Bazazeh, D. and Shubair, R., 2016, December. Comparative study of machine learning algorithms for breast cancer detection and diagnosis. *In 2016 5th international conference on electronic devices, systems and applications (ICEDSA) (pp. 1-4). IEEE.*

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

9. Kumar, U.K., Nikhil, M.S. and Sumangali, K., 2017, August. Prediction of breast cancer using voting classifier technique. *In 2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM) (pp. 108-114). IEEE.*

10. Amrane, M., Oukid, S., Gagaoua, I. and Ensari, T., 2018, April. Breast cancer classification using machine learning. *In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) (pp. 1-4). IEEE.*

11. Kumar, V., Mishra, B.K., Mazzara, M., Thanh, D.N. and Verma, A., 2020. Prediction of malignant and benign breast cancer: A data mining approach in healthcare applications. *In Advances in data science and management (pp. 435-442). Springer, Singapore.*

12. Rashmi, G.D., Lekha, A. and Bawane, N., 2015, December. Analysis of efficiency of classification and prediction algorithms (Naïve Bayes) for Breast Cancer dataset. *In 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 108-113). IEEE.*

13. Basunia, M.R., Pervin, I.A., Al Mahmud, M., Saha, S. and Arifuzzaman, M., 2020, June. On Predicting and Analyzing Breast Cancer using Data Mining Approach. *In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1257-1260). IEEE.*

14. Laghmati, S., Cherradi, B., Tmiri, A., Daanouni, O. and Hamida, S., 2020, September. Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques. *In 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet) (pp. 1-6). IEEE.*

15. Gayathri, B.M. and Sumathi, C.P., 2015, December. Mamdani fuzzy inference system for breast cancer risk detection. *In 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-6). IEEE.*

16. Elouedi, H., Meliani, W., Elouedi, Z. and Amor, N.B., 2014, August. A hybrid approach based on decision trees and clustering for breast cancer classification. *In 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR) (pp. 226-231). IEEE.*

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

18. Ara, S., Das, A. and Dey, A., 2021, April. Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. *In 2021 International Conference on Artificial Intelligence (ICAI) (pp. 97-101). IEEE.*

19. Alyami, R., Alhajjaj, J., Alnajrani, B., Elaalami, I., Alqahtani, A., Aldhafferi, N., Owolabi, T.O. and Olatunji, S.O., 2017, February. Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural

network and support vector machines. *In 2017 International Conference on Informatics, Health & Technology (ICIHT) (pp. 1-7). IEEE.*

20. Bhardwaj, A., Tiwari, A., Chandarana, D. and Babel, D., 2014, October. A genetically optimized neural network for classification of breast cancer disease. *In 2014 7th International Conference on Biomedical Engineering and Informatics (pp. 693-698). IEEE.*

21. Addeh, A., Demirel, H. and Zarbakhsh, P., 2017, September. Early detection of breast cancer using optimized ANFIS and features selection. *In 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 39-42). IEEE.*

22. Yassi, M., Yassi, A. and Yaghoobi, M., 2014, February. Distinguishing and clustering breast cancer according to hierarchical structures based on chaotic multispecies particle swarm optimization. *In 2014 Iranian Conference on Intelligent Systems (ICIS) (pp. 1-6). IEEE*.

23. Kummel, H., Abdulah, D. and Al-Tuwaijari, J.M., 2019, June. Cancer classification using gaussian naive bayes algorithm. *In 2019 International Engineering Conference (IEC) (pp. 165-170). IEEE.*

24. MThilagaraj,N.Arunkumar,and Petchinathan Govindan,2022 January Classification of Breast Cancer Images by Implementing Improved DCNN with Artificial Fish School Model

25. Lamiss Mohamed Abd El Aziz Sad, Naglaa Lofty Dabees, Dareen Abd El-Aziz Mohamed, Amr Tageldin, and Samar Galal Younis 2022 January Assessment of Suspected Breast Lesions in Early-Stage Triple-Negative Breast Cancer during Follow-Up after Breast-Conserving Surgery Using Multiparametric MRI