ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

Article Title: **A Comprehensive Review of Breast Cancer Detection Using Machine Learning and Deep Learning Classifiers**

A Comprehensive Review of Breast Cancer Detection Using Machine Learning and Deep Learning Classifiers

Abishajey J.B[1], T.DharmaRaj[2], L. Nisha Evangelin[3]

[1]Assistant professor, Department of Electronics and Communication Engineering, Narayanaguru College of Engineering, Tamil Nadu, India. abishajey1995@gmail.com

[2]Professor, Department of Electrical and Electronics Engineering, MuthayammalEngineering College, Rasipuram, Namakkal District, Tamil Nadu, India. vinseeedharma@gmail.com

[3]Associate professor, Noorul Islam Center for Higher Education, Kumaracoil, Tamil Nadu, India. nishakingston607@gmail.com

ABSTRACT

Breast cancer is the major critical disease and suffered many people around the world. Recent progress in Machine Learning (ML) and Deep Learning (DL) techniques has established as valuable tools in breast cancer detection, enhancing both precision and effectiveness. This paper contribute a comprehensive review of breast cancer prediction utilize ML algorithms and DL model. ML is frequently employed in classifying breast cancer pattern due to exhibiting a crucial feature detection from composite breast cancer datasets. ML techniques such as Random Forests (RF), K-Nearest Neighbors (KNN), Logistic Regression (LR) and Support Vector Machines (SVM) are evaluate to classify and predict breast cancer based on various features. DL techniques including, Artificial Neural Network (ANN), Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), models have extracting attribute from raw imaging data, autonomously learning and offering high accuracy and sensitivity. In this analyses determine the most effective method for handling large datasets while maintaining high prediction accuracy.

Keywords: Machine Learning, Deep Learning, Random Forests, K-Nearest Neighbors, Logistic Regression, Support Vector Machine, Artificial Neural Network, Convolutional Neural Network, Generative Adversarial Network.

1 Introduction

Breast cancer is the severe and prevailing cancers affecting women, leading to substantial number of deaths globally. It is a condition where cancer cells develop in the breast tissues. Cancer begins in the human cells, which are fundamental units of tissues in the breast or elsewhere in the body. While it's important to monitor any breast lump or changes, it is best to have them evaluated by a doctor. Nevertheless, early diagnosing breast cancer is vital to cancer related mortality and elevate patient's overall well-being. Age, sex, genetics, and having dense breasts are the most recognized and common factors that increasing the prospect to breast

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

cancer in women. Women with dense breast tissue face a greater risk of breast cancer and this risk is associated with age. Typically younger women have denser breast tissue compared to older women [1].

In the earlier stage various types of imaging are used for breast cancer including Mammograms, Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), Ultra sound and Histology Imaging. Early detection of breast cancer, which reduces mortality rates, is accomplished through mammograms. CT scan involves capturing images of the breast from multiple perspective as the patients enters in a closed machine, and a computer compiles these images to generate a detail view of the breast. MRI is often used to monitor to assess how well breast cancer patients are responding to chemotherapy before considering breast amputation. Ultrasound imaging which relies on the reflection or echo of sound waves is examine as more secure and efficient compared to X-rays. Micro anatomy of cells, organs, and tissues is examined through analysing the correlation between function and structure with images produced using a microscope in histological studies. [2].

Mammogram are employed for early detection of breast cancer because they are comparatively inexpensive and strong sensitive to small abnormalities. However, in the diagnosis process, the accuracy is frequently compromised by various factors, including interruptions, radiologist personal proficiency, the elaborate structure, and the difficulty of identifying early-stage affliction. Computer Aided Diagnosis (CAD) used for breast cancer aims to resolve these challenges. CAD systems primarily concentrate on detecting and diagnosing breast cancer [3].

Nowadays, Breast cancer treatment enhance the prognosis and survival rates by enabling patients to accept prompt medical care. Several researchers have explored the forecasting of breast cancer through ML techniques including Extreme Gradient Boosting (XGB), RF, SVM and KNN to detecting breast cancer in earlier stage. These classification has been evaluated with regard to proficient and efficacy. Furthermore, a classifier based on association learning which integrates the LR and stochastic gradient descent model to enhance the accuracy for detecting breast cancer [4].

DL techniques is the fundamental data analysis technology and has proven increasingly effectiveness in detecting breast cancer. DL such as, CNN, ANN, Recurrent Neural Networks (RNN), ADA BOOST, GANs. CNN area deep super vised learning technology that efficiently classify breast cancer prognoses. CNN-based algorithms are designed with lesions illustration to improve specialist's capacity for identifying and accurately analyzing the early stage of breast cancer. These model has enhanced the precision and decrease the False- Positive Ratio (FPR) for diagnosing breast cancer [5].

This study produce the comprehensive review of breast cancer detection using ML and DL classifiers from mammogram images. By evaluating these approaches and estimating their benefits and limitations, the performance and accuracy of the system is increased.

The most significant goal of this review is to identify breast cancer by using ML techniques

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

and DL techniques. The optimization of these approaches are intensively discussed below.

1. To conduct the comprehensive analyse of breast cancer through ML techniques.
2. To evaluate the comprehensive analyse of breast cancer through DL techniques.
3. To forecast prognosis and survival rate of breast cancer for the patients.
4. To make better accuracy and performance of these techniques.

2 Review of Comprehensive Analysis of Breast Cancer by Using Machine Learning Methods

2.1 Logistic Regression (Lr) Classifiers

Logistic regression is the ML techniques used to detect breast cancer by predicting tumor malignancy based on various input features. This model trained using a dataset that includes tumor characteristics, such as age, tumor size, and whether the tumour is malignant. The concept behind the LR algorithm is based on linear regression, and it solve the numerical sensitivity challenges of linear regression. In contrast of linear regression which produces continuous and unbounded values, logistic regression maps the output result using a sigmoid function. The equation for this function expressed as follows:

$$(Z) = \frac{1}{1+e^{z}} \tag{1}$$

Let z = W^T. X, weight is represented by W and the feature vector represent by X. (z) signifies the forecasted value associated to the occurrence with obtained from the sample. Specially, the recall rate effectively represents the proportion of deadly tumors correctly identified individuals with breast cancer. A more accurate prediction by the model is indicates by a higher recall rate. Nevertheless, accurately analysing breast cancer is complex and time- consuming task [6].

2.2 Support Vector Machine (Svm) Classifier

SVM is a type of super vised learning method employed for classifying breast cancer. It delivers the greatest precision when applying large datasets. Performance metrics are used to assess the precision rate in the SVM model across several kernels including, linear kernel, Radial Basis Function (RBF) kernel, polynomial kernel and sigmoid kernel. The sigmoid kernel function is,

$$(x_i,x_j)=\tan h\ (\gamma x_i.\ x_j+r), \gamma>0, r\geq 0 \tag{2}$$

It transforms the input data into a higher dimensional space, allowing the SVM to handle nonlinearly separable data. When sigmoid kernel function is applied, SVM operates similarly to a multilayer perceptron neural network. However, a significant challenge in using ML techniques for breast cancer detection is obtaining access to datasets [7].

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

2.3 K-Means and Gaussian Mixture Model (GMM) Classifiers

The author of [8] describes a K-means and GMM algorithms is accurately identify and distinguish between cancerous and non-cancerous breasts tissue. The primary focus is on predicting the precise location of tumor with in the breast and classifying images as benign and malignant using these algorithms. Mammogram images, enhanced through K-means and GMM segmentation improve the qualitative identification of breast cancer there by aiding physicians diagnosing cancer at early stage. The Analysis of Variance (ANOVA) is a statistical procedure used to identify significant difference among variable within a dataset. The below equation evaluate their fluence of one or more elements by analogize theme and values of various sample groups,

$$\sigma = \sqrt{\frac{1}{\sum(x-\bar{x})^2}} \tag{3}$$

The ANOVA propose an enhance in the accuracy rate of the suggested breast cancer evaluation metrics. The combined model increasing a better breast cancer detection. The implication of this analysis is aimed at reducing error rate, boosting accuracy and achieve a superior Signal-to-Noise Ratio (SNR). However, there is a need to improve the precision and segmentation models in subsequent studies.

3 Review of Comprehensive Analysis of Breast Cancer by Using Deep Learning Methods

3.1 Deep Learning Assisted Efficient Ada Boost Algorithm (DLA-EABA)

DLA-EABA is an advanced computational technique for identify and predicting breast cancer. The DLA-EABA model is predict tumors although they are locate within dense tissue over the pectoral muscles. This approach offer high level perfection in identifying breast cancer masses and improves patient's survival rates. It aims to perform binary classification, effectively distinguishing between positive and negative cases. Ada Boost is an effective technique that enhance classifier accuracy and simplifies the learning algorithm by converting a less effective classification model into a more robust one. The output of $h\ (y')$ is 1 if$y'$is categorized as a positive instance or else 0. This differing limits confines lower classifiers to site on just one featuref_i, resulting in a model where each lower classifier is defined by a single featuref_i, a threshold $\theta_i$and a equality $q_i$which both −1.

$$h(y') - y^1 = \begin{cases} if\ q \\ if\ (y') < q_i\theta_i\ N, \begin{cases} i = 1 \\ i = i \end{cases} \\ 0\ otherwise \end{cases} \tag{4}$$

The boosting algorithm evaluate θ_i is a positive value and q_i represent for weak classifier $h\ (y')$. However, the boosting algorithm require a high quality and large data sets [9].

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

3.2 Convolutional Neural Network (CNN)

CNN are a type of deep neural networks primarily apply for image categorization, including the detection of breast cancer. CNNs are utilized to attain high accuracy rates in distinguishing cancerous and non-cancerous breast tissue. This CNN architecture comprises two types of convolutional layers, the first layer detect low-grade features like corners, curves in 2D-images and edge while the second layer develop high-grade features based on the low- level characteristic identified by the initial layer.

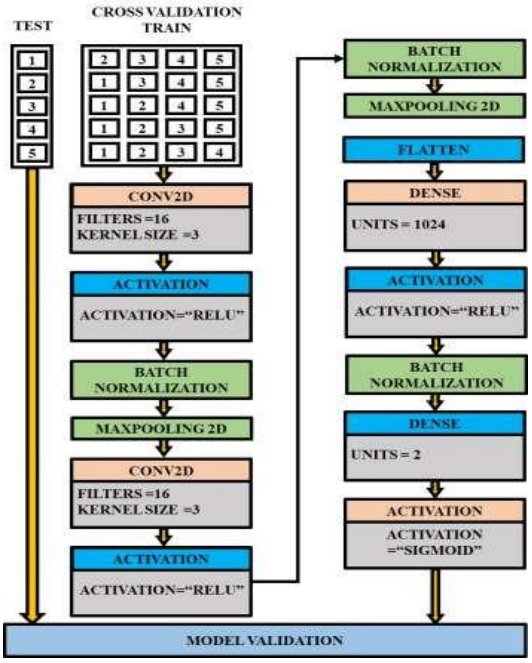

Figure1: *CNN diagram*

The CNN architecture employ two activation function namely the Rectified Linear Unit (ReLU) and sigmoid functions. ReLU is highly successful and excellent experimental outcomes. It features a straight forward formula and expressed as:

$$(x)=max\,(0,x) \tag{5}$$

This formula indicates the output 0 for negative values and return the input for positive values x. The sigmoid activation function expressed as follows:

$$(x)=\frac{1}{1+e^{-x}} \tag{6}$$

Therefore, the sigmoid function is inherently nonlinear, with *f(x)* significantly reply to swap in *x* within the scope of −2 to 2. This characteristic tends to push *f(x)* towards the extremes of the curve, facilitating a clear distinction in classification. However, CNN require a large data set and time [10].

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

3.3 Artificial Neural Network (ANN)

The work [11] propose an ANN, are involve an essential function in the diagnosing of breast cancer. It identifies diagnosing breast cancer through mammograms from the Breast Cancer Digital Repository (BCDR) dataset and it is simple to optimize the training parameters. It is important to emphasize that ANNs are not intended to substitute radiologists but to assist in enhancing their accuracy and reliability. ANN models generally include three layers: the input layer, one or more hidden layers, and the output layer. An ANN with several hidden layers excels at tackling intricate issues. Conversely, a neural network with fewer layers is simple to design and train, and optimizing the training parameters is easier. Additionally, with a limited dataset, it achieves superior generalization efficiency. Nevertheless, it is struggles with high-dimensional data.

4 Comparative Analysis

An evaluative comparison of breast cancer detection using various ML and DL classifiers are discussed below in Table.1 and Table.2.

Table1: *Comparable analysis of breast cancer detection by using ML techniques*

SI. NO	Author / Year of publication	Methodology	Advantages	Limitations
1.	Ahmad S. Elkorany el at(2022) [12]	This study propose a hybrid ML model that combines the Whale Optimization Algorithm (WOA) and Dragon fly Algorithm(DA) For diagnosing breast cancer.	This model increase the accuracy of breast cancer classification.	This method required a large datasets for long training time.
2.	Bassam m. Kanberelat (2024) [13]	This paper presented a Light GBM (LGBM) algorithm to analyze histopathology images for breast cancer diagnosis.	This techniques is highly effective in handling large datasets.	However, dataset makes hard to find different population and imaging condition.

Table 2: *comprehensive analysis of breast cancer detection using DL techniques*

SI. NO	Author / Year of publication	Methodology	Advantages	Limitations

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

1.	Jingyao Li etal(2021) [14]	A Multi-Scale U-net(MF U-NET) model developed For detect breast cancer From ultrasound images.	This model Achieves the highest recall on the Breast Ultrasound Image Segmentation Benchmark (BUSIS) dataset.	However, this Model require to better robustness and improve spatial complexity.
2.	A beer Saber etal(2021) [15]	A Visual Geometry using Mammographic Image Analysis Society (MIAS) dataset.	This model achieves the best accuracy, and F-Score of brain tumor diagnosis.	Nevertheless, the small size of Dataset not suitable To train the model.

Table 3: *Comparative analysis of accuracy in ML techniques*

Methods	Accuracy
SVM[7]	96.4%
K-MEAN Sand GMM [8]	95.50%
LGBM[13]	95.16%

Table. 3representsanaccuracycomparisonofbreastcanceranalysisusingMLtechniques. From the table that the SVM [7] model exceeds the LGBM [13] model and k-means and GMM [8] models interms of accuracy with 96.4%.

Table 4: *Comparative analysis of accuracy in DL techniques*

Method	Accuracy
ADABOOST[9]	97.2%
CNN[10]	98.76%
MFU-NET [14]	93.45%

Table.4 represents an accuracy comparison of breast cancer analysis using DL techniques. From the table that the CNN [13] model exceeds the ADABOOST [9] and MF U-NET [14] models in terms of accuracy with 98.76%.

5 Conclusion

The frequency and death rate of breast cancer are rising annually, making it the most prevalent and perilous cancer among women globally. Therefore, society requires advanced techniques for the prompt identification of breast cancer. To address this issue researchers have established a numerous methods to enhance the categorization for detecting breast cancer. The methods aims to identify the most effective algorithm and the number of training data sets to predict the more precise classification outcomes for breast cancer. In this research, numerous ML and DL

techniques are reviewed for the breast tumor detection from mammography images with high accuracy and performance. Initially, ML models suchas, SVM, LR, K-MEANS and GMM are compared for the analysis of breast cancer. Likewise, DL techniques including, CNN, ADABOOST, ANN, MF-UNET are compared for breast cancer analysis. It is verified that the CNN based on DL model results better performance and higher accuracy rate when compared to several ML techniques. The future prospects of this approach indicate improved results inaccuracy and plan to employ the use of modern optimization algorithm for multi class classification. This advanced health care model is transforming the medical field by addressing societal issues, particularly, for the early detection of breast cancer in women.

References

1. Usman Naseem; Junaid Rashid; Liaqat Ali; Jungeun Kim; Qazi Emad Ul Haq; Mazhar Javed Awan; Muhammad Imran, Year: 2022, “An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers”, IEEE Access, Vol: 10, pp. 78242 – 78252.

2. Mohammed Abdulla Salim Al Husaini; Mohamed Hadi Habaebi; Shihab A. Hameed; Md Rafiqul Islam; Teddy Surya Gunawan, Year: 2020, “A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks”, in IEEE Access, Vol: 8, pp. 208922 – 208937.

3. Zhiqiong Wang; Mo Li; Huaxia Wang; Hanyu Jiang; Yudong Yao; Hao Zhang; Junchang Xin, Year: 2019, “Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features”, IEEE Access, Vol: 7, pp. 105146 – 105158.

4. Amreen Batool; Yung-Cheol Byun, Year: 2024, “Towards Improving Breast Cancer Classification using an Adaptive Voting Ensemble Learning Algorithm”, IEEE Access.

5. Noreen Fatima; Li Liu; Sha Hong; Haroon Ahmed, Year: 2020, “Prediction of breast cancer, comparative review of machine learning techniques, and their analysis”, IEEE Access, Vol: 8, pp. 150360 – 150376.

6. Hua Chen; Kehui Mei; Yuan Zhou; Nan Wang; Guangxing Cai, Year: 2023, “Auxiliary Diagnosis of Breast Cancer Based on Machine Learning and Hybrid Strategy”, IEEE Access.

7. P. Esther Jebarani; N. Umadevi; Hien Dang; Marc Pomplun, Year: 2021, “A novel hybrid K-means and GMM machine learning model for breast cancer detection”, IEEE Access, Vol: 9, pp. 146153 – 146162.

8. Jing Zheng; Denan Lin; Zhongjun Gao; Shuang Wang; Mingjie He; Jipeng Fan, Year: 2020, “Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis”, IEEE Access, Vol: 8, pp. 96946 – 96954.

9. Murtada K. Elbashir; Mohamed Ezz; Mohanad Mohammed; Said S. Saloum, Year: 2019,

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 10-18

"Lightweight convolutional neural network for breast cancer classification using RNA-seq gene expression data", IEEE Access, Vol: 7, pp. 185338 – 185348.

10. Muhammad Firoz Mridha; Md Abdul Hamid; Muhammad Mostafa Monowar; Ashfia Jannat Keya; Abu Quwsar Ohi; Md Rashedul Islam; Jong-Myon Kim, Year: 2021, "A comprehensive survey on deep-learning-based breast cancer diagnosis", Cancers, Vol: 13, no: 23, pp. 6116.

11. Ahmed S. Elkorany; Mohamed Marey; Khaled M. Almustafa; Zeinab F. Elsharkawy, Year: 2022, "Breast cancer diagnosis using support vector machines optimized by whale optimization and dragonfly algorithms", IEEE Access, Vol: 10, pp. 69688 – 69699.

12. Bassam M. Kanber; Ahmad AL Smadi; Naglaa F. Noaman; Bo Liu; Shuiping Gou; Mutasem K. Alsmadi, Year: 2024, "Light GBM: A Leading Force in Breast Cancer Diagnosis Through Machine Learning and Image Processing", IEEE Access.

13. Jingyao Li; Lianglun Cheng; Tingjian Xia; Haomin Ni; Jiao Li, Year: 2021, "Multi-scale fusion U-net for the segmentation of breast lesions", IEEE Access, Vol: 9, pp. 137125 – 137139.

14. Abeer Saber; Mohamed Sakr; Osama M. Abo-Seida; Arabi Keshk; Huiling Chen, Year: 2021, "A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique", IEEE Access, Vol: 9, pp. 71194 – 71209.