IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006155|

Survey on Breast Cancer Classification using Machine Learning

Sundari David, Prof. Sandeep Kumar

M.Tech Student, Dept. of Computer Science & Engineering, Cambridge Institute of Technology, Bangalore, India

Professor, Dept. of Computer Science & Engineering, Cambridge Institute of Technology, Bangalore, India

ABSTRACT: One of the most deadly diseases impacting women worldwide is breast cancer. The primary reason why women get breast cancer is because of the growth of malignant cells in the breast area. The life expectancy of the affected women would enhance with timely breast cancer mass (BCM) detection. This study draws on data from a survey of several journals, the majority of which use Machine Learning (ML), while a lesser number of studies identify BCM using Deep Learning (DL). On various image datasets made up of mammograms and ultrasound images, ML techniques like Support Vector Machine (SVM), Naive Bayes, Random Forest, and DL algorithms like Artificial Neural Network (ANN) are employed for classifying the BCM.

KEYWORDS: Machine Learning (ML), Deep Learning (DL), Breast Cancer Masses (BCM)

I. INTRODUCTION

The abnormal mammary cell development leads to BCM. In the women's body, there are cancerous cells forming a malignant or harmful mass which spreads throughout the mammary gland. Indian Council of Medical Research (ICMR) claims, breast cancer recorded for about two lakh cases i.e., 14.8 percent in the year 2020. BCM is determined by the malignant cells and could appear in any place within the mammary gland. The majority of BCM occurs in the breast lobules or ducts. As a result, timely recognition of BCM is critical in improving patient's chances of survival. Some procedures like mammography, biopsy and ultrasound are used for detecting BCM.

In this work, the survey is done taking various papers used for predicting the BCM. Most of them have employed the traditional ML algorithms, and some have used ensemble learning and DL techniques and even concepts like data mining are employed to detect the BCM and predict the breast cancer disease based on the different mammogram and ultrasound image datasets. The traditional ML algorithms perform two tasks such as classification and regression. They can be used to identify texts or detect diseases from the medical images. The input & output layer, and an extra layer can all be added by the DL algorithms in order to learn from the data. Random Forest, SVM, Naive Bayes (NB), k-Nearest Neighbor (kNN), Logistic Regression, AdaBoost, and Decision Trees (DTs) are a few examples of ML algorithms. Long Short Term Memory (LSTM), Convolution Neural Networks (CNN), and ANN are examples of DL algorithms.

II. LITERATURE SURVEY

Based upon dataset & parameters settings, every algorithm functions differently. The KNN method produced the finest outcome for the research on the whole. In the screening of BCM, logistic regression & NB also showed promising results. Given how SVM is a powerful tool in prediction assessment, we draw the conclusion as to SVM with a Gaussian kernel is the best method for predicting whether the breast cancer will reoccur or not [1].

We employed a repository from the University of Wisconsin Hospital for this study, which contained 30 parameters that describes the characteristics of the mammary mass's nucleus. Like we demonstrated in the findings part, the DNN classifier performs excellently in terms of accuracy score (92 percent), showing superior outcomes compared to conventional methods. Comparable to other documented research from the past, Random Forest 50 & 100 produced the greatest outcomes for the ROC curvature measure [2].

This research proposes a system that compares various ML methods for the assessment of BCM. The Wisconsin Diagnosis Breast Cancer data set has been used to compare the efficiency of various ML classifiers methodologies. Every classifier's accuracy in distinguishing malignant or benign tumor calls was found to be greater than 94 percent. It

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006155|

was discovered that kNN is the most efficient method for detecting breast malignancy because it outperformed the other methods in terms of precision, F1 score & accuracy [3].

Evaluating the efficiency of several implemented ML approaches using important performance indicators including precision, accuracy, ROC & recall area. According to the indicators used to measure the effectiveness of the used ML methods, SVM based Sequential Minimal Optimization (SMO) method demonstrated the finest result with an accuracy of 96.9957 percent for detection & forecasting out from WBC datasets [4].

Comparative assessment of various ML classifiers, including NB, DTs, SVMs, Vote (DT+NB+SVM), AdaBoost & Random Forest was done. Employing the ML application rapidminer, empirical samples were retrieved from the Wisconsin database about breast cancer malignancy. We discovered that DL via the Exprectifier activator mechanism has an accurate results of 96.99 percent with10-fold cross - validation [5].

The majority of women experienced breast cancer as a result of being unaware about it. Here, we employed two ML approaches. Employing the approaches, we were able to get accuracy rates of 73.63 percent for XGBoost & 74.73 percent for Random Forest. We contrasted the findings of our investigation with those of other models and discovered that our method outperformed the existed models [6].

The outcomes would help in selecting the optimal ML technique to build an automated approach for diagnosing breast cancer masses. Our research leads us to the conclusion, as to both SVM & Random Forest gave the highest accurate results, at 96.5 percent. In the upcoming time, we'll endeavor to improve our attempt by managing a sizable dataset & including further features like breast cancer stage identification etc [7].

The answer to treating breast cancer is the earliest identification of it. This work suggested a classifying method for predicting breast cancer called stacked classifier, which had a 97.20 percent accuracy rate. According to the accuracy %, a malignant tumor can be found in the optimum situation [8].

Whereas if malignancy region is encircled or compressed by another region which is colored around the same moment as the segmentation stage, this research of mammary segmentation employing the watersheds transformation method includes multiple faults. According to these findings, the overall accuracy for the malignancy region is 88.65 percent. Thus, it is concluded in this study that mammary segmentation using ultrasound pictures could be processed using the watersheds transformation technique methodology [9].

Employing the Wisconsin database, we suggested an adaptable ensemble voting mechanism in this study for confirmed breast cancer. This study compares & explains why ANN & logistical methods perform well if used in combination with ensembles ML techniques to detect breast carcinoma, while the amount of parameters would be decreased. We used the voting method after implementing the logistical & neural networking techniques on these 16 parameters to reach a 98.50 percent accuracy rate. 699 entries with categories features of 30 characteristics can be found in the Wisconsin Database for breast cancer [10]

According to a study on mammary cancer danger in India, 1 in every 28 women will develop it at some point in their lifespan. BCM starts after cancerous cells in the mammary area initiate to develop uncontrollably or start to do so. It then spreads from the mammary tissue. The tumour that results from these cells is typically visible on a mammography or palpable as lumps. This platform employs a number of techniques, including PNN classifiers, adaptive mean & GMM segmentation to forecast if the specified mammography has malignant or benign cell, enabling the individual to identify the condition more quickly & undertake the necessary treatment [11].

A neural categorization approach for the identification of mammary cancer was developed in this work. The suggested methodology comprises of an only one buried neural net with an extreme learning foundation. For the buried neurons, we employed sigmoid activate function having various settings. In order to solve the optimizing issue of determining the ideal ranges for these variables, we employed a genetic technique. This aids in overcoming the ELM method's primary issue, which is its enormous amount of neurons in the buried layer. The suggested methodology was also evaluated against other studies, and the majority of them were excelled [12].

On Wisconsin datasets for mammary cancer, we applied categorization methodology to forecast mammary cancer. Utilization of the four classifiers was made: KNN, GANN, C4.5 & CPSO. We found out that KNN is the top ML method and that GANN is the highest evolving method, scoring 95%. However, GANN was the superior method available, regardless of whether it uses ML or evolving [13].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006155|

III. LITERATURE SURVEY DETAILS

Sl. no	Paper Title	Year	Methodology used	Findings
1	Using Machine Learning algorithms for breast cancer risk prediction and diagnosis	2018	KNN, NB and SVM	Benign (B) and malignant (M) are the two types of cancer cells identified.
2	An Analysis of Machine Learning Classifiers in Breast Cancer Diagnosis	2019	The DTs, Multilayer Perceptron, SVM, Random Forest, and Deep Neural Network are the 6 various categorization techniques that are presented here for comparison.	In order to assess the precision, objectivity, and reproducibility of the detection of malign carcinoma using fine syringe suction, this paper uses ML classifying forecast methods.
3	Breast Cancer Detection Using Machine Learning Algorithms	2018	Comparative assessment of the three most widely employed ML methods—Random Forest, kNN and NB—for predicting malignancy in breast.	With a training set, the Wisconsin Diagnosis BC database was employed for assessing how well each ML methods performed in regards of important metrics like precision & accuracy.
4	Comparison of Machine Learning Methods for Breast Cancer Diagnosis	2019	Consequently, ANN & SVM methods were employed to forecast the categorization of BCM in order to determine which of the ML techniques performed best.	The Wisconsin database for mammary cancer was classified employing the most prevalent ML methods, and the categorization efficiency of these methods was contrasted utilizing the scores of the performance indicators such as of accuracy, recall, ROC Area and precision.
5	Deep Learning Algorithms for Predicting Breast Cancer Based on Tumor Cells	2019	AdaBoost, NB, DTs, SVMs and Random Forest	We evaluated the effectiveness of several DL activate mechanism in the categorization methods.
6	Breast Cancer Risk Prediction using XG-Boost and Random Forest Algorithm	2020	Random Forest and XGBoost	Employing 2 well-known ensembles ML methods, the BC database was examined to provide predictions about mammary cancer.
7	Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms	2021	Classifiers like Logistic Regression, DTs, NB, SVM, kNN, and Random Forest.	In order to examine the effectiveness of several ML techniques for forecasting mammary cancer by analyzing the database.
8	On Predicting and Analyzing Breast Cancer using Data Mining Approach	2020	Methods like Random Forest, kNN, Logistic Regression, Stacking & SVM Classifier.	For assessment & forecasting of BCM employing several categorization methods.
9	Segmentation of Breast Using Ultrasound Image for Detection Breast Cancer	2020	Employing the watersheds transformation technique in ultrasound images segmentation.	Comparing the region of malignancy calculated using healthcare records with test findings based on the outcomes of these experiments.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006155|

10	Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm	2018	An logistic algorithm & ANN	When used with ensembles ML methods in identifying mammary cancer, also with fewer parameters, this research compares, explains, & offers a preferable answer.
11	Breast Cancer Detection using Neural Network Mammogram	2020	GMM segmentation, PNN classifier, & Adaptive Mean	The work employs a number of methods, including as one that predicts if a given mammography contains malignant or benign cells, allowing the individual to identify the condition more quickly & perform the necessary treatment.
12	Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network	2018	Extreme Learning Machine (ELM) method	The suggested approach is an ELM trained neural network having only one buried layer. The usage of various activating mechanism for the buried neurons & its refinement utilizing genetic methods are the key accomplishments of this research.
13	Analogizing of Evolutionary and Machine Learning Algorithms for Prognosis of Breast Cancer	2018	Particle Swam Optimization (CPSO), kNN, C4.5, and Genetic Algorithm for Neural Networks (GANN)	This work compares the accuracy percentage of several techniques using a categorization strategy employing k-fold cross-validation. The outcome can be examined based on the tumor, that is divided as 2 categories such as Malignant & Benign.

Table 1 I. Literature Survey Details:

IV. CONCLUSION

In this paper, survey on various papers is done; most of them use the ML algorithms in classifying the BCM. Most of the algorithms give accuracy of more than 90 percent and some give low accuracy based on the image dataset used. The different ML and DL are efficient and can be used for classifying of BCM using various datasets consisting of mammogram and ultrasound images. Not many have touched the concept of transfer learning with the ensemble model. Many ML algorithms are used to provide good accuracy and even comparisons are done among the algorithms to see which one of them provides the best accuracy. All these ML methods will help in the early breast cancer classification and employ in improving the survival rate of the women affected by it. Future efforts can use the Transfer Learning approach with ensemble for obtaining better results.

REFERENCES

1. Bharat, A., Pooja, N. and Reddy, R.A., 2018, October. Using machine learning algorithms for breast cancer risk prediction and diagnosis. In 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C) (pp. 1-4). IEEE.
2. Teixeira, F., Montenegro, J.L.Z., da Costa, C.A. and da Rosa Righi, R., 2019, September. An analysis of machine learning classifiers in breast cancer diagnosis. In 2019 XLV Latin American computing conference (CLEI) (pp. 1-10). IEEE.
3. Sharma, S., Aggarwal, A. and Choudhury, T., 2018, December. Breast cancer detection using machine learning algorithms. In 2018 International conference on computational techniques, electronics and mechanical systems (CTEMS) (pp. 114-118). IEEE.
4. Bayrak, E.A., Kırcı, P. and Ensari, T., 2019, April. Comparison of machine learning methods for breast cancer diagnosis. In 2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT) (pp. 1-3). IEEE.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006155|

5. Mekha, P. and Teeyasuksaet, N., 2019, January. Deep learning algorithms for predicting breast cancer based on tumor cells. In 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON) (pp. 343-346). IEEE.
6. Kabiraj, S., Raihan, M., Alvi, N., Afrin, M., Akter, L., Sohagi, S.A. and Podder, E., 2020, July. Breast cancer risk prediction using XGBoost and random forest algorithm. In 2020 11th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-4). IEEE.
7. 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.
8. 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.
9. Khasana, U., Sigit, R. and Yuniarti, H., 2020, September. Segmentation of Breast using ultrasound image for detection breast cancer. In 2020 International Electronics Symposium (IES) (pp. 584-587). IEEE.
10. 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.
11. Nagpure, R., Chandak, S. and Pathak, N., 2020, February. Breast Cancer Detection using Neural Network Mammogram. In 2020 International Conference on Convergence to Digital World-Quo Vadis (ICCDW) (pp. 1-6). IEEE.
12. Nemissi, M., Salah, H. and Seridi, H., 2018, November. Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network. In 2018 International Conference on Signal, Image, Vision and their Applications (SIVA) (pp. 1-4). IEEE.
13. Sethi, A., 2018, August. Analogizing of evolutionary and machine learning algorithms for prognosis of breast cancer. In 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 252-255). IEEE.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details