Review of Breast Cancer Prediction using Machine Learning Techniques

Aman Shakya
Research Scholar
Department of CSE
TIEIT, Bhopal

Kaptan Singh, PhD
Assistant Professor
Department of CSE
TIEIT, Bhopal

Amit Saxena
Assistant Professor
Department of CSE
TIEIT, Bhopal

ABSTRACT
Breast cancer is a malignant neoplasm that develops in the breast tissues. Among the top causes of mortality in women, it is the most frequent form of cancer in females worldwide. Machine and deep learning approaches based on AI can successfully locate the prediction model. In this study, we examine the state of the art in applying machine learning to forecast breast cancer. The prediction model's important parameters are its accuracy and error rate.

Keywords
Breast Cancer, AI, Machine Learning, Accuracy, Model, Prediction.

1. INTRODUCTION
As a major cause of mortality among women, breast cancer is a serious health concern. The rapid growth of cancerous breast cells is what causes the disease. If breast cancer is detected at an early stage, more treatment options will be available, and survival chances will improve. A variety of screening procedures using PC-supported location frameworks have been developed for effective analysis and treatment of breast cancer. In the medical and health care field, visual data plays a crucial role. Deep learning is used to extract highlights from image collections, since deep learning algorithms can rapidly and accurately filter out irrelevant data. Existing methods, such as mammography screening and biopsy, for examining and detecting breast cancer have been greatly aided by significant advancements [1]. After heart disease, breast cancer is the leading cause of mortality among women worldwide. It might be a step forward for radiologists to diagnose and treat breast cancer. Because of this, sickness prediction and mortality have progressed farther due to critical consideration. The main goal of this study is to determine whether or not earlier diagnosis increases treatment options and the likelihood of a full recovery. By combining traditional methods of subdividing with cutting-edge machine learning techniques, this investigation demonstrates the flexibility of the concept and opens up new research frontiers. It is common practise to employ a flexible middle channel in the pre-handling process for purposes including noise removal, image enhancement, edge preservation, and smoothing [2].

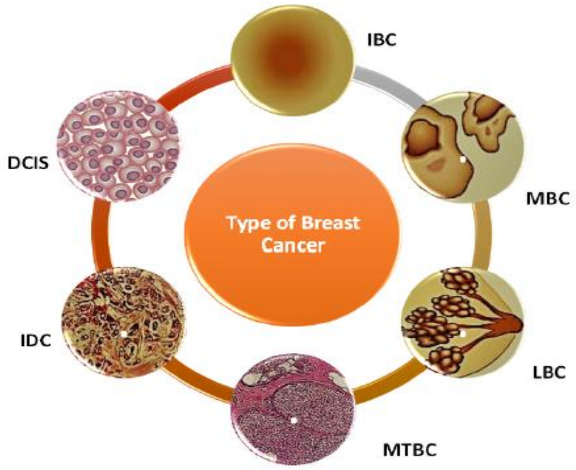

Figure 1: Major types of Breast Cancer [1]

Understanding how much of a model's dynamic interaction can be deduced by both model developers and end users is what we call "interpretability" in the context of machine learning (ML) models. For example, cancer prediction models based on transcriptomics have been shown to be very accurate; nevertheless, they are notoriously difficult to interpret because to the large layered highlight space and complex nature of the models themselves. Several methods for improving the interpretability of inconsistent classifiers have been presented, since this feature is fundamental to the transparency and integrity of ML models. In spite of this, evaluating these estimates often calls very substantial spatial data [3].

When it comes to cancer, breast cancer (BC) is a major cause of death in women. When diagnosed early, people with BC have a better chance of receiving effective treatment and living a long, healthy life. To help with the automated detection and analysis of the BC thinking area based on the 80-20 and cross-approval procedures, this study develops a new deep learning (DL) model based on the transfer learning (TL) approach. It is shown that DL architectures are problem explicit. Learning from past experiences allows TL practitioners to apply lessons learned in one context to future problems [4]. Clinicians and scientists have a great test case in the breast cancer grouping. In the field of cancer research, neural networks have recently gained popularity as a useful tool for organising data. Using state-of-the-art computational methods, this study proposes the numerical implementation of Deep Learning to Enable Effective Adaboost Analysis (DLA-EABA) for the detection of breast cancer. Despite the advances in traditional PC vision, advanced convolutional neural networks (CNNs) have shown to be the most effective method for cancer classification using motion analysis [10].

2. LITERATURE REVIEW

C. Chen et al.,[1] to put out a unique diagnostic approach using CEUS movies. A three-dimensional convolutional neural network forms the basis of the model. To be more precise, we find that radiologists often use one of two distinct patterns while watching CEUS recordings. Two commonalities emerge: (1) a fixation on discrete intervals of time, and (2) an interest in contrasting CEUS and US frames. In order to account for these two trends in our deep learning model, we create a temporal attention module informed by domain knowledge and a channel attention module that focuses on specific channels. We use our own Breast-CEUS dataset of 221 instances to verify our model's accuracy. These findings demonstrate that our model has a sensitivity of 97.2% and an accuracy of 86.3%. In instance, adding domain knowledge increases sensitivity by 3.5% and specificity by 6.0%.

According to P. E. Jebarani et al.[2] This study makes a significant commitment by suggesting a new limit for evaluating the performance of K-implies with a Gaussian mixture model (GMM). In the case of breast cancer, a hybrid approach combining categorization and classification was used. It is crucial to distinguish between benign and dangerous malignancies, and the suggested approach would do just that. In order to determine the efficacy of this method for the prompt diagnosis of breast cancer, the replicated data are analysed and evaluated. Professional medical personnel can detect breast cancer more rapidly and with more precision using this method. The suggested strategy's multi-variable analysis and expected rate were determined using an ANOVA test.

N. Adnan et al. [3] offer a model for predicting the spread of breast cancer that makes use of a few easily recognisable features and a simple but novel approach to comprehending models that allows for individualised interpretations. We also provided what seems to be the first comprehensive method for statistically comparing and contrasting various estimations of comprehension. Experimental findings demonstrate that our model not only met but exceeded expectations for precision, but also achieved greater levels of translation consistency amongst classifiers than state-of-the-art comprehension strategies. Our translated findings have the potential to improve the generalizability of expectation models, which is of paramount importance.

Saber, A., et al.,[4] The suggested model uses pre-trained convolutional neural network (CNN) technology, such as the Birth V3, ResNet50, Visual Math Gathering Organizations (VGG)-19, VGG-16, and Origin V2 ResNet, to remove the highlights from the Mammographic Image Analysis Society (MIAS) dataset. The suggested model's performance has been evaluated using six metrics: accuracy, F-score, ROC area, precision, responsiveness, and responsiveness to outliers (sensitivity). Trial findings reveal that the VGG16 model's TL is robust for BC conclusion by clustering mammography breast images with respective values of 98.96%, 97.83%, 99.13%, 97.35%, 97.66%, F-score, and AUC for the 80-20 method, and 98.87%, 97.27%, 98.2%, 98.84%, 98.04%, and 0.993.

Successful image pre-handling algorithms are proposed by A. R. Beeravolu et al. [5] to generate datasets that save processing time for brain organisation and further improve precision and arrangement rates. This investigation suggests methods for doing so, including the removal of the foundation, the removal of the pectoral muscles, the addition of noise to the images, and the enhancement of the images themselves. Information visualisations for the brain network may be made more vivid by adding noise without altering the nature of subtlety in the visuals, which may improve the presentation of the brain network model in practise.

An analogous investigation of breast cancer prediction using machine learning, deep learning, and data mining is presented by N. Fatima et al. [6]. While many researchers have focused on breast cancer research and visualisations, multiple approaches have been used, each with its own accuracy rate that varies with context, methodology, and data collection. Our primary focus is on critically dissecting current Machine Learning and Information Mining processes to identify the best suitable technique that would maintain the massive dataset with high prediction accuracy.

Using a total of nine separate credits—age, weight list, glucose, insulin, and an assessment of a homeostasis model—H. - J. Chiu et al. [7] offered a different handling technique for predicting breast cancer. In the first place, principal component analysis (PCA) was used to identify crucial parts of the data and further reduce the components of the data. The sum of the five most important elements amounted to 99.89%. As a follow-up, we used the multi-facet perceptron network (MLP) method to get rid of characteristics remembered for the data, and we built the structure to look at how the data changed as the aspects grew or shrank.

Proposed models by Y. Yari et al.[8] have undergone in-depth hyperparameter analyses in both amplification-ward and amplification-free characterisation settings. The suggested framework has achieved an accuracy of up to 98% in multiclass settings. The suggested framework ensures up to 100% accuracy at the double order level. The results surpass the accuracy of previous studies in fully described performance metrics for breast cancer CAAD models derived from histology images.

Benefits and risks of breast multi-imaging modalities, division plans, inclusion extraction, and arrangement of breast abnormalities using state-of-the-art deep learning approaches are the topic of study by T. Mahmood et al. [9]. This study also examines different noteworthy data sets with the "Breast Cancer" keyword to present an exhaustive study on existing indicative plans to open-up new examination provokes for radiologists and specialists to intervene as soon as possible to foster a proficient and solid breast cancer visualisation framework utilising unmistakable profound learning plans.

Attractive Reverberation Imaging (X-ray), Ultrasound (US), computerised breast tomosynthesis, and mammography are just few of the imaging modalities that J. Zheng et al.,[10] examine in order to learn how to best depict breast masses for different analytic, prescient tasks or predictive purposes. Some of the layers in the deep learning system include convolutional, long short-term memory (LSTM), and max-pooling. The focus of this study is on combining these machine learning methods with highlight selection and removal techniques by evaluating their outcomes using order and division techniques to determine the best appropriate strategy. Comparing the trial results with other current frameworks reveals a high degree of accuracy (97.2%), awareness (98.3%), and particularity (96.5%).

Z. Wang, et al. [11] The detection, diagnosis, and treatment of breast cancer may be facilitated by using a computer-aided design (CAD) framework that takes mammograms into account at an early stage. However, existing computer-aided design frameworks continue to lack sufficient accuracy. In this study, we examine a breast CAD method that combines the power of convolutional neural networks (CNNs) with the flexibility of component-based CAD. At first, we suggest a population-wide tracking method dependent on deep convolutional neural network (CNN) features and unsupervised extreme learning machine (ELM) clustering. The second step is to compile a set of skills by merging features such as deep highlights, morphological elements, surface elements, and thickness highlights. Third, using the linked training set, an ELM classifier is developed to distinguish between benign and malignant breast tumours.

Using an exchange learning approach and a pre-handling computation, J. Zhang et al. [12] improved the characterisation performance of breast cancer images. In addition, the area under the curve, awareness, and explicitness of AlexNet and GoogLeNet support vector machines suggest that the combination of deep learning with photoacoustic imaging may have a major impact on clinical diagnosis.

3. PROPOSED STRATEGY

- Load the dataset from the Kaggle repository and load it. [13]

The dataset will be downloaded from the kaggle source at this phase of the process. It is a corporation that specialises in offering enormous datasets. After that, load this dataset into the environment that Python is running in.

- Creating a Visualization of the Dataset
- At this point, you should access the dataset files and examine the different data in terms of attributes such as temperature and solar panels.
- Perform preparatory work on the dataset

The data preprocessing stage has now been performed, and at this point, the data is complete and ready to be processed. During this stage, missing data will either be removed or replaced with the form constant one or zero.

- Separating the dataset into testing and training versions

This stage involves separating the final preprocessed dataset into two distinct datasets: the training dataset and the testing dataset. During the training phase of machine learning, a dataset is used to instruct the computer. After that, the machine is put through its paces using the remaining datasets.

- Classification Through the Use of an Algorithm for Machine Learning

At this point, you should use the method of machine learning to determine the performance parameters. The work that was done before used a variety of methods. The concepts of machine learning and deep learning are implemented in the approach that we have suggested.

- Metrics Regarding Performance

(Accuracy, Precision, Recall, F1 - Score) (Accuracy, Precision, Recall, F1 - Score)

At this point, the performance metrics are being determined by utilising the following formulae in terms of precision, recall, f-1 measure, accuracy, and so on.

- TP stands for "true positive," which means that both the prediction and the occurrence were accurate.
- TN stands for "true negative," which means that both the occurrence and the prediction came true.
- A false positive, often known as an FP, is when the event in question turns out to be positive.
- False negative (FN): When the occurrence does not match the prediction and the prediction is false.

$$
\begin{aligned}
Precision &= \frac{|TP|}{|TP| + |FP|} \\
Recall &= \frac{|TP|}{|TP| + |FN|} \\
F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \\
Accuracy &= \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}
\end{aligned}
$$

4. CONCLUSION

Breast cancer patients are more likely to have true health-related ambiguities, which have been linked to increased mortality. One possible reason is that the number of false positives and negatives is higher because of specialised difficulties in imaging features and diverse breast densities. Early intervention is crucial in developing a state-of-the-art forecasting mechanism that may efficiently reduce the complexities of disease and speed up the recovery process. In this study, we use ML techniques to poll women's breast cancer expectations.

5. REFERENCES

[1] C. Chen, Y. Wang, J. Niu, X. Liu, Q. Li and X. Gong, "Domain Knowledge Powered Deep Learning for Breast Cancer Diagnosis Based on Contrast-Enhanced Ultrasound Videos," in IEEE Transactions on Medical Imaging, vol. 40, no. 9, pp. 2439-2451, Sept. 2021, doi: 10.1109/TMI.2021.3078370.

[2] P. E. Jebarani, N. Umadevi, H. Dang and M. Pomplun, "A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection," in IEEE Access, vol. 9, pp. 146153-146162, 2021, doi: 10.1109/ACCESS.2021.3123425.

[3] N. Adnan, M. Zand, T. Huang and J. Ruan, "Construction and evaluation of robust interpretation models for breast cancer metastasis prediction," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, doi: 10.1109/TCBB.2021.3120673.

[4] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk and H. Chen, "A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique," in IEEE Access, vol. 9, pp. 71194-71209, 2021, doi: 10.1109/ACCESS.2021.3079204.

[5] 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," in IEEE Access, vol. 9, pp. 33438-33463, 2021, doi: 10.1109/ACCESS.2021.3058773.

[6] N. Fatima, L. Liu, S. Hong and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," in IEEE Access, vol. 8, pp. 150360-150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[7] H. -J. Chiu, T. -H. S. Li and P. -H. Kuo, "Breast Cancer–Detection System Using PCA, Multilayer Perceptron,

Transfer Learning, and Support Vector Machine," in IEEE Access, vol. 8, pp. 204309-204324, 2020, doi: 10.1109/ACCESS.2020.3036912.

[8] Y. Yari, T. V. Nguyen and H. T. Nguyen, "Deep Learning Applied for Histological Diagnosis of Breast Cancer," in IEEE Access, vol. 8, pp. 162432-162448, 2020, doi: 10.1109/ACCESS.2020.3021557.

[9] T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran and K. U. Rehman, "A Brief Survey on Breast Cancer Diagnostic With Deep Learning Schemes Using Multi-Image Modalities," in IEEE Access, vol. 8, pp. 165779-165809, 2020, doi: 10.1109/ACCESS.2020.3021343.

[10] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He and J. Fan, "Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis," in IEEE Access, vol. 8, pp. 96946-96954, 2020, doi: 10.1109/ACCESS.2020.2993536.

[11] Z. Wang et al., "Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features," in IEEE Access, vol. 7, pp. 105146-105158, 2019, doi: 10.1109/ACCESS.2019.2892795.

[12] J. Zhang, B. Chen, M. Zhou, H. Lan and F. Gao, "Photoacoustic Image Classification and Segmentation of Breast Cancer: A Feasibility Study," in IEEE Access, vol. 7, pp. 5457-5466, 2019, doi: 10.1109/ACCESS.2018.2888910.

[13] https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data