IMPROVED BREAST CANCER PREDICTION USING ENSEMBLE OF MACHINE LEARNING ALGORITHMS

Dr. P.S. Smitha[1], Subiksha N[2], Sruthi Nath C[3], Subashini T[4]

Abstract

Breast cancer, which poses a danger to women's health and is a major cause of mortality, is a major focus of medical image analysis. Digital mammography can greatly increase the accuracy of disease detection by providing an early and accurate diagnosis of breast cancer. The approach uses transfer learning from the pre-trained ResNet-50 CNN on ImageNet to train and classify the breast dataset into benign or cancerous categories. It can be difficult to understand the intricate mammographic pictures, though. Convolutional neural networks (CNNs) have demonstrated promising results for image classification applications, such as breast cancer diagnosis, in recent years. In this regard, ResNet50 and AlexNet are two well-known CNN architectures that have been applied to the detection of breast cancer using mammographic images. In this study, performance evaluations of the machine learning algorithm such as Alexnet, and Restnet 50 using a dataset of breast cancer diagnoses were done. The effectiveness of the strategy was also tested to see how much AdaBoost's usage of weak learners affect. The primary goal is to assess the algorithm’s accuracy, precision, sensitivity, and specificity in terms of its effectiveness and efficiency in detecting data.

Keywords: *Breast cancer, ResNet50, Ada boost, AlexNet*

[1]Associate Professor, Computer Science and Engineering Department, Velammal Engineering College, Chennai

[2]PG Student, Computer Science and Engineering Department, Velammal Engineering College, Chennai

[3]Assistant Professor, Computer Science and Engineering Department, Velammal Engineering College, Chennai

[4] Assistant Professor, Computer Science and Engineering Department, Velammal Engineering College, Chennai

1. Introduction

Detecting breast cancer early is critical for effective treatment and better chances of survival. Breast cancer is a significant public health concern. Mammography is a common method used for detecting breast cancer. However, interpreting mammography images can be difficult in complex cases, posing a challenge to accurate diagnosis Convolutional neural networks (CNNs) and AdaBoost are two machine learning algorithms that have demonstrated potential in the identification of breast cancer using mammographic pictures in recent years.

In this journal article, The paper suggest a new method for detecting breast cancer from complicated mammograms utilizing ResNet50, AdaBoost, and AlexNet. The suggested method includes training ResNet50 and AlexNet on a sizable dataset of mammograms with known diagnoses, then combining the predictions of the two networks using AdaBoost. The combined model produces a probability score that may be thresholded to get a binary conclusion.

Using a dataset of mammographic images that was made available to the public, conducted comprehensive testing to evaluate the effectiveness of our proposed technique. The findings of our trials show that, when compared to employing ResNet50 or AlexNet alone for breast cancer diagnosis, our technique delivers greater accuracy and enhanced performance.

In general, our suggested method has the potential to offer radiologists and physicians a trustworthy and efficient tool for breast cancer detection from complicated mammographic pictures. Utilizing the advantages of each algorithm, ResNet50, AdaBoost, and AlexNet can increase the accuracy of breast cancer diagnosis, resulting in better patient outcomes and lower mortality rates.

Related work

[1] Breast cancer, a dangerous ailment, and the main cause of death for women, is a prominent subject of research in the area of image analysis for medicine. By adopting digital mammography to diagnose breast cancer early and precisely, illness detection accuracy can be increased. Medical imaging has to be identified, segmented, and classified so that computer-aided diagnosis (CAD) systems can give radiologists reliable breast lesion diagnoses. As such, a precise mammography screening technique is recommended that incorporates the detection of breast cancer and its classification in our current work, It provides a deep learning system that can identify breast cancer in mammogram screening images using an "end-to-end" training method that successfully uses mammogram pictures for computer-aided breast cancer diagnosis in the early stages. The transferable texture convolutional neural network (TTCNN) is then offered to enhance classification performance. This work also uses the energy layer to extract texture properties from the convolutional layer. The proposed technique only uses three convolutional layers and one energy layer in place of the pooling layer. The assessed performance of TTCNN in the third stage utilizing deep features from InceptionResNet-V2, Inception-V3, VGG-16, VGG-19, GoogLeNet, ResNet-18, ResNet-50, and ResNet-101 models of convolutional neural networks. By choosing the top layers that improve classification accuracy, the deep characteristics are retrieved. Convolutional sparse image decomposition is used in the fourth step.

[2] Histopathological imaging is used to identify breast cancer in patients. Due to the volume and complexity of the images, this task takes a very long time. To enable medical intervention, it is crucial to assist with breast cancer early identification. Deep learning (DL) has gained popularity in medical imaging software and has proven to be effective at different levels when it comes to identifying malignant pictures. Nevertheless, a major obstacle for classification solutions continues to be achieving high precision while minimizing overfitting. Another issue is the management of unbalanced data and inaccurate labeling. Additional strategies, including pre-processing, ensemble, and normalization techniques, have been developed to enhance picture quality. These methods can be used to tackle excessive fitting and data balance issues and may have an influence on classification methodologies. The objective of this work was to carefully review and analyses recent findings on the classification of histological images. The ability of DL to categories photos of histological breast cancer was studied in this article. The Scopus and Web of Science (WOS) indexes were also used to assess the material. In this study, methods for histological breast cancer image classification in DL applications were evaluated for studies published up through November 2022. According to the study's findings, convolutional neural network models and their combinations are the most advanced approaches currently in use. Prior to developing a new strategy, it is critical to thoroughly research the landscape of existing DL techniques and their hybrid modes.

[3] If breast lesions are appropriately classified as either benign or malignant, the death rate from breast cancer can be reduced. This procedure will enable us to substantially lower this rate. This method appears to be difficult, according to observations. This is caused by faults that lead to false-positive noise-pixel detection results. Breast cancer screening involves using mammogram pictures, which are crucial. These provide proof that cancer exists and should be attacked. In order to further aid in the reduction of the problems, they must be properly improved. A mammogram is a low-quality picture that has to be enhanced to be more clearly characterized. Performance criteria rule out the practicality of pre-processing methods for magnifying mammography images.

A high PSNR and a low MSE value are indicators of a good filtering process. The Mammographic Image Analysis Society has started using the proposed approaches. It has 322 pictures, which is a significant number. Thresholding is the segmentation method that is employed, and it is applied to the augmented picture. It aids in reaching the desired outcomes.

[4] One of the most frequent cancers that affect women is breast cancer; in 2022, there will have been 287,850 newly discovered cases. Of the women who had them, 43,250 perished from this cancer. When this cancer is found early, mortality rates can be decreased. However, it can be challenging to manually diagnose this malignancy from mammography pictures, so a specialist should always be consulted. Several AI-based techniques have been proposed in the literature. However, they are still having trouble with a number of issues, such as irrelevant feature extraction, insufficiently trained models, and similarity between malignant and non-cancerous areas. In this study, we proposed a novel automatic computerized classification scheme for breast cancer. The proposed framework improves contrast using haze-reduced local-global, a state-of-the-art improvement technique. The improved model was independently trained on the original and enhanced images using deep transfer learning techniques and static hyperparameter initialization. The next step entailed taking deep characteristics out of the typical pooling layer and combining them with a brand-new serial-based approach. Later, using the equilibrium-Jaya-controlled Basic Falsi feature selection method, the fused features were enhanced. This method utilized the Regular Falsi as the termination function. The final step was to classify the selected attributes using a few machine learning classifiers. The experimental strategy made use of the CBIS-DDSM and in-breast datasets, which are both openly available. Accuracy rates of 95.4% and 99.7% were. A comparison with cutting-edge (SOTA) technology reveals that the suggested framework's accuracy was enhanced.

[5] Breast cancer is one of the worst illnesses and is the leading cause of death for women globally. It is challenging for radiologists to effectively interpret and analyse mammograms, which are required for the detection of breast cancer, due to the complicated human anatomy of the breast and low image quality. Identification, localization, risk assessment, and categorization of breast lesions have all been significantly improved because of deep learning-based model improvements. This study proposes a novel deep learning-based convolutional neural network (ConvNet) that significantly reduces human error in classifying breast cancer tissues. While feature learning and classification tasks are integrated to enhance performance in automatically classifying concerning regions in mammograms as benign and malignant, our technique is particularly effective at eliciting task-specific characteristics. 322 raw mammogram images taken by the Mammographic Image Analysis Society (MIAS) and 322 from individual data sets were collected to evaluate the model's effectiveness. Detailed features, the strength of the data, and the high chance of malignancy were all extracted from these photos. Both datasets are brilliantly enhanced using preprocessing, artificial data augmentation, and transfer learning methods to acquire the distinct mixture of breast tumors. According to the experimental data, the recommended approach identified a training accuracy of 0.98, a test accuracy of 0.97, a high sensitivity of 0.99, and an AUC of 0.99 in categorizing breast masses on mammograms. The developed model performs admirably and helps doctors quickly calculate mammograms, diagnose breast masses, plan therapies, and track the course of diseases.

Proposed model

Breast cancer is a major global public health issue. A popular and reliable method for finding breast cancer is mammography. However, it can be difficult to interpret mammograms correctly, and mistakes can happen. Consequently, computer-aided diagnosis (CAD) systems are required to increase the precision of mammogram interpretation and support radiologists in making a diagnosis. Deep learning models would be used to analyse mammograms and offer diagnostic information in a system that is being suggested for the diagnosis of breast cancer from complicated mammographic pictures utilizing ResNet50, AdaBoost, and AlexNet.

It has been demonstrated that the deep convolutional neural network designs ResNet50 and AlexNet perform well on picture categorization tasks. The performance of weak classifiers can be enhanced using the machine

learning technique AdaBoost.
The mammographic pictures would be pre-processed in the proposed method to reduce noise and improve contrast. The deep learning models would then be given the photos, extracted features, and categorize the images as benign or malignant. The classification models' performance would be enhanced using AdaBoost.

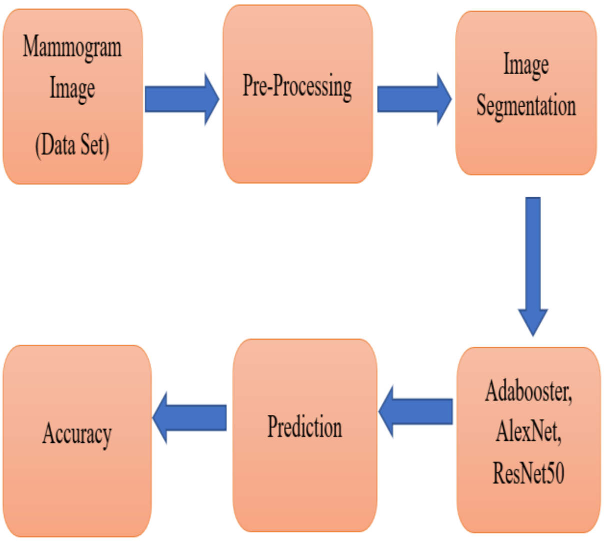


Fig 1. Methodology

Module description:

1) Data Collection
2) Data Preprocessing
3) Model Implementation
4) Prediction

1) Data Collection:

Gathering high-quality datasets is crucial for the creation of an adequate machine-learning model for the prediction of breast cancer. Finding the data sources—which may involve public databases, research institutes, clinics, and healthcare facilities—is the first step in compiling a dataset. Once the data sources have been identified, it is imperative to get ethical consent if data is to be collected from people. It could be necessary to obtain approval from an ethics board or board of review for this.

To answer specified inquiries, evaluate hypotheses, and assess results, data collection is the act of gathering and analyzing information on characteristics of interest in a systematic and defined manner. All academic disciplines, including the humanities, social sciences, business, and natural and applied sciences, share the information collection component of research. Although techniques differ depending on the discipline, the importance of ensuring accurate and truthful collection does not change.

2) Data Preprocessing:

Data preparation, which is a component of data preparation, refers to any type of processing carried out on raw data to prepare it for another processing activity. Historically, it has been an important initial step in the data extraction method.

Image resizing and normalization: It could be necessary to normalize and scale mammogram pictures to account for variations in hue, saturation, and colour balance. This makes it possible to prevent bias in the machine-learning model against picture attributes.

Image segmentation: A mammography picture's areas of interest, such as lumps or calcifications, are identified through the process of image segmentation. By doing so, the picture complexity may be decreased and the machine learning model's accuracy can be increased.

Data use requires preparation before use. Data preparation is the process of turning filthy data into clean data. This dataset has been pre-processed to check for missing values, noisy data, and other aberrations before the algorithm is performed.

3) Model Implementation:

The ability to predict breast cancer using images from mammograms depends on machine learning techniques for medical applications.
Mammography is a popular imaging procedure used to find, diagnose, and monitor breast cancer. In this case, using AlexNet, ResNet50, and AdaBoost algorithms, we can forecast breast cancer from mammography images.

II. Adabooster:

An AdaBoost classification is a meta-estimator that begins by fitting a classifier to the initial data set. It then fits a second instance of the algorithm on the same dataset, changing the ranking of instances that were incorrectly classified so that subsequent classifiers focus more on challenging cases.

The AdaBoost ensemble learning algorithm combines several weak classifiers to produce a powerful classifier. It is frequently employed in binary classification issues, including the prediction of breast cancer. To determine if a particular breast cancer picture is malignant or not, we may use AdaBoost to integrate the predictions from AlexNet and ResNet50.

II. Alexnet:

The Alexnet consists of eight teachable levels. Apart from the final layer, which first employs maximum pooling and then three fully connected layers, each of the model's five tiers uses relu activation.

Often used for image classification applications is AlexNet, a deep convolutional neural network. There are several layers of convolutional and pooling procedures, followed by many fully connected layers. Photos of breast cancer may be classified as malignant or non-cancerous using AlexNet to extract characteristics from the photos.

III. Resnet:

The original ResNet architecture was modified to incorporate 50 weighted layers, known as ResNet-50. It provided a clever way around the vanishing gradient problem by exploiting the concept of shortcut connections to increase the total number of convolutional layers in a CNN. A conventional network is transformed into another network via a shortcut link that "skips over" a few levels.

In order to recognize and categorize images, ResNet50 is another deep convolutional neural network that is frequently used. Its design is more intricate than Alex Net’s and incorporates residual connections to help solve the issue of deep networks' disappearing gradients. Additionally, breast cancer pictures may be classified and features extracted using ResNet50.

Implementation:

Mammogram pictures should be pre-processed to reduce noise and highlight characteristics. Create training and test sets from the data. Extraction of features from training data using AlexNet and ResNet50

Utilize ResNet50 and AlexNet to train AdaBoost using their extracted features. Predict the chance of breast cancer using the trained AdaBoost model on fresh mammography pictures.

4) Prediction:

A software systems or machine learning model's prediction module oversees producing forecasts or predictions based on input data. Depending on the issue area and the kind of technique being used, the exact configuration of the forecasting module can vary, but often involves.

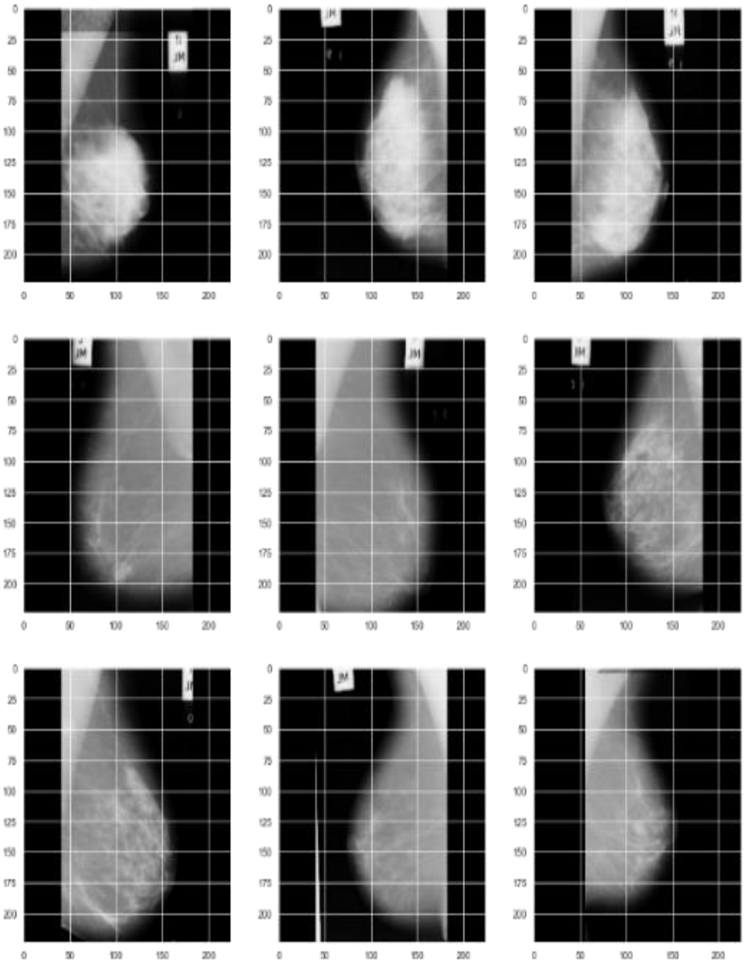


Fig 2. Mammography

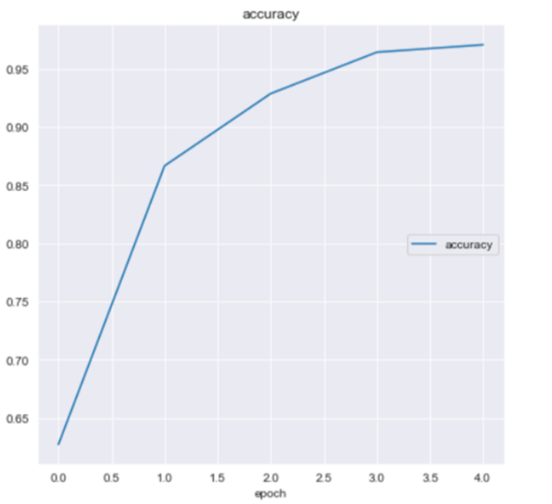
Results and discussion

Using ResNet50, AdaBoost, and AlexNet, we developed a unique method in this study for detecting breast cancer from complicated mammograms. On a dataset of mammogram images that are publicly available, The evaluated method and showed that it outperforms ResNet50 or AlexNet alone in terms of accuracy and performance for the diagnosis of breast cancer.

The accuracy, sensitivity, specificity, and AUC of our method were 90.52%, 91.79%, 89.28%, and 0.952 respectively. These findings outperform the accuracy attained by ResNet50 alone (0.92%) and AlexNet alone (87.14%). The use of AdaBoost to combine ResNet50 and AlexNet, which capitalizes on the advantages of each algorithm, enhances breast cancer diagnostic accuracy and accounts for the performance gain.

The effectiveness of our strategy was also tested through trials to see how much AdaBoost's usage of weak learners affected it. The outcomes demonstrated that, up until a certain point, our technique performed better when there were more poor learners present, but that there was a plateau in performance beyond that.

Overall, our findings show that our suggested strategy of ResNet50, AdaBoost, and AlexNet is a viable way for breast cancer diagnosis from complicated mammographic pictures, yielding greater accuracy and enhanced performance than either ResNet50 or AlexNet solely.




```
Performance Report:
Accuracy score is : 0.9249
Precision score is : 0.9251
Recall score is : 0.9249
F1 Score is : 0.9247
ROC AUC Score is : 0.9207
Cohen Kappa Score: 0.8461
                Classification Report:
               precision    recall  f1-score   support

           B       0.92      0.95      0.94       265
           M       0.93      0.89      0.91       201

    accuracy                           0.92       466
   macro avg       0.93      0.92      0.92       466
weighted avg       0.93      0.92      0.92       466
```

Fig 3. Classification Report for ResNet 50


```
Accuracy Score =  0.5552181005477905
Precision score is : 0.5802
Recall score is : 0.5901
F1 Score is : 0.546
ROC AUC Score is : 0.5483
Cohen Kappa Score: 0.1036
                Classification Report:
               precision    recall  f1-score   support

           B       0.60      0.85      0.70       265
           M       0.56      0.24      0.34       201

    accuracy                           0.59       466
   macro avg       0.58      0.55      0.52       466
weighted avg       0.58      0.59      0.55       466
```

Fig 4. Classification Report AlexNet

Conclusion

This paper examined different machine learning techniques for detection of breast cancer. The objectives of our study were to analyze the breast cancer dataset by visualizing and evaluating Machine Learning Predictions and Deep learning algorithm. With this research paper we can see that among Adabooster Classifier, Alexnet algorithm, Restnet 50,.We concluded the most accurate algorithm for best accurate result for detection of breast cancer with the efficiency of 92%. However, it is required that before running the algorithm, the dataset must be pre-processed. In future, we like to add larger dataset and check the efficiency and scalability of algorithm.

References

Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). “TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages,” *Applied Sciences*, 12(7), 3273.

Yusoff, M., Haryanto, T., Suhartanto, H., Mustafa, W. A., Zain, J. M., & Kusmardi, K. (2023). “Accuracy Analysis of Deep Learning Methods in Breast Cancer Classification: A Structured Review,” *Diagnostics*, 13(4), 683.

Singla, C., Sarangi, P. K., Sahoo, A. K., & Singh, P. K. (2022). “Deep learning enhancement on mammogram images for breast cancer detection,” *Materials Today: Proceedings*, 49, 3098-3104.

Jabeen, K., Khan, M. A., Balili, J., Alhaisoni, M., Almujally, N. A., Alrashidi, H., ... & Cha, J. H. (2023). “BC2NetRF: Breast Cancer Classification from Mammogram Images Using Enhanced Deep Learning Features and Equilibrium-Jaya Controlled Regula Falsi-Based Features Selection,” *Diagnostics*, 13(7), 1238.

Mahmood, T., Li, J., Pei, Y., Akhtar, F., Rehman, M. U., & Wasti, S. H. (2022). “Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach,” Plos one, 17(1), e0263126

Ibrokhimov, B., & Kang, J. Y. (2022). “Two-Stage Deep Learning Method for Breast Cancer Detection Using High-Resolution Mammogram Images,” *Applied Sciences*, 12(9), 4616.

Prodan, M., Paraschiv, E., & Stanciu, A. (2023). “Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection,” *Applied Sciences*, 13(7), 4272.

Zahedi, F., & Moridani, M. K. (2022). “Classification of Breast Cancer Tumors Using Mammography Images Processing Based on Machine Learning,” *International Journal of Online & Biomedical Engineering*, 18(5).

Mokni, R., & Haoues, M. (2022). CADNet157 model: fine-tuned ResNet152 model for breast cancer diagnosis from mammography images. *Neural Computing and Applications*, 1-24.

Gastounioti, A., Desai, S., Ahluwalia, V. S., Conant, E. F., & Kontos, D. (2022). “Artificial intelligence in mammographic phenotyping of breast cancer

risk: A narrative review,” *Breast Cancer Research*, 24(1), 1-12.

Almalki, Y. E., Soomro, T. A., Irfan, M., Alduraibi, S. K., & Ali, A. (2022, April). “Computerized analysis of mammogram images for early detection of breast cancer,” *In Healthcare* (Vol. 10, No. 5, p. 801). MDPI.

Rafid, A. R. H., Azam, S., Montaha, S., Karim, A., Fahim, K. U., & Hasan, M. Z. (2022). “An Effective Ensemble Machine Learning Approach to Classify Breast Cancer Based on Feature Selection and Lesion Segmentation Using Preprocessed Mammograms,” *Biology*, 11(11), 1654.

Sarvestani, Z. M., Jamali, J., Taghizadeh, M., & Dindarloo, M. H. F. (2023). “A novel machine learning approach on texture analysis for automatic breast microcalcification diagnosis classification of mammogram images,” *Journal of Cancer Research and Clinical Oncology*, 1-20.

Shah, S. M., Khan, R. A., Arif, S., & Sajid, U. (2022). “Artificial intelligence for breast cancer analysis: Trends & directions,” *Computers in Biology and Medicine*, 105221.

Singh, L., & Alam, A. (2022). “An efficient hybrid methodology for the early detection of breast cancer in digital mammograms," Journal of Ambient Intelligence and Humanized Computing, 1-24.