ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5,May 2024

Early Detection of Breast Cancer Using Machine Learning Algorithm And a Customized AlexNet

Rajya Lakshmi Jonnalagadda[1], Sabeena Patan[2], Afroz Syed[3], Meghana Peddi[4], Lokesh Gosala[5]
[1] Assistant Professor, [2, 3, 4, 5] Undergraduate. Department Of ECE,
PSCMR College Of Engineering And Technology, Vijayawada, Andhra Pradesh, India
rajimtech12@gmail.com[1], patansabeena07@gmail.com[2], afrozsyed568@gmail.com[3],
meghanapeddi123@gmail.com[4],gosalalokesh11@gmail.com[5]

Abstract—-Breast cancer, the deadliest and most destructive type of the disease, is frequently discovered for the first time when a lump develops in the breast. The reason behind this is a normal cell proliferation in the mammary glands. Early diagnosis increases the prognosis for breast cancer. Therefore early detection of breast cancer is crucial for preventing serious complications. Ultrasound, Computed tomography, mammography are three major imaging techniques used for breast cancer screening and early detection. Digital image processing techniques such as image pre-processing, image segmentation, feature extraction and image classification are applied in this project on the digital mammogram images to achieve early and automated detection of breast cancer. On mammograms, convolutional neural networks (CNNs) may identify breast cancer. CNNs may be trained on huge datasets to enhance accuracy and deal with more complicated visual information than standard approaches. This study utilizes AlexNet, a convolutional neural network with eight layers: five convolutional layers and two max-pooling layers. Furthermore, we used support vector machines (SVM) for categorization. We conducted this experiment using the widely famous Django benchmark dataset. It is shown that the proposed model achieves accuracy of 98.71%, Precision of 99.01%, Recall of 98.32%, FScore of 98.66%. Overall, this project offers a comprehensive solution to early breast cancer detection using machine learning techniques to improve healthcare.

Index Terms—Breast Cancer, Mammograms, Django dataset, CNN, AlexNet, Support vector machine.

I. Introduction

Breast cancer claims the lives of most women in developed countries. The most effective tool in the fight against breast cancer is early detection. According to WHO (World Health Organization), breast cancer accounted for maximum deaths (2.26 million cases), worldwide in 2020 out of the 10 million cases of cancer. Breast cancer starts when cells in the breast begin to grow out of control. These accumulations of cells are called tumours and they can often be seen on an x-ray or felt as a lump. Breast cancer can spread when the cancer cells get in to the blood or lymph system and are carried to other parts of the body making them prone to cancer. Before invasive surgery, physicians need to be able to distinguish between benign and malignant breast tumors with high accuracy and reliability [1]. Classifying patients into two groups based on benign and malignant illnesses is the aim of these projections. Only a biopsy can definitively determine if there is cancer in the breast; doctors, self-examination, and imaging tools cannot. Nowadays, the primary method for the early detection of breast cancer is the use of imaging technologies, such as computed tomography, magnetic resonance imaging, ultrasound, and mammography. The breast is imaged using X-rays in this procedure. Breast cancer treatment choices and survival rates are improved when the disease is discovered early by screening and diagnostic tomography. According to research, radiologists make mistakes in the diagnosis of cancer during screening studies at a rate of 10% to 30% . 52% of mistakes are the result of incorrect interpretation of breast cancer symptoms, and 43% of errors are the result of failing to notice symptoms in abnormal scans
[2]. One of the main causes of these errors is radiologist's dependence on visual inspection. While manually screening a large number of mammograms, radiologists may soon become tired and miss crucial signs while interpreting the pictures. To mitigate these effects, significant efforts are being made to automate the mammography screening procedure. One popular machine learning technique that has shown great success across numerous domains is deep learning. NLP, robotics, agriculture, medicine, and other industries are among them. A potent Support Vector Machine (SVM) - based Deep Learn-ing Architecture (DLA) for breast cancer diagnosis. CNN is a popular and successful deep learning technique. Trained convolutional neural network (CNN) models are frequently used in data science applications, such as the classification of medical pictures and the feature extraction from training photos. Stated differently, CNN's purpose is to identify and extract high-level features that allow it to discriminate between several class labels in a classification task [5]. The research presents a novel model that combines SVM with the Modified AlexNet architecture for the aim of diagnosing breast cancer. Deep learning-based feature extraction and the SVM classifier can work together to diagnose breast cancer more rapidly and reliably. This study looks into ways to change the original AlexNet architecture in order to recognize breast cancer in mammograms. The Modified AlexNet is adjusted and tailored to identify relevant and note.

II. Literature Survey

A. Related works on Breast cancer

The following research papers have been evaluated for finalizing the objectives of our project work. The research papers discussed in this collection are mostly relevant to our project Early detection of Breast Cancer Using Machine Learning Algorithm And a Customized AlexNet

Sadhukhan Subham, Nityasree Upadhyay, and Prerana Chakraborty proposed a comparative study of two different algorithms KNN and SVM is conducted where the accuracy of each classifier is measured. After this, they analyze a digital image of a fine needle aspirate (FNA) of breast tissue using image processing to find out the features of nuclei of the cells. And then apply the feature values to our trained model to find whether the tumor developed is benign or malignant [1].

Ganesan, Karthikeyan, proposed a review aims at providing an overview about recent advances and developments in the field of Computer-Aided Diagnosis (CAD) of breast cancer using mammograms, specifically focusing on the mathematical aspects of the same, aiming to act as a mathematical primer for intermediates and experts in the field [2].

Escorcia-Gutierrez, Jose, et al proposed ML techniques, particularly the wavelet neural network (WNN) is used for the classification of digital mammograms for the detection of breast cancer. The ADL-BCD method is evaluated using a benchmark dataset and the results are analyzed under several performance measures. The simulation outcome indicated that the ADL-BCD model outperforms the state of art methods in terms of different measures [3].

Houssein, Essam H., et al suggested this review reflects on the classification of breast cancer utilizing multi- modalities medical imaging. It first provides an overview of the different approaches to machine learning, then an overview of the different deep learning techniques and specific architectures for the detection and classification of breast cancer. Finally, this review summarizes the future trends and challenges in the classification and detection of breast cancer [4].

Dina A. Ragab, maha sharkas, Jinchang Ren proposed model, the first approach involves determining the region of interest (ROI) manually, while the second approach uses the technique of threshold and region based. The deep convolutional neural network (DCNN) is used for feature extraction. A well-known DCNN architecture named AlexNet is used and is fine-tuned to classify two classes instead of 1,000 classes. [5].

jayandhi G, S.mary joans proposed an automated computerized system with high accuracy is needed. In this study, an efficient Deep Learning Architecture (DLA) with a Support Vector Machine (SVM) is designed for breast cancer diagnosis. It combines the ideas from DLA with SVM. The state-of-the-art Visual Geometric Group (VGG) architecture with 16 layers is employed in this study as it uses the small size of 3× 3 convolution filters that reduces system complexity [6].

YJ Tan, Sim Kok Swee, Sicily Ting Fung Fung proposed Breast cancer detection using convolutional neural networks (BCDCNN) is aimed to speed up the diagnosis process by assisting specialist to diagnosis and classification the breast cancer. A series of mammogram images are used to convert a human visual image into a computer visual image. After that, all changed images are assigned into CNN classifier as training source. The CNN classifier will then produce a model to recognize the mammogram image. The proposed method has higher accuracy than other existing models [7].

MANJUNATHAN N, N.Gomathi proposed an application of ML approaches in breast cancer diagnosis has resulted in higher accuracy and specificity, making it a suitable tool for assisting clinicians in clinical decision-making. Overall, the findings indicate that more study may be required to address the problems and limitations of ML in breast cancer diagnosis, as well as to create standardized processes for data collection and analysis [8].

S Anklesaria, U Maheshwari, R Lele, P Verma proposed a model aimed to incorporate several machine learning (ML) algorithms, including Support Vector Machine (SVM), Logistic Regression, k-Nearest Neighbour (KNN), Decision Tree (DT), Importance Method for feature selection. The research concluded that the Support Vector Machine Algorithm proved to be the most effective model which fit our dataset with an Accuracy of 95.8% followed by KNN with an accuracy of 95.3% [9].

III. METHODOLOGY

As can be seen as that the proposed model consists of training and a validation stage. The Django dataset, which consists of Mammograms from women who have been diagnosed with breast cancer, is the first source used. Firstly the image pre-processing of the mammogram is carried out which helps in removing noise in the image, if any. In this first step of data preparation, the input photos (1560) are divided into three categories benign, malignant, and normal and they are processed to enhance the quality of the data before the modeling stage. This stage involves filtering outliers, removing noise, and transforming the input images. Furthermore, there are three possible sizes for the input patch: 16 by 16, 32 by 32, or 48 by 48. Following cleaning and organization, the data is divided into two sets: the training set and the validation set. Twenty percent of the cleaned and processed data is reserved for other purposes, and eighty percent is used for training. The training data is comprised of two distinct components, namely the application layer and the performance layer. As advised, features have been retrieved at the application layer by using the Breast-SVM modified convolutional neural network. Important information is retrieved from the input photos during the feature extraction process, and it is then sent to the next phase. A fundamental component of a convolutional neural network (CNN), the con-volution layer is essential for identifying prominent features in the input data. The Breast-SVM, CNN network architecture is shown in Fig. 1 for the purpose of identifying breast cancer. After training of Breast-SVM the validation process begins, the trained Breast-SVM is used to access 20% of the validation data set. The trained model exhibits a benign classification when cancerous cells are lightly detected in breast cancer, where as a malignant classification is displayed when cancer

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5,May 2024

cells are detected and normal displayed when no cancer cells are detected. During testing and evaluation, the hybrid model

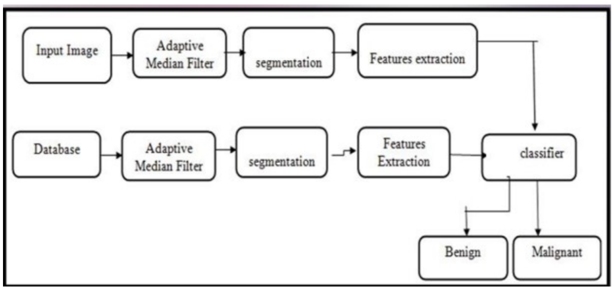


Fig. 1. Block Diagram of suggested Breast-SVM Model

demonstrates superior performance metrics such as accuracy, sensitivity, specificity, and area under the ROC curve. This indicates its effectiveness in accurately distinguishing between malignant and benign cases, which is crucial for timely and accurate diagnosis in clinical settings.

A. Image Pre-Processing

The primary purpose of pre-processing is to improve picture quality and prepare it for subsequent processing by eliminating or minimizing irrelevant and unnecessary portions in the back-ground of mammography images.This process involves the following stages where,input for this stage is a mammogram image which was converted into a gray scale image. Adaptive median filter is applied to the image to remove the noise from the image. Repeat the above steps for a dataset that consists of several mammogram images. Output for this stage is also an image without noise.

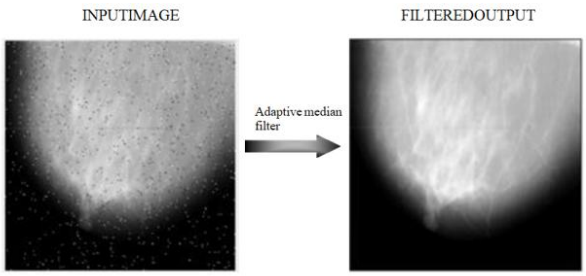


Fig. 2. Pre-Processing of Breast Cancer Tissue

B. Segmentation

The goal of segmentation is to simplify and/or change the representation of an image into something that is more mean-ingful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same

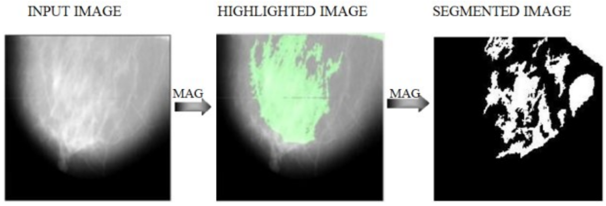


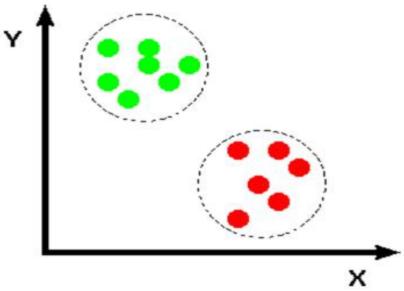
Fig. 3. Segmentation of tissue

C. Feature Extraction

Feature extraction is a part of the dimensionality reduction process, in which, an initial set of the raw data is divided and reduced to more manageable groups. So when you want to process it will be easier. The most important characteristic of these large datasets is that they have a large number of variables. These variables require a lot of computing resources to process them. So Feature extraction helps to get the best fea-ture from those big data sets by select and combine variables into features, thus, effectively reducing the amount of data. These features are easy to process, but still able to describe the actual dataset with the accuracy and originality.Some of the features we extracted from an image are Mean, Variance, Standard deviation, Skewness, Kurtosis, Entropy, Correlation, Inverse difference moment(IDM).

D. Classification

Image classification analyzes the numerical properties of various image features and organizes data into categories. In the subsequent testing phase, these feature-space partitions are used to classify image features. The description of training classes is an extremely important component of the classifica-tion process. The objective of image classification is to identify and portray, as a unique gray level (or color), the features occurring in an image in terms of the object or type of land cover these features actually represent on the ground. Image classification is perhaps the most important part of digital image analysis.



IV. Results and Discussions

In this study context, the suggested Breast-SVM model undergoes training using 874 from benign category. Among 874 samples, the model correctly predicted 870 samples, with 4 samples was misclassified. In the case of malignant samples, the model was trained using a dataset of 420 samples. The model accurately predicted 413 samples while 7 samples are inaccurately classified. For validation process, a sample size of 310 was used to access the proposed model which is divided into three categories: benign, malignant and normal. The model exhibits a high accuracy in predicting benign, malignant and normal cases. The model achieves Accuracy of 98.71%, Precision of 99.01%, Recall of 98.32%, and Fscore of 98.66%. A web-based application has been developed to assist patients

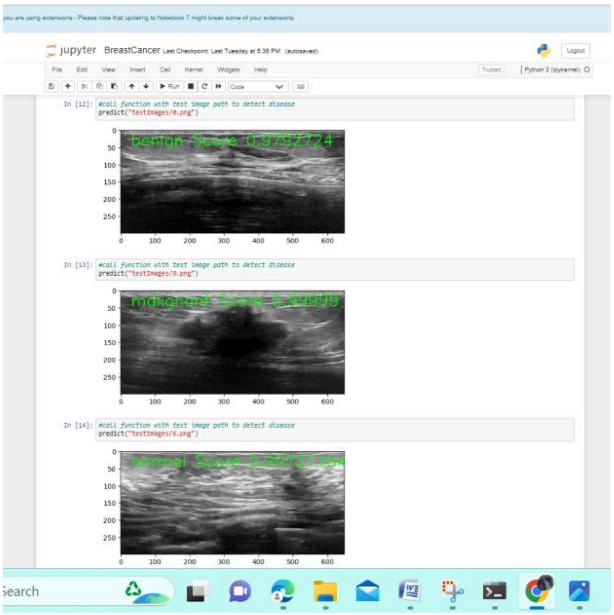


Fig. 5. Breast cancer detected output

in evaluating breast abnormalities detected in mammogram images. The application employs a machine learning algorithm trained to classify these abnormalities as benign, malignant, or normal. Patients can easily upload their mammogram images through the web interface and receive an immediate prediction regarding the nature of the detected abnormality.

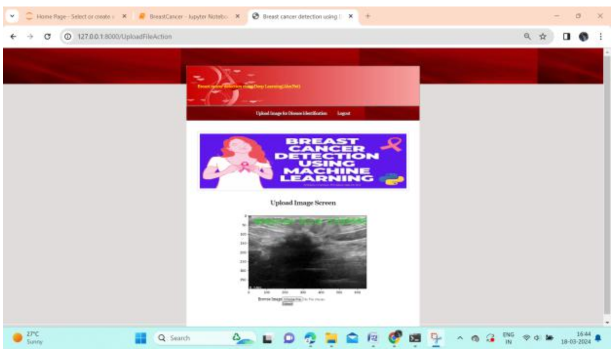


Fig. 6. User Interface

V. Conclusion

In conclusion, this project successfully detected breast can-cer on different test images. The proposed system represents a significant advancement in breast cancer detection lever-aging machine learning algorithms specifically a customized AlexNet architecture. Early detection of the tumor is a vital process that benefits the diagnosis of Breast cancer. This project is achieved by using four image processing techniques, namely image pre-processing, image segmentation, feature extraction, and classification. The final classification process is done using the ensemble learning method and it determines whether the tumor is normal, benign, or malignant with an output image of the segmented part of the breast. Through, meticulous data pre-processing, feature extraction, and model training, we have demonstrated the efficacy of our approach in accurately classifying breast cancer cases from mammography images. Our results showcase the potential of machine learning in revolutionizing early cancer detection, a critical aspect in improving patient outcomes and reducing mortality rates associated with breast cancer. The utilization of deep learning techniques, such as the customized AlexNet, not only enhances classification accuracy but also streamlines the process by automating the analysis of medical imaging data. Hence, the project helps in detecting the cancerous tumor before its spreads to other parts of the body and increases the chances of successful diagnosis. In this project, we aim to expand our scope for future by incorporation of additional data sources, advanced deep learning architectures, and interdisciplinary collaborations with healthcare professionals. Together, these efforts will contribute to the ongoing advancement of ma-chine learning in healthcare, ultimately leading to more CNN, SVM ,ALEX-NET Data pre-processing, Splitting input image data,Threshold Resize,Normalization,Training data Model im-plementation Testing data, Model Evaluation, Accuracy, pre-cision, Recall, F-Score effective and personalized diagnostics strategies for breast cancer and beyond.

References

[1] Sadhukhan, Subham, Nityasree Upadhyay, and Prerana Chakraborty. ”Breast cancer diagnosis using image processing and machine learning.” Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018. Springer Singapore, 2020.

[2] Ganesan, Karthikeyan, et al. ”Computer-aided breast cancer detection using mammograms: a review.” IEEE Reviews in biomedical engineering 6 (2012): 77-98.

[3] Escorcia-Gutierrez, Jose,´ et al. ”Automated deep learning empowered breast cancer diagnosis using biomedical mammogram images.” Com-puters, Materials and Continua 71 (2022): 3- 4221.

[4] Houssein, Essam H., et al. ”Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review.” Expert Systems with Applications 167 (2021): 114161.

[5] Ragab, Dina A., et al. ”Breast cancer detection using deep convolutional neural networks and support vector machines.” PeerJ 7 (2019): e6201.

[6] Jayandhi, G., J. S. Jasmine, and S. Mary Joans. ”Mammogram Learn-ing System for Breast Cancer Diagnosis Using Deep Learning SVM.” Computer Systems Science Engineering 40.2 (2022).

[7] Tan, Y. J., K. S. Sim, and Fung Fung Ting. ”Breast cancer detection using convolutional neural networks for mammogram imaging system.” 2017 International Conference on Robotics, Automation and Sciences (ICORAS). IEEE, 2017.

[8] Manjunathan, N., N. Gomathi, and S. Muthulingam. ”Early Detection of Breast Cancer using Machine Learning.” 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS). IEEE, 2023.

ISSN (ONLINE): **2454-9762**
ISSN (PRINT): **2454-9762**
Available online at www.ijarmate.com

IJARMATE

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5,May 2024

[9] Sharma, Harsh, Pooja Singh, and Ayush Bhardwaj. ”Breast cancer detec-tion: comparative analysis of machine learning classification techniques.” 2022 International Conference on Emerging Smart Computing and Infor-matics (ESCI). IEEE, 2022.
[10] Anklesaria, Saniya, et al. ”Breast cancer prediction using optimized machine learning classifiers and data balancing techniques.” 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA. IEEE, 2022.
[11] Antor, Jahidul Hasan, Rajneesh Kumar, and Pooja Rani. ”Breast Cancer Prediction Using Machine Learning Classifiers.” 2022 5th International Conference on Advances in Science and Technology (ICAST). IEEE, 2022.
[12] Tavakoli, Nasrin, et al. ”Detection of abnormalities in mammograms using deep features.” Journal of Ambient Intelligence and Humanized Computing (2023): 1-13.
[13] Moon, Woo Kyung, et al. ”Computer-aided diagnosis of breast ul-trasound images using ensemble learning from convolutional neural networks.” Computer methods and programs in biomedicine 190 (2020): 105361.
[14] Rovshenov, Atajan, and Serhat Peker. ”Performance Comparison of Different Machine Learning Techniques for Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset.” 2022 3rd International Informatics and Software Engineering Conference (IISEC). IEEE 2022.
[15] Rovshenov, Atajan, and Serhat Peker. ”Performance Comparison of Different Machine Learning Techniques for Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset.” 2022 3rd International Informatics and Software Engineering Conference (IISEC). IEEE, 2022.
[16] Gupta, Madhuri, and Bharat Gupta. ”A comparative study of breast can-cer diagnosis using supervised machine learning techniques.” 2018 second international conference on computing methodologies and communication (ICCMC). IEEE, 2018.
[17] Tewari, Yajush, Eshant Ujjwal, and Lalit Kumar. ”Breast Cancer Classi-fication Using Machine Learning.” 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2022.
[18] Anklesaria, Saniya, et al. ”Breast cancer prediction using optimized machine learning classifiers and data balancing techniques.” 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA. IEEE, 2022.
[19] Ara, Sharmin, Annesha Das, and Ashim Dey. ”Malignant and benign breast cancer classification using machine learning algorithms.” 2021 International Conference on Artificial Intelligence (ICAI). IEEE, 2021.
[20] Khan, Razib Hayat, et al. ”A comparative study of machine learning algorithms for detecting breast cancer.” 2023 IEEE 13th Annual Com-puting and Communication Workshop and Conference (CCWC). IEEE, 2023.
[21] Abdollah Jafari Chashmi and Mehdi Chehelamirani “Using Adaptive Median Filter for Noise Removal from Image to Diagnose Breast Cancer”, Merit Research Journal of Engineering, Pure and Applied Sciences Vol. 5(1) pp. 014-018, August-2019.
[22] Osama R.Shahin , Hamdy M. Kelash , Gamal Attiya , “Breast cancer de-tection based on dynamic template matching”, Vol 20, No. 12, Dec2013.