Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 9, Issue 1, 2022
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Research JoAIRA

Breast Cancer Detection Using Deep Learning

Nilima Bhosale[1,*], Jagruti Bhagat[1], Shivani Bhoir[1], Rajashree Gadhave[2]

Abstract

Breast cancer, according to the Breast Cancer Institute, is one of the most difficult kinds of cancer that is extremely threatfulfor women all over the world. Early cancer discovery, followed by appropriate treatment, can help minimize the risk of deaths in the first stage of respective cancer only and increase the likelihood of survival by up to 8%. Medical imaging research has become increasingly reliant on machine learning and deep learning. In this research, we are utilizing a deep learning system to identify breast cancer on mammograms screening to get high accuracy. CNN and deep learning algorithms are commonly employed for image classification. The tumors are identified using a convolutional neural network approach. The final stage in determining whether the lesion under observation is normal or cancerous, is classification. If it is determined to be a cancerous component, additional categorization is performed to establish cancer's future behavior, i.e., benign or malignant. Steps like segmentation and feature extraction are required for classification. Our study made use of mammograms from the Mammographic Image Analysis Society (MIAS) dataset that contains a total of 322 mammographic images. It is a work in progress, and additional progress is being made by optimizing the CNN architecture and using pre-trained networks, both of which should result in greater accuracy metrics.

Keywords: Mammographic Image Analysis Society (MIAS), breast cancer, CNN, image processing, machine learning, tumor, benign, malignant

INTRODUCTION

Cancer is a condition in which aberrant cells grow out of control and cause tissue destruction. Breast cancer is a kind of cancer that arises from breast tissue. Breast cancer is the second most often diagnosed malignancy in women, after skin cancer. Both men and women can get breast cancer, although women get it more commonly.

Being female, obesity, a lack of physical activity, alcoholism, hormonal imbalance, and radiation are all risk factors for developing breast cancer. About 5–10% of cases are caused by a genetic predisposition inherited from a person's parents.

Breast cancer was expected to afflict 2.3 million persons globally in 2020, with 6,85,000 deaths. As of the end of 2020, 7.8 million women have been diagnosed with breast cancer in the previous 5 years, making it the most frequent disease on the globe. Breast cancer affects women of all ages after puberty in every country, with rates increasing as they become older [1].

Breast cancer is usually treated with surgery, hormone treatment, radiation, and chemotherapy. If identified early, it is not hazardous, and imaging tools such as ultrasound imaging, MRI, and digital imaging can identify it. However, one of the most

***Author for Correspondence**
Nilima Bhosale
E-mail: nilima3421@gmail.com

[1]Student, Department of Computer Science, Pillai Hoc College of Engineering and Technology, Rasayani, Mumbai, Maharashtra, India
[2]Assistant Professor, Department of Computer Science, Pillai Hoc College of Engineering and Technology, Rasayani, Mumbai, Maharashtra, India

Received Date: April 18, 2022
Accepted Date: April 25, 2022
Published Date: April 30, 2022

Citation: Nilima Bhosale, Jagruti Bhagat, Shivani Bhoir, Rajashree Gadhave. Breast Cancer Detection Using Deep Learning. Journal of Artificial Intelligence Research & Advances. 2022; 9(1): 33–42p.

effective techniques for detecting breast abnormalities is mammography. A novel method for detecting breast cancer using a photo identification system is suggested. In this image recognition system, a convolutional neural network is a type of artificial neural network that has been successfully used to determine the visual description of a picture.

Tumors are divided into two groups: A benign tumor is one that does not infect surrounding tissues or spread throughout the body. Malignant (cancerous abnormality), benign (non-cancerous abnormality), and normal are the three classifications for mammography images [2].

As a result, offering clinical therapy to patients as soon as feasible after the early discovery of breast cancer can considerably enhance prognosis and survival prospects.

DIAGRAMMATIC REPRESENTATION

Figure 1 shows the CNN approach flowchart, whereas Figure 2 shows the MRI mammographic image.

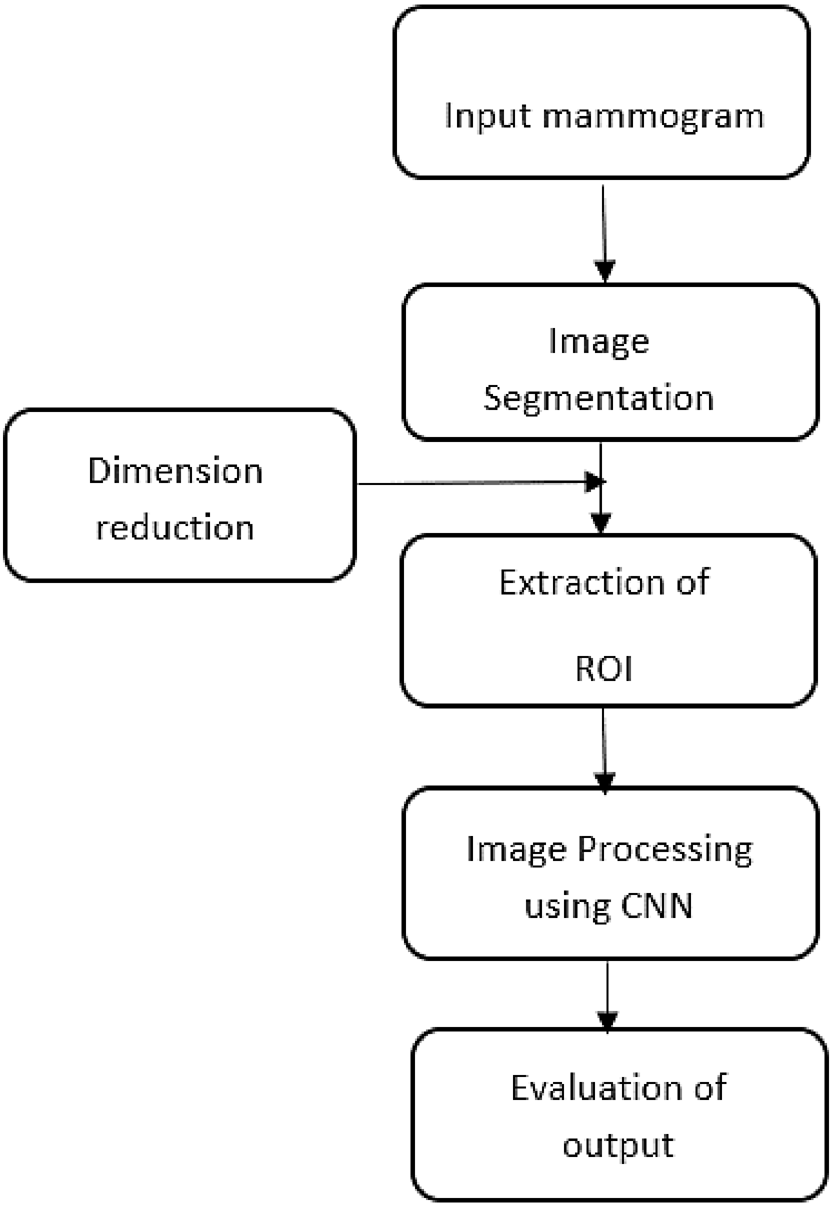

Figure 1. Flowchart for CNN approach.

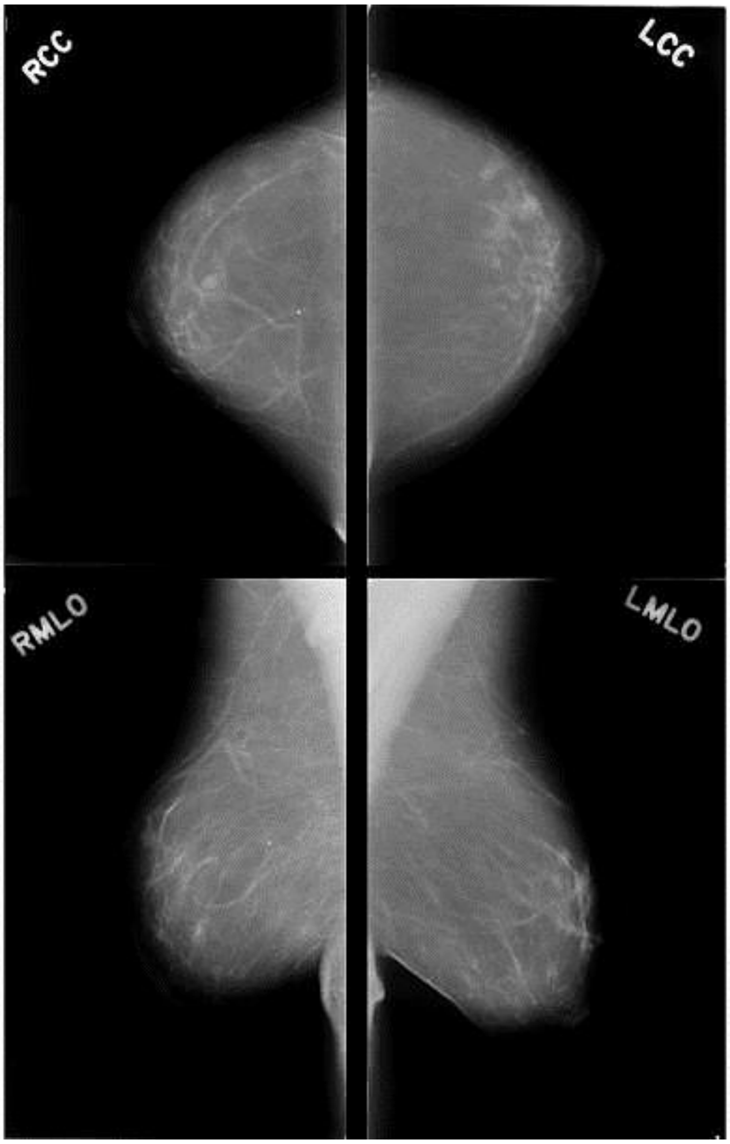

Figure 2. Mammographic Image.

LITERATURE SURVEY

There are various image classification techniques to achieve maximum possible accuracy. Let us review some of them in short:

Naive Bayes

For classification tasks, a probabilistic machine learning model known as a Naive Bayes classifier is used. The classification algorithm is based on the Bayes theorem.

We can calculate the probability of an occurrence $P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$ using Bayes' theorem as long as B has occurred. Here the evidence given in the above explained Figure 2 is the value of "B", and the hypothesis. The assumption here is that the predictors/features are self-contained. That is the existence of one attribute that has no effect on the opposite. As a result, it is known as Naïve. Classifiers enhance accuracy while lowering costs.

It takes a lot of time to compute and it is really fast. However, Naive Bayes is not very good at dealing with unknown characteristics. It is based on conditional probability, and if a circumstance has never occurred before, it is basically a very broad prediction that would not be particularly accurate [3].

Support Vector Machine

The SVM is the hyperplane with the greatest margin in a given space. The SVM is a linear classifier in its original form. In SVM, a hyperplane is a decision border that separates the two classes. Different classifications can be assigned to data points that lie on either side of the hyperplane.

Given a collection of coaching examples that are individually marked as belonging to one of two categories or the inverse, an SVM training approach develops a model that assigns new instances to at least one of two categories, making it a non-probabilistic binary linear classifier. New instances are subsequently mapped into the same area and given to a category that supports the gap's side. The SVM approach aims to reduce the distance between the data points and the hyperplane as much as possible [4].

KNN

KNN (K-Nearest Neighbors) classifier method is the algorithm which falls under supervised learning used in data processing and machine learning. It is a classifier technique where the training is based on "how similar" two sets of data are. It is a method that takes a long time to complete. KNN works by calculating the distance between a point and all of the data's examples, picking the needed number of instances (K) closest to the point, and then voting for the most frequent label. If there are lumps in the breast, the mammography will show you a picture of it. A biopsy will be taken if the lumps are discovered. The model then determines whether the cancer is benign or malignant [5].

Random Forest

A random forest is a group of decision trees that work together to solve a problem and is used as a classifier to address regression and classification issues. It avoids the overfitting problem by taking the average of all predictions, which cancels out the biases. The random forest model computes the variability of responses by constructing numerous decision trees, modeling each item under each tree, and then combining multiple trees to forecast the class of the dataset. In the event of classification, the element is given to the most likely class. Some decision trees may correctly anticipate the result, while others may not. However, when all of the trees are joined, they correctly forecast the outcome. Also, Random Forest takes a long time to make predictions since it uses several decision trees [6].

CNN

CNN is a more powerful and accurate approach of tackling classification issues in general. Because of its great accuracy, Convolutional Neural Networks are utilized for picture categorization and identification.

The process of image classification entails removing elements from an image in order to see other patterns in the dataset. The capacity of CNNs to enhance the internal representation of a two-dimensional image is one of its benefits. CNN detects important features without the assistance of a human.

This is done with an untrained CNN, which implies that every pixel of all features, as well as all weights in all fully linked layers, is assigned to a random value. You send an image into the network as an input, and it travels through many convolutions, subsampling, a completely connected layer, and lastly a fully connected layer [7].

The dataset was divided into two parts for the sake of ML algorithm implementation: the testing set and the training set. A comparison of all four methods will be performed. The method that produces the best results will be sent to the program as a model. Then it finally outputs something. Each image CNN processes results in a vote. Accuracy = number of correct predictions/total number of predictions.

The following Table 1 describes the various theories and their accuracy as proposed by various authors using various methodologies.

Table 1. Literature survey of various theories and their accuracy.

S.N.	Author	Methodology	Published in	Description	Accuracy
1	Bazila Banu A1, and Ponniah Thirumalaikolundu subramanian	Naïve Bayes	Journal of imaging	In this research, a comparison of several bayes classification strategies, as well as the boosting method, is explored in terms of accuracy %, sensitivity, and specificity. Naive Bayes Classifier with Gradient Boosting gives the highest accuracy with the lowest Mean Squared Error.	92.81%
2	Adel S. Assiri, Saima Nazir and Sergio A. Velastin [8]	SVM	PubMed Central® (PMC)	They examined the effectiveness of eight alternative classification algorithms on the WBCD testing dataset in this research. SVM learning performed well with two classifiers as well, therefore ensemble classification utilizing a voting method was devised.	97.08%
3	Melanie A. Sutton, James C. Bezdek	k-NN	Research gate	In this study, the crips k-nearest neighbor (k-NN) algorithm is used to aid in the diagnosis of breast cancer in digital mammograms. To obtain the final result, both supervised and unsupervised segmentation approaches, such as k-NN and fuzzy c-means, are used.	97.08%
4	Bin Dai, Rung-Ching Chen, Shun-Zhi Zhu, Wei-Wei Zhang	Random Forest	IEEE Xplore	The random forest technique is employed in this study to examine a medical case diagnosis of breast cancer. For accuracy prediction, the random forest technique is used to integrate the results of many decision trees.	94.90%
5	Maleika Heenaye-Mamode Khan, Nazmeen Boodoo-Jahangeer, Wasiimah Dullull, Shaista Nathire, Xiaohong Gao, G. R. Sinha, Kapil Kumar Nagwanshi	CNN	PLOS one	Of this research, network layers in a CNN are used as a detection filter to detect the presence of certain patterns or characteristics. CNN layers perform thorough categorization by combining all the characteristics discovered by previous levels.	88%

PROPOSED METHOD

There are various image classification techniques to achieve maximum possible accuracy. We have used the Convolutional Neural Network approach for identification of the tumor. CNN is a more powerful and accurate approach for tackling classification issues in general. Because of its great accuracy, Convolutional Neural Networks are utilized for picture categorization and identification. The process of image classification entails removing elements from an image in order to see other patterns in the dataset. The capacity of CNNs to enhance the internal representation of a two-dimensional image is one of its benefits. For the sake of algorithm implementation, the dataset was separated into two portions, with 85% of the data used for training and 15% used for testing.

Convolution layer, pooling layer, and fully connected layer are the three main layers. The input layer is where the datasets are learnt, and it is the initial layer. The input layer takes data that is being sent and adds some weight to it before it is sent to the hidden levels. To detect a pattern, the neurons of the hidden layer separate the features from the data. The pattern is then applied as the foundation for output layers that choose relevant classes. Training a convolutional neural network on raw pictures will almost certainly result in poor classification results. Therefore, the data is pre-processed and images are resized. Now, we have to eliminate the images that do not have class labels. We shall now input this pre-processed data into our network [8].

In the convolution layer, the image matrix is multiplied with the filter matrix to obtain the featured matrix. The pooling layer is used to minimize or reduce image size and to avoid the overfitting problem. Then, to extract more complicated characteristics, we use convolutions and pooling once again.

The characteristics are flattened down to a single layer so that the model may be fed into a fully connected neural network. After applying the SoftMax activation function, the desired outcome, whether benign or cancerous, is discovered; this is a classifying phase.

In this project, the input or original mammographic image will go through the process of Image Segmentation. Image segmentation is the process of lessening image complexity by extracting the area of interest. After extracting the area of interest, the cropped image will undergo the various image processing techniques while being fed to the CNN model.

We employed VGG19, a highly deep neural network that extracts more parameters from images [9].

These parameters then get passed through various layers that we have mentioned below:

- In Dropout layer, randomly selected neurons/nodes will be dropped out with a given probability at each iteration which helps to prevent the problem of overfitting. Now, data will be passed through a flatten layer.
- The process of flattening data into a one-dimensional array for input into the next layer is known as flattening.
- Batch normalization is a technique used to automatically normalize the inputs to each layer.
- Dense is usually used to modify the dimensionality of a vector.
- Conv2d are the convolution layers that will handle our inputs, which are represented as 2-dimensional matrices.
- Max Pooling layer will perform operations that calculate maximum value in each route.

So, our image data gets passed through all these layers and we get results into two categories as benign and malignant. We utilized Adam optimizer in our project to minimize mistakes or losses when fitting the model. The systematic CNN technique is presented in Figure 3 at several levels [10].

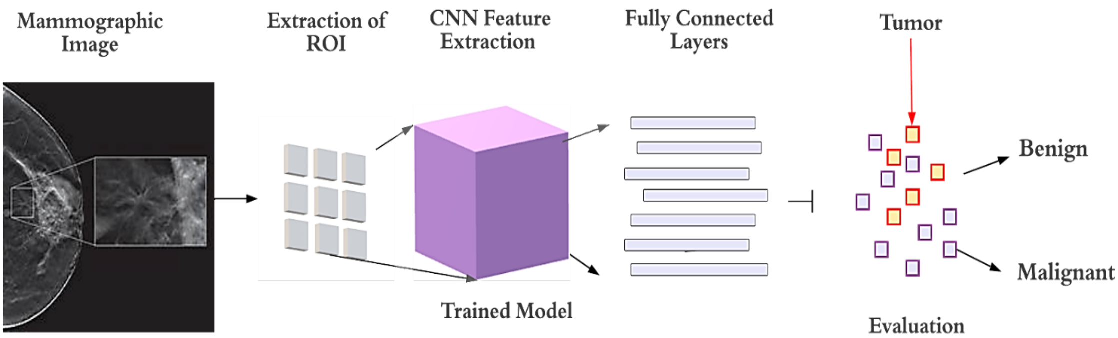

Figure 3. Diagrammatic representation for CNN approach.

RESULT AND DISCUSSION

Our system can estimate if an image is malignant or not and then show the appropriate category i.e., benign or malignant.

It will also provide the percentage of how certain the system is about the category that has been predicted by the algorithm. We have achieved the accuracy of 94.2% for our system.

It also provides a link to the official Breastcancer.org website, which is dedicated to delivering the most accurate, full, and easy-to-understand information about breast cancer and breast health to patients and caregivers.

This aids in the segmentation of a mammography image and can be utilized to construct an expert system for breast cancer early detection. This system can be used in a variety of applications, including natural scene segmentation and detection, as well as the segmentation and classification of blood cells.

We will also be able to focus on other types of cancer. Also, the system could be enhanced by using ensemble learning techniques. Patients spend less time in unpleasant conditions because digital technology is quick. The input mammogram will undergo various image filtering and convolutional layers, as a result of which the mammogram will be classified into two categories benign or malignant.

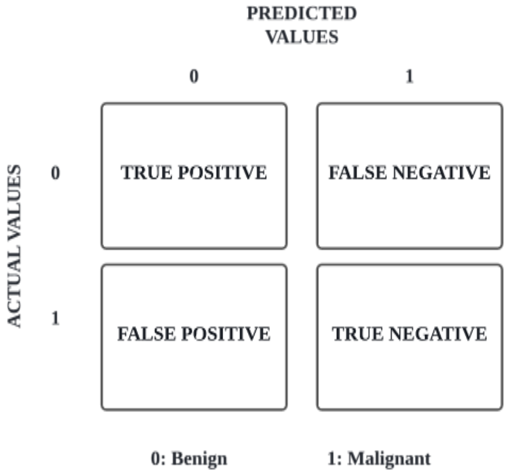

Figure 4. Confusion matrix I.

There are four different situations in this scenario (Figure 4).

Let us assume that 0 is holding a condition: Positive

- True positives (TP): These are examples in which we predicted 0 or condition is being satisfied (they have the condition) and the prediction is correct.
- True negatives (TN): Our model predicted that there is no condition, and the prediction is correct.
- False positives (FP): Our model projected that the condition is being satisfied but the prediction is false (A "Type I error" is a term used to describe this).
- False negatives (FN): We projected that there is no condition satisfied and the prediction is wrong (A "Type II error" is a term used to describe this).

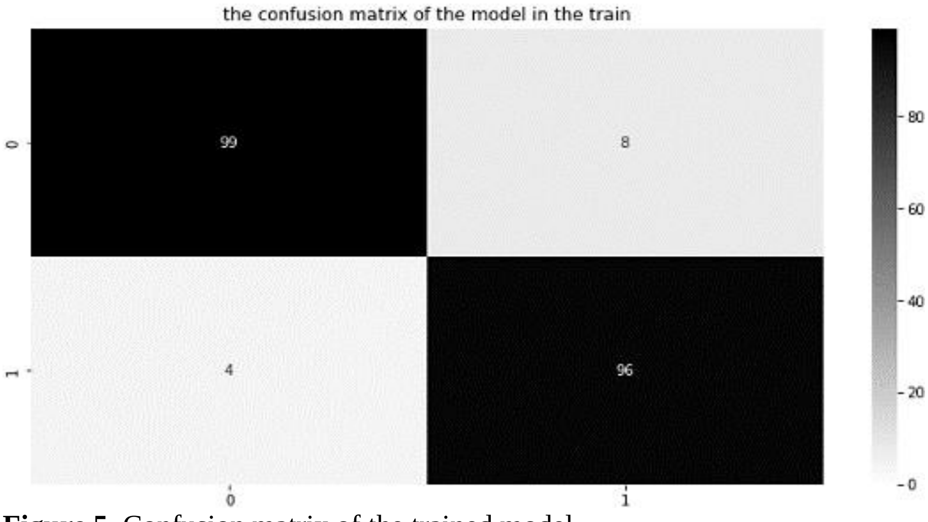

Figure 5. Confusion matrix of the trained model.

Accuracy (ACC)=TP+TN/P+N

Here, 0 indicates the benign image and 1 indicates the malignant image, where 99 images are benign and are predicted as truly benign, 96 images are malignant and are predicted as truly malignant whereas 8 images are benign but got predicted as malignant and 4 images are malignant but predicted as benign (Figure 5).

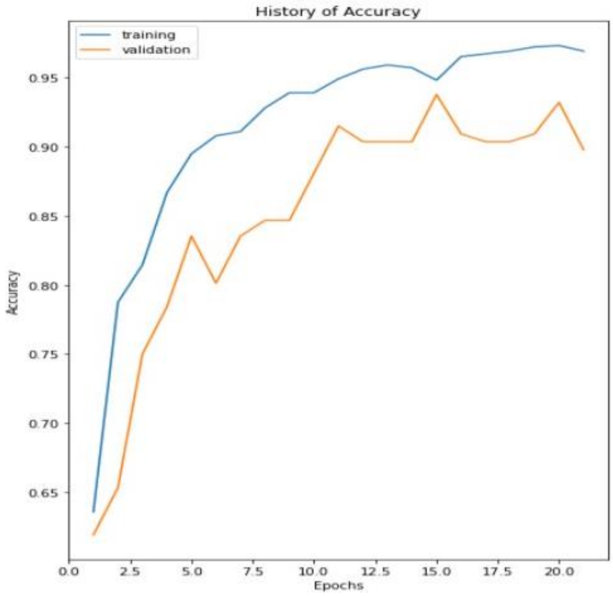

Figure 6. Epochs vs. Accuracy.

The epoch and accuracy are depicted in this Figure 6. When the epoch is elevated, the accuracy rises with it.

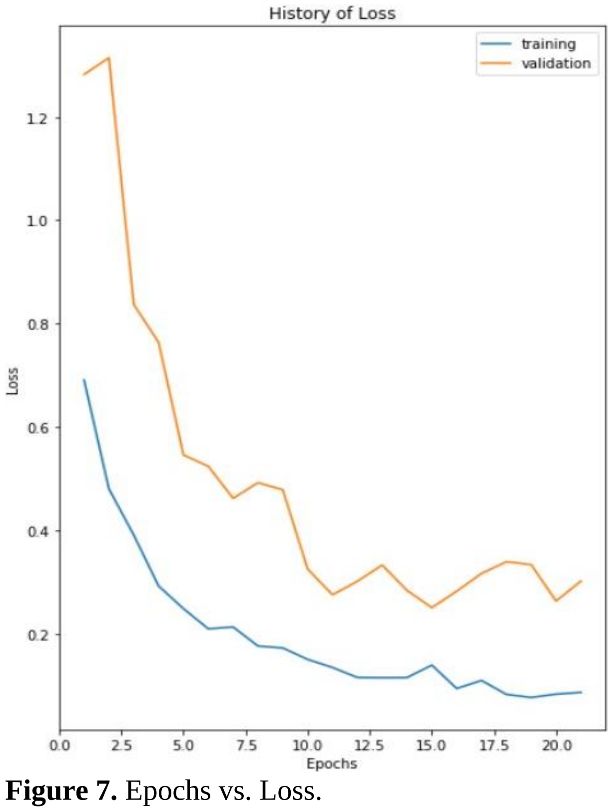

Figure 7. Epochs vs. Loss.

This figure represents the loss as a function of epoch. Here, Loss is inversely proportional to epochs (Figure 7).

Furthermore, the system helps both individuals and specialists diagnose tumors in the simplest manner possible (Figures 8 and 9).

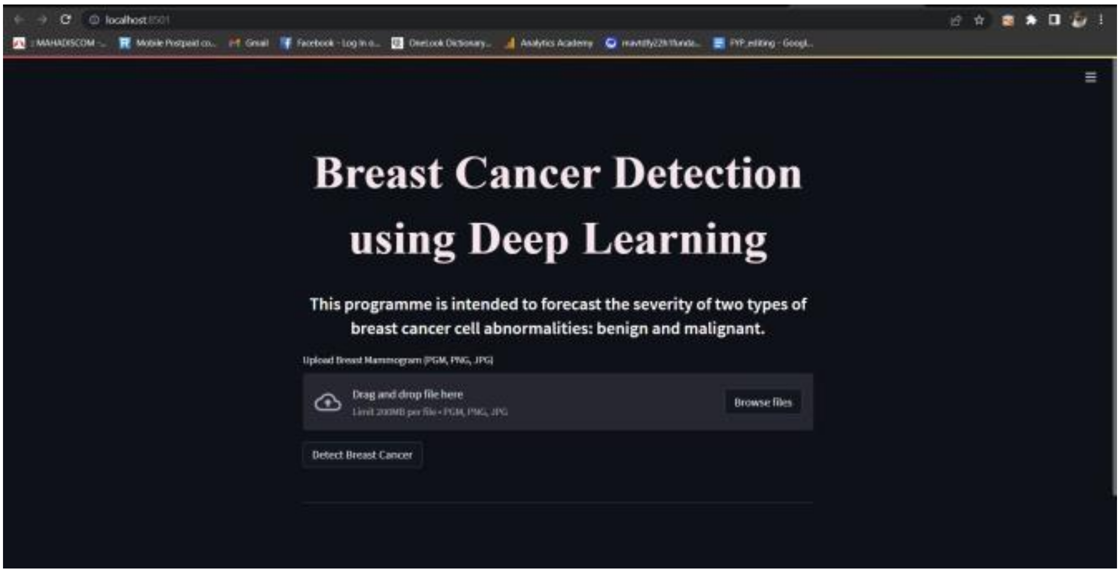

Figure 8. Home page of the system.

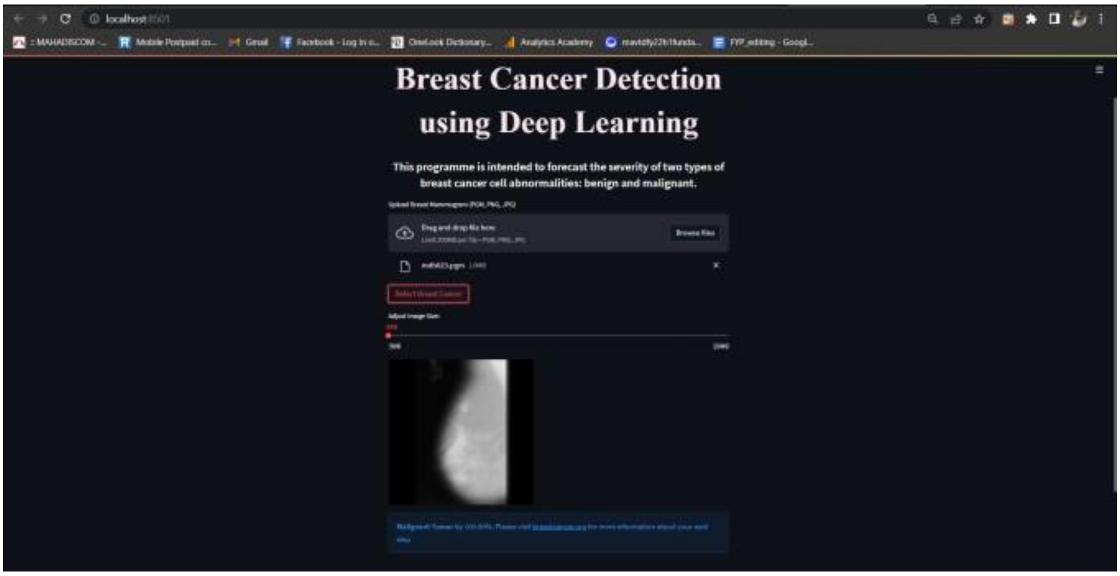

Figure 9. Prediction made by the system.

CONCLUSION AND FUTURE SCOPE

To minimize future problems and illness, any disease must be treated at an early stage. Tests for breast cancer detection systems are intended to detect cancer at an early stage, when it may be treated and perhaps cured. This will aid in the detection of cancers that are tiny and slow growing. This will decrease their risk of dying from breast cancer by up to 20%.

This project would aid in the medical industry by assisting real-world patients in gathering information concerning any tumor detection that may be there. The research of 8 publications has assisted us in gaining information and studying briefly about this issue, allowing us to collect data and propose a proposal on it. We will be able to categorize and forecast whether a tumor is benign or malignant by applying these deep learning techniques. It is a work in progress, and additional progress is being made by optimizing the CNN architecture and using pre-trained networks, both of which should result in greater accuracy metrics. As a consequence, we may draw conclusions by comparing the data given by various algorithms. We will try to implement in such a manner that the Convolutional Neural Network (CNN) will provide more accurate output for our research.

REFERENCES

1. Alanazi SA, Kamruzzaman MM, Islam Sarker MN, Alruwaili M, Alhwaiti Y, Alshammari N, Siddiqi MH. Boosting breast cancer detection using convolutional neural network. J Healthc Eng. 2021 Apr 3; 2021: Article ID 5528622.
2. Amin SU, Alsulaiman M, Muhammad G, Bencherif MA, Hossain MS. Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification. IEEE Access. 2019 Jan 29; 7: 18940–50.
3. Cahoon TC, Sutton MA, Bezdek JC. Breast cancer detection using image processing techniques. In 9th IEEE International Conference on Fuzzy Systems. FUZZ-IEEE 2000 (Cat. No. 00CH37063). 2000 May 7; 2: 973–976.
4. Rane N, Sunny J, Kanade R, Devi S. Breast cancer classification and prediction using machine learning. Int J Eng Res Technol. 2020; 9(02): 576–80.
5. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. Sci Rep. 2019 Aug 29; 9(1): 1–2.
6. Shen Y, Wu N, Phang J, Park J, Liu K, Tyagi S, Heacock L, Kim SG, Moy L, Cho K, Geras KJ. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. Med Image Anal. 2021 Feb 1; 68: 101908.
7. Chagpar AB, Coccia M. Factors associated with breast cancer mortality-per-incident case in low-to-middle income countries (LMICs). J Clin Oncol. 2019; 37(15 Suppl): 1566–1566.
8. Assiri AS, Nazir S, Velastin SA. Breast tumor classification using an ensemble machine learning method. J Imaging. 2020 Jun; 6(6): 39.
9. Charan S, Khan MJ, Khurshid K. Breast cancer detection in mammograms using convolutional neural network. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE. 2018 Mar 3; 1–5.
10. Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks and support vector machines. Peer J. 2019 Jan 28; 7: e6201.

