Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

ANALYSIS ON OPTIMISED NEURAL NETWORK-BASED CLASSIFICATION FOR BREST CANCER DISEASE DETECTION

[1]V. Kusuma Priya, [2]K. Sravani, [3]Katamreddy. Mahendra, [4]V.Venkata Sai Karthik
[1,2,3]Dept of Computer Science and Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.
[4]Dept of Electronics and Communication Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.

ABSTRACT
The convolutional neural network-based method for identifying breast detecting cells (CNN) that we present in this study is innovative. Breast cancer can cause benign or malignant tumours. For medical diagnosis, a breast cancer tumour must be given the appropriate categorization. In this study, a (CNN) method for breast cancer detection is suggested. It looks at various convolutional neural network (CNN) designs that could be applied to automatically detect breast cancer. [1] Making a breast cancer diagnosis is a difficult task that must be finished to improve patient treatment. According to a recent survey, using CNN is the best way to locate and get trustworthy results. This may lower human error in diagnosis and lower the price of cancer detection. [2] Our approach can identify regions of masses and classify them as benign or cancerous.[3] On the test dataset, the model demonstrated 95% detection accuracy and a 94% AUC-ROC.
Keywords: CNN, Benign, Malignant, Resnet50, SVM, LR.

1. INTRODUCTION

A form of cancer known as breast cancer starts in the breast cells and spreads to other body organs. Women are more likely than men to develop breast cancer. The stage of a cancer determines how it is treated. A family of disorders characterised by benign changes to breast tissue are referred to as "benign" diseases. The majority of benign breast conditions don't increase the risk of breast cancer. Cancerous tumours are referred to as malignant tumours. If left untreated, malignant cells have the capacity to spread from the main tumour to other parts of the body. Breast cells gave rise to the malignant tumour known as breast cancer.[4]

At some time in their lives, one in every eight women will be diagnosed with breast cancer. The best approaches to reduce the mortality rate from breast cancer are early detection and effective treatment because there is currently no preventive intervention that is widely accepted. If breast cancer is found early enough, before it becomes localised tumours, it can be successfully treated. As a result, accurate breast cancer diagnosis has emerged as a critical and popular research issue.[5]

Cancer is now more common than ever and is a serious public health concern. According to data from the Global Burden of Disease Cancer Collaboration and the WHO's International Agency for Research on Cancer, there were 28% more cancer diagnoses between 2006 and 2016 and 2.7 million more cases are anticipated by 2030. Breast cancer is one of the mainly prevalent and dangerous type of tumor, secretarial for 1.7 million incident cases, 535,000 fatalities, and 14.9 million years of life with a disability. Therefore, early detection of breast cancer is crucial. Breast cancer has been detected using X-rays, MRIs (Magnetic Resonance Imaging), and ultrasounds for more than 40 years.[6]

Classification is a supervised learning approach in machine learning and statistics in which a programmed learns from provided data and then applies what it has learnt to classify fresh observations.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

2. RELATED WORK

2.1 Support Vector Machine (SVM): SVMs, or vector support machines. In three-dimensional spaces, it functions well. [7] This method displays each data point as a point in an n-dimensional space, where n represents the feature number and each feature value indicates a unique coordinate value. Once the hyper plane that best distinguishes the two classes has been identified, classification can begin.[8,9]

2.2 K-Nearest Neighbour (KNN): With the aid of training datasets for pattern recognition, it is a technique for locating k's closest relatives in subsequent samples. In order to define a large number of training samples near the new point and use them to predict the label, the nearest-neighbor method is used. In k-nearest neighbour (k-NN) learning, the sample number can be manually specified, but it can also fluctuate dependent on the point density in the area.[10]

Any metric unit can be used to express the distance; the usual choice is the standard Euclidean distance. Because of its core structure, the nearest neighbour is also available for a wide range of datasets and can produce better results for complex border patterns. The greater the value of K, the smoother the boundary appears to be.

2.3 Logistic Regression(LR):"Yes/No,""Pass/Fail,""Alive/Dead," and so on are examples of binary outcomes that can be forecasted using logistic regression.[11] If we consider IDC (+) to be 1 and IDC () to be 0, the result is a categorical 0 or 1, which may be defined as

$P(Y = 1|X)$

or $P(Y = 0|X)$.

Depending on the sequence of the polynomials, the decision border of logistic regression may be either linear or nonlinear, creating a complex system. A decision boundary for a logistic regression may be linear or nonlinear.[12, 13] The cost function cannot be an R-squared function since it is not convex. The cost function's distribution matches that of the dependent variable in logistic regression, which uses the Bernoulli distribution.

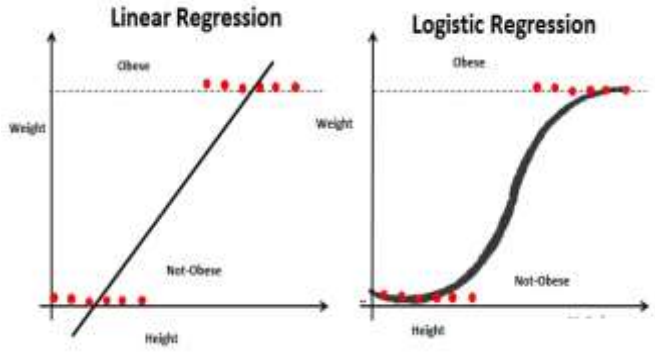

Figure 1: Linear and Logistic Regression

3. PRELIMINARIES

3.1 Preliminary concept:

3.1.1 CNN: A Convolutional Neural Network (Conv Net/CNN) is a Deep Learning system that can take an input image, give different aspects and objects in the image weights and biases that can be learned, and distinguish between them. Comparatively less pre-processing is needed for a convolutional network than for traditional classification techniques. With enough practise, Conv Nets may learn these filters and attributes, whereas with crude approaches, filters must be hand-engineered. To extract certain features from a picture while keeping the spatial arrangement information, we can take the input image, create a weight matrix, and then convolve the input.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

3.2.2 CNN Architecture: A Conv Net's structure is modelled after how the visual cortex is organised and reflects how neurons connect in the human brain. Only when they are in the Receptive Field, a small area of the visual field, can individual neurons respond. A number of equivalent fields can be stacked on top of one another to span the whole visual field.

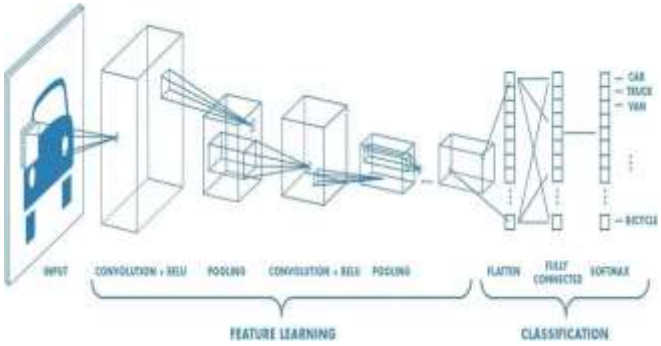

Figure 2: Architecture of CNN

To define a basic convolutional network, we need three components:

- The convolutional layer
- The Pooling layer
- The output layer

The features of the input image are initially extracted using the convolution layer. In order to maintain the connection between pixels, the convolutional layer learns visual properties from a small square of input data. A kernel or filter and an image matrix are the two inputs required by this mathematical technique.

To properly pre-process images, the pooling layer is essential. When the photos are too big, the pooling layer minimises the number of parameters. Pooling is the process of "downscaling" an image composed of prior layers. Maximum and average pooling are the two forms. The essential data is preserved. After many layers of padding and convolution, we must obtain the output as a class. to produce a product with the same number of classes as what we need.

4. METHODOLOGY

4.1 Convolutional Neural Network Model: The main advantage of CNN is automatic feature extraction. Hubel et al.'s 1968 experimental study served as the foundation for this idea. The fitness function that was utilized to create this model is as follows:

$$\text{Fitness Function} = (TP+TN)/TP+FP+FN+TN$$

Where, TP=True Positive FP=False Positive

TN=True Negative FN=False Negative

In this experimental attempt, learning transfer was optimised for a multi-layered deep CNN architecture known as Resnet50. The pre-built model ResNet-50 was trained to distinguish different photographs in 1,000 classes using the Image Net dataset. Initial weights for the proposed deep neural network were pre-trained weights from ImageNet. High gradient values must be transferred from ResNet50's residual layers to its preceding and adjoining levels.The vanishing gradient problem can be solved and complicated and useful patterns can be extracted using this layer [35,36]. All pre-train model layers in our experimental setup are left open to pick up new characteristics from biopsy images.The CNN layers sent the feature matrices to the fine-tuned FC layer, which then applied the sigmoid function to the output layer.

Additionally, accuracy was increased by using the Adam optimiser. The loss function displays the discrepancy between the actual and expected value. Twenty was chosen for Epoch.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

4.2 Metrics and Evaluation Parameters: The model's classification accuracy rate is primarily utilised to determine its value. Additional quality metrics like precision, recall, and F1 score are examined using the confusion matrix. The binary classification performance indicators are represented using a 2*2 confusion matrix. True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) values are displayed in each column. The terms "true positive class" and "true negative class" refer to situations in which a true positive class is determined to exist while a true negative class is actually present. TN stands for a correctly predicted negative class, while FN stands for a positive class that was misinterpreted as a negative class. A key metric for any CNN model is the accuracy outcome.

The algorithm used to describe the CNN model using RESNET 50 is depicted in the figure. Networks with several layers—even thousands—can be easily trained without increasing the training error percentage. ResNet-50 is the name of a convolutional neural network with 50 layers. A pre-trained version of the network that has been trained on more than a million photos can be loaded from the Image Net database.

4.3 RESNET 50 MODEL:

Algorithm1: Algorithm for Convolutional Neural Network (CNN)

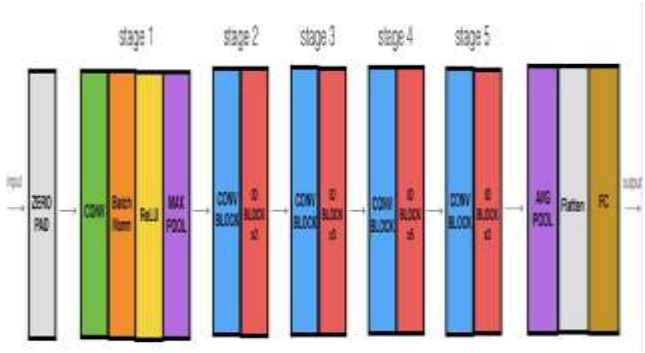

Figure 3: Resnet 50 Architecture

Figure provides a detailed illustration of the neural network's design. "ID BLOCK" in the diagram stands for "Identity block," while "ID BLOCK x3" denotes the stacking of three identity blocks.

- Zero-padding: a pad of zero is applied to the input (3,3)

Stage 1: The 2D A stride of (2,2) and 64 different shape (7,7) filters are used in convolution. Its name is "conv1".On the input's channels axis, Batch Norm is applied. Max Pooling uses a window of (3,3) and a stride of (2,2).

Stage 2: The convolutional block employs three sets of 64x64x256, f=3, and s=1 filters. The two identity blocks employ three sets of filters with a combined size of 64x64x256 and f=3.

Stage 3: Three sets of filters, each with a size of 128x128x512, f=3 and s=2, are used in the convolutional block.The three identity blocks utilise a set of three filters with a size of 128x128x512 and f=3.

Stage 4: Three sets of 256x256x1024, f=3, s=2 filters are used in the convolutional block. The size 256x256x1024, f=3 filters are used by the five identity blocks.

Stage 5: The convolutional block employs three sets of 512x512x2048, f=3 and s=2 filters.The two identity blocks employ a three-set of 256x256x2048 (f=3) filters.

In 2D Average Pooling, a window with the shape (2,2) is known as the "Avg pool".

4.4 BLOCK DIAGRAM:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

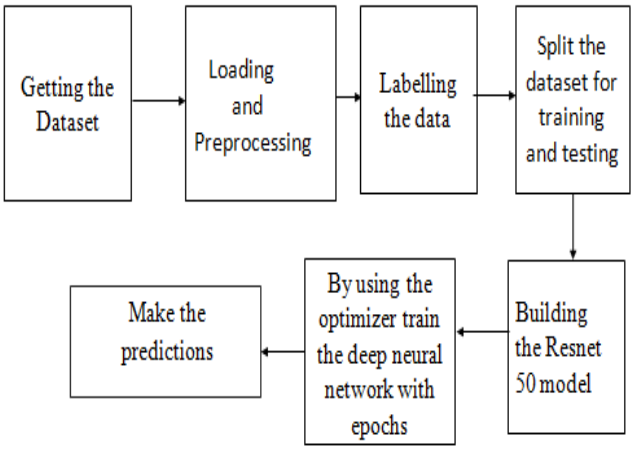

Figure 4: Block diagram of breast cancer classification process

4.4.1 GETTING THE DATASET:

A data set is an ordered collection of linked items and values that apply to a particular business and may be accessed individually or collectively.

There are two ways to gather data for our model:

a. Rely on open source data
b. Collect your data in a right way.

4.4.2 LOADING AND PRE-PROCESSING:

Loading:

In Google Co-Lab, add the relevant libraries and dataset to the working directory.

Pre-processing:

Pre-processing data is a technique for turning unclean data into a clean data set. To put it another way, whenever data is gathered from numerous sources, it is done so in a form that prevents analysis.

4.4.3 LABELLING THE DATA:

Data that includes samples that have been marked with one or more labels is known as labelled data. A labelled dataset can be used to feed more unlabeled data into machine learning models, which can then estimate or predict a plausible label for that unlabeled piece of data.

4.4.4 SPLIT THE DATASET FOR TRAINING AND TESTING:

We normally split the data between testing and training stages in the range of 20% to 80%. In Python ML, supervised learning divides a dataset into training and test data.

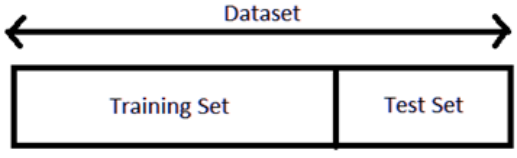

Figure 5: Splitting the dataset

4.4.5 BUILDING THE MODEL:

This stage entails running Resnet 50, which is a needed algorithm.

4.4.6 RESNET-50:

The ResNet-50 model contains five stages, each of which has a unique convolution and identity block. There are three convolution layers in each identity block and three convolution layers in each convolution block.
There are about 23 million trainable parameters in the ResNet-50.
The flatten has no hyper parameters or a name.
The Fully Connected (Dense) layer reduces the number of classes in its input by employing a softmax activation.

4.4.7 Learning rate:

The step size, also referred to as the "learning rate," is the increment by which the weights are changed during exercise. The learning rate, a hyper parameter that may be modified and has a moderately positive value typically between 0.0 and 1.0, is used to train neural networks.
A factor called learning rate affects how quickly the model adapts to its surroundings. Greater learning rates necessitate fewer training epochs due to the smaller weight changes with each update, whereas smaller learning rates necessitate more training epochs.

4.4.8 *Epoch:*

The learning algorithm's frequency of iterations through the full training dataset is controlled by the number of epochs, a hyper parameter. Every epoch, the underlying model parameters have been modified once for each training dataset sample. An era consists of one or more batches. For instance, the batch gradient descent learning algorithm is named after an era in which there was only one batch. It is common to see line graphs with time represented by epochs along the x-axis, and the model's skill or error on the y-axis. These diagrams are often known as learning curves. These graphs can indicate if the model has learned too much, too little, or is well fitted to the training set of data.

4.4.9 Make the predictions:

The accuracy, confusion matrix, ROC curve, and AUC value must all be predicted in this step. The machine will be put to the test with a test dataset, and the actual and anticipated outcomes will be compared.

4.5.0 Dropout rate:

A training method called dropout involves rejecting a sample of randomly chosen neurons. Random events cause them to "disappear." As a result, their temporally temporal temporal contribution to the activity of downstream neurons is removed on the forward pass, and on the backward pass, no weight changes are added to the cell.
Since a fully connected layer makes up the majority of the parameters, co-dependency between neurons during training reduces the unique power of each neuron and causes training data to be over-fit.

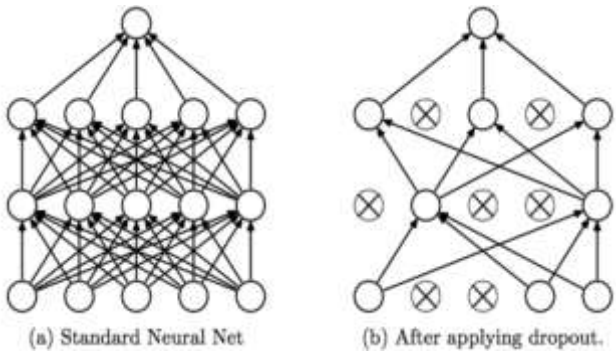

Figure 6: Drop out rate

5. RESULT AND DISCUSSION

5.1 CONFUSION MATRIX

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

In the field of machine learning, and more specifically the issue of statistical classification, an error matrix, also known as a confusion matrix, is a table that is frequently used to describe how well a classification model (or "classifier") performs on a set of test data for which the true values are known. The effectiveness of an algorithm can be seen through visualisation. It also makes it simpler to identify instances of class confusion, such as when one class is frequently mislabeled as the other. The confusion matrix is used to generate the bulk of performance metrics.

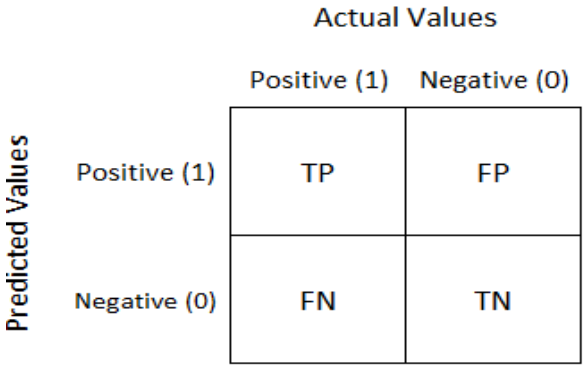

Figure 7: Confusion Matrix

5.2 DATA SET

1000 FNA photographs made up the research's input data. 200 images are utilised for testing and 800 images are used for training in that data set. The samples were taken from human breast tissue using fine needle aspirates. Only a small number of benign and malignant photos have ever been taken for testing; there are many benign and malignant photos taken for training..

5.3 PERFORMANCE MATRICES:

5.3.1 Accuracy:

The ability of the classifier to deliver accurate diagnoses is measured by its accuracy.

Accuracy= (TP+TN)/(TP+FP+TN+FN)

5.3.2 F-Score:

Precision and recall are functions of the F1-score. When a balance of precision and recall is required, it is calculated.

F-Score=(2*Recall*Precision)/(Recall+Precision)

5.3.3 Precision:

It's the proportion of positively predicted items that were accurately forecasted to the total number of items predicted.

Precision=TP/(TP+FP)

5.3.4 Recall:

It determines how many true positives the model has recorded and labels them as such.

Recall=TP/(TP+FN)

5.3.5 ROC Curve:

The Receiver Operating Characteristic Curve (ROCC) is a useful tool for assessing diagnostic tests. Sensitivity vs. 1-Specificity is shown by a graph with two dimensions. Each point on the ROC curve has a sensitivity/specificity pair that relates to a specific decision threshold. A parameter's capacity to distinguish between two diagnostic groups (diseased/normal) is measured by the AUC. The AUC range is [0,1]. AUC values near 1 signify a highly trustworthy diagnostic test. The AUC in this case was 0.940.

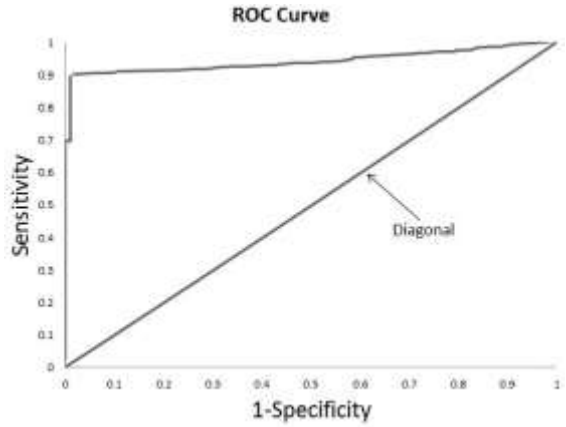

Figure 8: ROC Curve

TABLE 1: PERFORMANCE COMPARISION

Method	Accuracy (%)	Precision (%)	Recall (%)
LR	71	59	80
KNN	71.2	72	91
SVM	78.5	82	92
CNN(thisstudy)	95	92	98

5.3.6 Comparison with other methods:

The suggested approach The accuracy, precision, and recall of CNN are compared to those of logistic regression (LR), k-nearest neighbour (k-NN), and support vector machine (SVM). The breast cancer classification accuracy of CNN's database is 95%. Our approach, which uses a neural network's topology to be genetically evolved, outperformed the other three approaches with an accuracy rate of 95%. Comparing this to the BPNN approach is a tremendous improvement. We also contrasted the results of our approach with past research on this issue.

Along with neural networks, we also compared our results to other machine learning algorithms, like support vector machines. The precision of the least squares support vector machine was very low.

OUTPUT RESULTS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 7, No. 1, July 2021

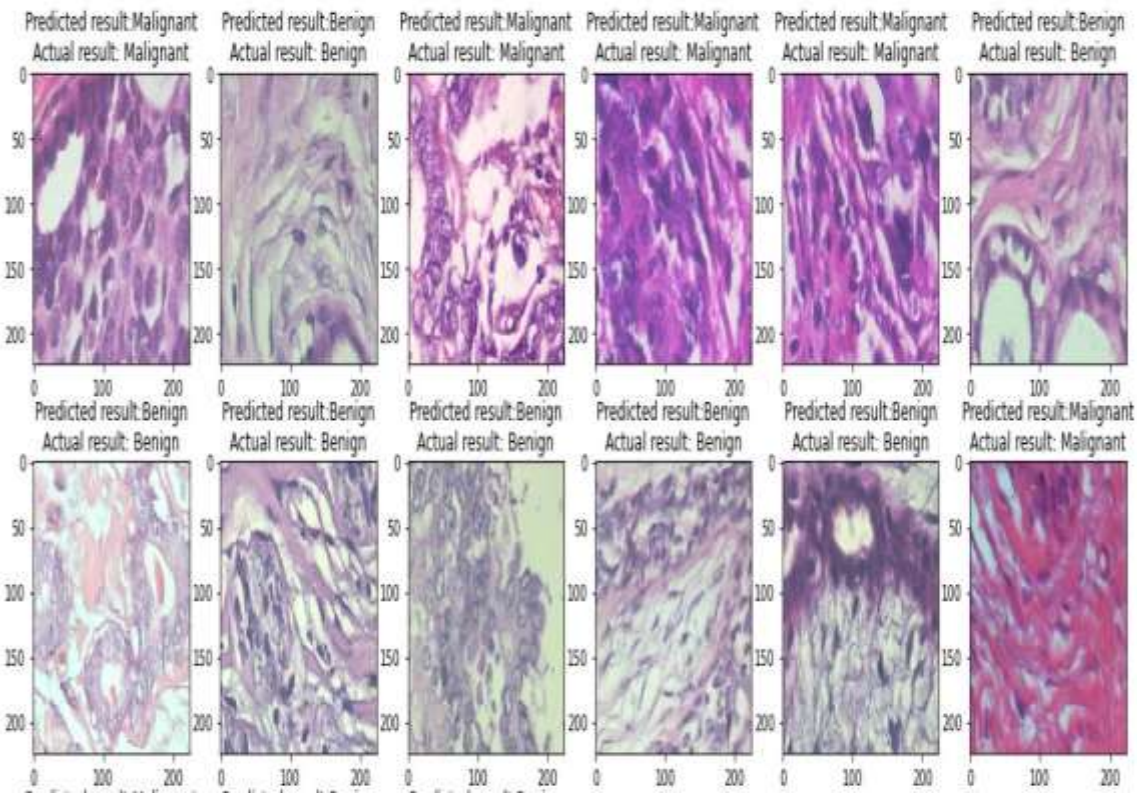

Figure 9: Output Images

CONCLUSION

This study's novel breast cancer detection method uses Python programming to genetically create a convolutional neural network in order to produce the optimal classification structure. This is a better choice when the ideal topology of a neural network is unknown or it is difficult to determine an ideal structure through trial and error. With regard to the Wisconsin Breast Cancer database, our method yields a 95% classification accuracy.

REFERENCES

[1]. Saad Awadh Alanazi, M. M. Kamruzzaman, Md Nazirul Islam Sarker Boosting Breast Cancer Detection Using Convolutional Neural Network , Journal of Healthcare Engineering Volume 2021, Article ID 5528622.

[2]. Ben-Ari et al. (2017) Rami Ben-Ari, Ayelet Akselrod-Ballin, Leonid Karlinsky, and SharbellHashoul. Domain specific convolutional neural networks. Pages. 552-556 in the IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).. IEEE, 2017.

[3]. Digital Image Processing with MATLAB, R.C. Gonzalez, R.E. Woods, and S.L. Eddins, prentice hall, New Jersey, USA, 2004.

[4]. Tavassoli FA and Devilee P (eds.) (2003) World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Tract. Lyon, France: IARC Press.

[5]. Martin and Weber (2000) Anne-Marie Martin and Barbara L Weber. Breast cancer risk factors include both genetic and environmental variables. National Cancer Institute Journal, 92(14), 1126-1135, 2000.

[6]. "Support Vector Networks," C. Cortes and V.N. Vapnik, Machine Learning Boston, Vol. 3, pg 273-297, September 1995.

[7]. J edge et. al. " Magnetic resonance imaging of the breast : A clinical Perspective ", South African journal of radiology , Vol 16, No. 2, pp:61-64,2012.

[8]. Apratim Sadhu Prediction of Breast Cancer using Support Vector Machines International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056Volume: 08 Issue: 02 | Feb 2021.

[9]. PramitBrata Chanda; Subir Kumar Sarkar Detection And Classification Technique Of Breast Cancer Using Multi Kernal SVM Classifier Approach Published in: 2018 IEEE Applied Signal Processing Conference (ASPCON) 07-09 December 2018.

[10]. Bonfring International Journal of Advances in Image Processing, "An Efficient CBIR Approach for Diagnosing the Stages of Breast Cancer Using KNN Classifier,"Vol. 2, No. 1, pp. 1–5, March 2012. Jini R. Marsilin and G. Wiselin Jiji, "An Efficient CBIR Approach for Diagnosing the Stages of Breast Cancer Using KNN Classifier," Bonfring International Journal of Advances in Image Processing.

[11]. LR-HIDS: Logistic Regression Host-based Intrusion Detection System for Cloud Environments (E. Besharati, M. Naderan, E. Namjoo) J. Ambient Intell. Hum. Comput., 10 (9) (2019), pp. 3669-3692.

[12]. Alok Chauhan ,HarshwardhanKharpate, Yogesh Narekar, Sakshi Gulhane, Tanvi virulkar, Yamini Hedau, 2021 third international conference on inventive research in computing applications for SVM, KNN IEEE 02-04 september 2021.

[13]. Ram MurtiRawat; Shivam Panchal; Vivek Kumar Singh; Yash Panchal Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning Published in: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, 02-04 July 2020.