Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 11, November : 2021

CLASSIFICATION OF A BREST CANCER DISEASE USING OPTIMIZED NEURAL NETWORK

[1]Lakshmi Deepthi Nukala, [2]Shephali Singh, [3]Sujana B, [4]Gummalla Aakanksha
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Computer science and Engineering, Rishi MS Institute of Engineering and Technlogy for Women, Kukatpally, Hyderabad.

ABSTRACT

In this study, we provide a novel convolutional neural network-based method for classifying breast detection cells (CNN). Tumours from breast cancer can be benign or malignant. The correct classification of a breast cancer tumour is essential for medical diagnosis. In this paper, a convolutional neural network (CNN) approach is suggested for the detection of breast cancer. It looks into the suggested method for automatically detecting breast cancer using several convolutional neural network (CNN) architectures. [1] It is a difficult task to diagnose breast cancer in order to improve patient treatment. A recent study suggests using CNN to find and get accurate findings, which may lessen human error in the diagnosis process and lower the cost of cancer detection. [2] Our model recognises a mass region and categorises them into benign or malignant.[3] The performance of the model on test dataset is found to be: detection accuracy 95%, and AUC-ROC of 94%.

Keywords: CNN, Benign, Malignant, Resnet50, SVM, LR.

1. INTRODUCTION

Breast cancer is a type of cancer that develops in the breast cells and spreads to other parts of the body. Breast cancer is more likely to affect women than men. How a cancer is treated depends on its stage. The term "benign" refers to a class of diseases marked by benign changes to breast tissue. Most benign breast diseases don't raise the chance of breast cancer. Malignant tumors are tumors that have cancer. Malignant cells have the ability to move outside of the primary tumor to other regions of the body if untreated. Breast cancer is a malignant tumor that originated from breast cells.[4]

At some time in her life, one out of every eight women will be diagnosed with breast cancer. Given that there is currently no universally accepted preventive method, early identification and good treatment are the key options for reducing breast cancer mortality. Localized tumors can be successfully treated before the cancer spreads if breast cancer is diagnosed early enough. As a result, accurate breast cancer diagnosis has emerged as a critical and in research area.[5]

Cancer has risen to prominence as a serious public health issue. According to data from the WHO's IARC (International Agency for Research on Cancer) and the GBD (Global Burden of Disease Cancer Collaboration), cancer diagnoses climbed by 28% between 2006 and 2016, with 2.7 million new cases expected by 2030. Breast cancer is one of the most frequent and lethal cancers in women among the numerous types of cancer (1.7 million incident cases, 535,000 deaths, and 14.9 million disability-adjusted life years). As a result, it is crucial to detect breast cancer early. Although X-ray, MRI (Magnetic Resonance Imaging), and ultrasound have all been used to diagnose breast cancer 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.

2. RELATED WORK

2.1 Support Vector Machine (SVM): Vector Support Machines (SVMs) (SVM). It works well in three-dimensional spaces.[7] This algorithm plots each data item as a point in an n-dimensional space, where n signifies the feature number and each feature value indicates a unique coordinate value. After obtaining the hyperplane that best distinguishes the two classes, classification can be performed.[8,9]

2.2 K-Nearest Neighbour (KNN): It is a pattern recognition technique that uses training datasets to find k's closest relatives in subsequent samples. The nearest-neighbor algorithm's theory is utilized to define many training samples next to the new point and use them to forecast the label. The sample number can be set by the user, as in k-nearest neighbour (k-NN) learning, or it can change based on the local point density.[10]

Any metric measure can be used for, the distance; standard Euclidean distance is a popular choice. Due to its basic structure, which can provide superior results for complicated borders, the nearest neighbour is also available for a wide number of datasets.The bigger 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 $P(Y = 0|X)$.

The decision border of logistic regression can be linear or nonlinear, with an increase in the polynomial order resulting in a complex systemLogistic regression's decision boundary might be either linear or nonlinear.[12, 13] Because of its non convex nature, the cost function cannot be an R-squared function. The dependent variable follows the same distribution as the cost function in logistic regression (which includes the Bernoulli distribution).

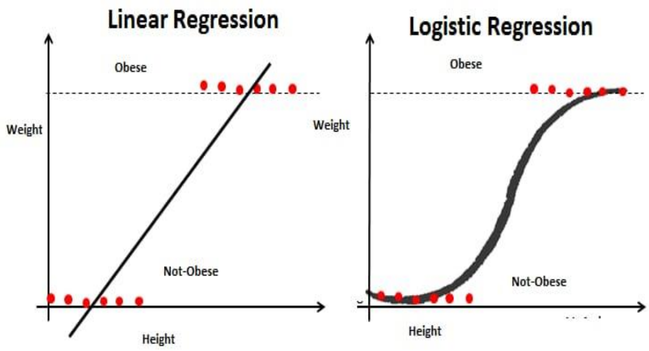

Figure 1: Linear and Logistic Regression

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, assign relevance (learnable weights and biases) to various aspects/objects in the image, and distinguish between them. When compared to other classification techniques, the amount of pre-processing required by a Conv Net is significantly less. While crude approaches require hand-engineering of filters, Conv Nets can learn thesefilters/characteristics with adequate training. We can take the input image, build a weight matrix, then convolve the input to extract certain features from the image while preserving the spatial arrangement information.

Through the use of necessary filters, a Conv Net is able to successfully capture the Spatial and Temporal dependencies in a picture. The architecture provides better fitting to the picture dataset because to the

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 11, November : 2021

reduced number of parameters involved and the reusability of weights.In other words, the network may be trained to better understand the image's sophistication.

3.2.2 CNN Architecture: The architecture of a Conv Net is inspired by the arrangement of the Visual Cortex and is akin to the connectivity pattern of Neurons in the Human Brain. Individual neurons can only respond to stimuli in the Receptive Field, a tiny portion of the visual field. To cover the entire visual field, a number of comparable fields can be piled on top of each other.

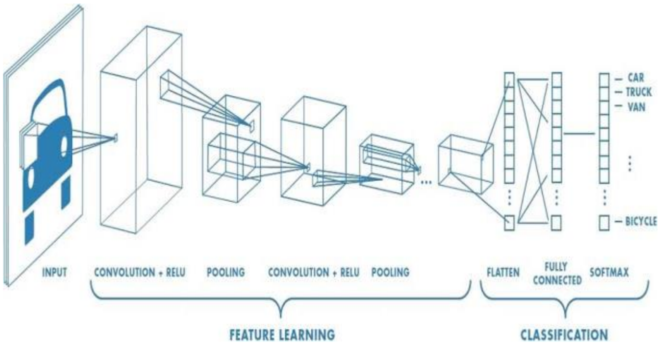

Figure 2: Architecture of CNN

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

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

The convolution layer is the first to extract features from an input image.The convolutional layer preserves the link between pixels by learning visual properties using a tiny square of input data. It's a mathematical process that accepts two inputs: a matrix of images and a kernel or filter.

In picture pre-processing, the pooling layer is very significant. When the photos are too huge, the pooling layer minimises the number of parameters. Pooling is the process of "downscaling" the image created by the previous layers. Maximum and average pooling are the two types of pooling.It retains the important information.

We need to retrieve the output in the form of a class after numerous layers of convolution and padding. To create an output with the same amount of classes as we require.

4. METHODOLOGY

4.1 Convolutional Neural Network Model: Automatic feature extraction is the prime benefit of using CNN. This concept was originated through experimental work done by Hubel et al. in 1968.

The fitness function that was utilized to create this model is as follows:

Fitness Function= (TP+TN)/TP+FP+FN+TN

Where, TP=True Positive FP=False Positive

TN=True Negative FN=False Negative

In this experimental work, a multi-layered deep CNN framework Resnet50 was tuned for transferring the learning process. ResNet-50 is a prebuilt model which has been trained on the Image Net dataset for identifying different images of 1,000 classes. As initial weights for the proposed deep neural network, ImageNet pre-trained weights were provided. ResNet50's residual layers play a vital function in transferring big gradient values to its preceding neighbouring layers.The model can extract complicated and relevant patterns and address the vanishing gradient problem thanks to this layer [35,36]. All pre-train model layers are maintained open in our experimental setup to learn new characteristics from biopsy photos.The feature matrices, acquired from CNN layers, were supplied to the fine-tuned FC layer, where the sigmoid function was used in the output layer.

Further, the Adam optimiser was applied for achieving better accuracy. Loss function indicates the difference between the actual and predicted value. Epoch was set to 20. The available data were highly imbalanced in their sample number, so for the data balancing, proper weights for each class were assigned. The whole dataset was divided into different subsets as follows:training dataset with multiple images to train the model, validation dataset has few images for tuning network parameters and test subsets have images to evaluate model performance.

4.2 Evaluation Parameters and Metrics: The model's value is primarily judged by its classification accuracy rate. Additional quality parameters such as precision, recall, and F1 score are examined using the confusion matrix. The performance measurements for binary classification are represented by a 2*2 confusion matrix. True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) values are represented in each cell. The TP stands for correct prediction of true positive class, while the FP stands for circumstances where a true negative class exists but is decided to be a positive class. TN denotes a correctly predicted negative class, whereas FN denotes a positive class mistakenly as a negative class. The accuracy result is a crucial metric for any CNN Model.

The CNN model using RESNET 50 step-by- step procedure is described in algorithm as shown in the figure. Networks with large number (even thousands) of layers can be trained easily without increasing the training error percentage. ResNet-50 is a convolutional neural network that is 50 layers deep. You can use the ImageNet database to load a pre-trained version of the network that has been trained on over a million photos.

4.3 RESNET 50 MODEL:

Algorithm1: Algorithmfor Convolutional Neural Network (CNN)

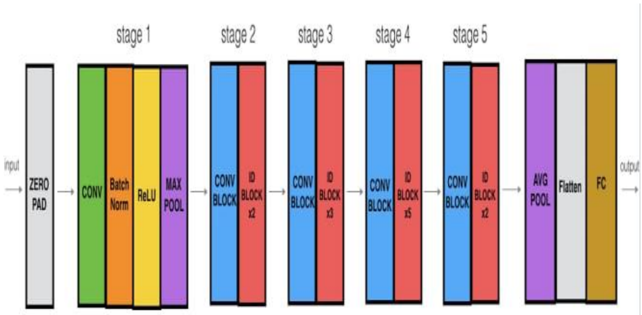

Figure 3: Resnet 50 Architecture

The architecture of this neural network is depicted in detail in Figure. In the diagram, "ID BLOCK" stands for "Identity block," while "ID BLOCK x3" indicates that three identity blocks are stacked together.

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

Stage 1: The 2D Convolution has 64 filters of shape (7,7) and uses a stride of (2,2). Its name is “conv1”.Batch
Norm isapplied to the channels axis of the input. Max Pooling uses a (3,3) window and a (2,2) stride.

Stage 2: The convolutional block uses three set of filters of size 64x64x256, f=3, s=1. The two identityblocks use
three set offilters of size 64x64x256, f=3.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 11, November : 2021

Stage 3: The convolutional block uses three set of filters of size128x128x512, f=3 , s=2 .The three identity blocks
use three set of filters of size 128x128x512, f=3.
Stage 4: The convolutional block uses three set of filters of size 256x256x1024, f=3, s=2 . The five identity blocks
use the threeset of filters of size 256x256x1024, f=3 .
Stage 5: The convolutional block uses three set of filters of size512x512x2048, f=3, s=2 .The two identity blocks
use the threeset of filters of size 256x256x2048, f=3.
The "Avg pool" is a window of shape (2,2) that is used in 2D Average Pooling.

4.4 BLOCK DIAGRAM:

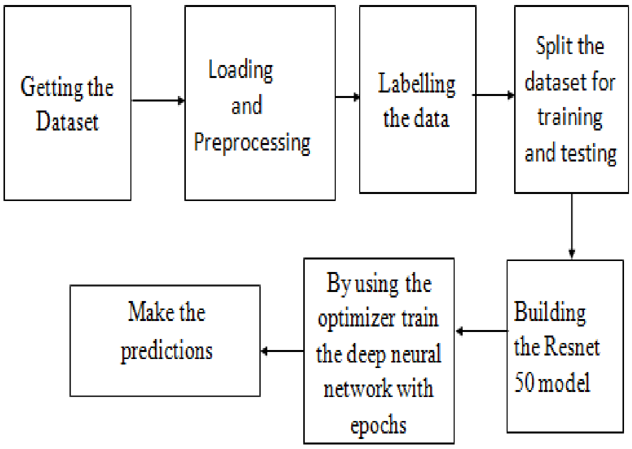

Figure 4: Block diagram of breast cancer classification process

4.4.1 GETTING THE DATASET:

A data set is a collection of connected entities and values pertaining to a specific business that can be accessed individually or as a single element and is organised using a data structure.
Two ways to collect a 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:
Data pre-processing is a method for transforming raw data into a clean data set. To put it another way, anytime data is received from many sources, it is collected in raw format, which makes analysis impossible.

4.4.3 LABELLING THE DATA:

A collection of samples that have been labelled with one or more labels is referred to as labelled data. Machine learning models can be used to a labelled dataset in order to supply additional unlabeled data to the model and estimate or predict a likely label for that piece of unlabeled 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.

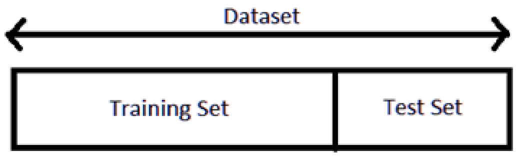

Figure 5: Splitting the dataset

4.4.5 BUILDING THE MODEL:

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

4.4.6 RESNET-50:

Each of the five stages of the ResNet-50 model has its own convolution and identity block. Each convolution block has three convolution layers, and each identity block has three convolution layers. The ResNet-50 has roughly 23 million trainable parameters.

There are no hyperparameters or a name for the flatten.

Using a softmax activation, the Fully Connected (Dense) layer lowers the number of classes in its input.

4.4.7 Learning rate:

The step size, often known as the "learning rate," is the amount by which the weights are changed during training. The learning rate is an adjustable hyperparameter that has a modest positive value, usually between 0.0 and 1.0, and is used in the training of neural networks.

The learning rate is a parameter that determines how quickly the model adapts to the situation. Given the smaller changes to the weights each update, lesser learning rates necessitate more training epochs, whereas greater learning rates necessitate fewer training epochs.

4.4.8 *Epoch:*

The number of epochs is a hyperparameter that controls how many times the learning algorithm runs over the whole training dataset. Each sample in the training dataset has had the opportunity to update the internal model parameters once each epoch. There are one or more batches in an epoch. The batch gradient descent learning algorithm, for example, is named after an epoch with only one batch.Line charts with epochs along the x-axis as time and the error or skill of the model on the y-axis are frequent. Learning curves are another name for these charts. These graphs can show if the model has over-learned, under-learned, or is well-fit to the training 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:

Dropout is a training approach in which randomly selected neurons are rejected. They are "disappeared" at random. This means that on the forward pass, their contribution to the activation of downstream neurons is removed temporally, and on the backward pass, any weight updates are not applied to the neuron.

Because a fully connected layer takes up the majority of the parameters, neurons develop co-dependency with one another during training, limiting each neuron's unique power and resulting in over-fitting of training data.

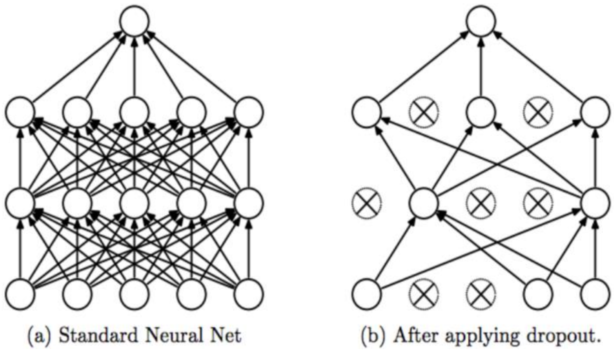

Figure 6: Drop out rate

5. RESULT AND DISCUSSION

5.1 CONFUSION MATRIX

In the field of machine learning, and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a table that is frequently used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. It allows the performance of an algorithm to be visualised. It also makes class confusion easier to spot, such as when one class is routinely mislabelled as the other. The majority of performance measures are calculated using the confusion matrix.

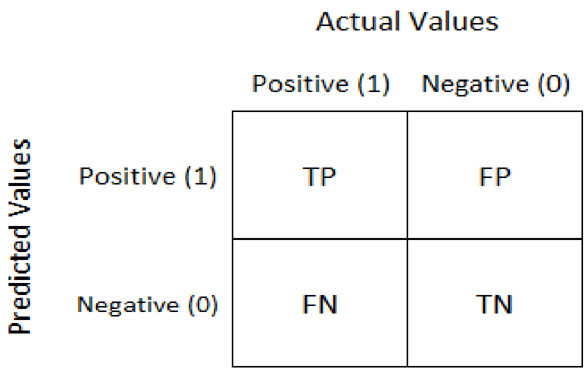

Figure 7: Confusion Matrix

5.2 DATA SET

The input data for this research consisted of 1000 FNA photos. In that data set, 800 photos are used for training and 200 images are used for testing. Fine Needle Aspirates of human breast tissue were used to get the samples. There are many Benign and Malignant photos taken for training, but only a few Benign and Malignant images taken for testing.

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.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 11, November : 2021

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. It's a Sensitivity vs. 1-Specificity graph with two dimensions (True Positive fraction vs. False Positive fraction). Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a certain decision threshold. The AUC (area under the receiver operating characteristic curve) is a measurement of a parameter's ability to differentiate between two diagnostic groups (diseased/normal). [0,1] is the AUC range. AUC values close to 1 indicate a very reliable diagnostic test. In our instance, the AUC was 0.940.

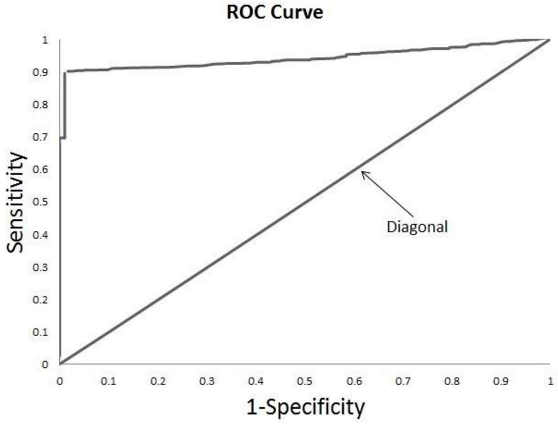

Figure 8: ROC Curve

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 proposed method CNN is compared to three different approaches in terms of accuracy, precision, and recall: logistic regression (LR), k-nearest neighbour (k-NN), and support vector machine (SVM). CNN's breast cancer database has a 95% classification accuracy. Our method, which involves genetically evolving the structure of a neural network, achieved a 95 percent accuracy rate, which is much greater than the other three methods. This is a huge improvement over the BPNN method. We also compared the outcomes of our method to earlier work on this problem.

In addition to neural networks, we compared our results to those of other machine learning algorithms such as support vector machines. The precision of the least square support vector machine was extremely low.

OUTPUT RESULTS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 11, November : 2021

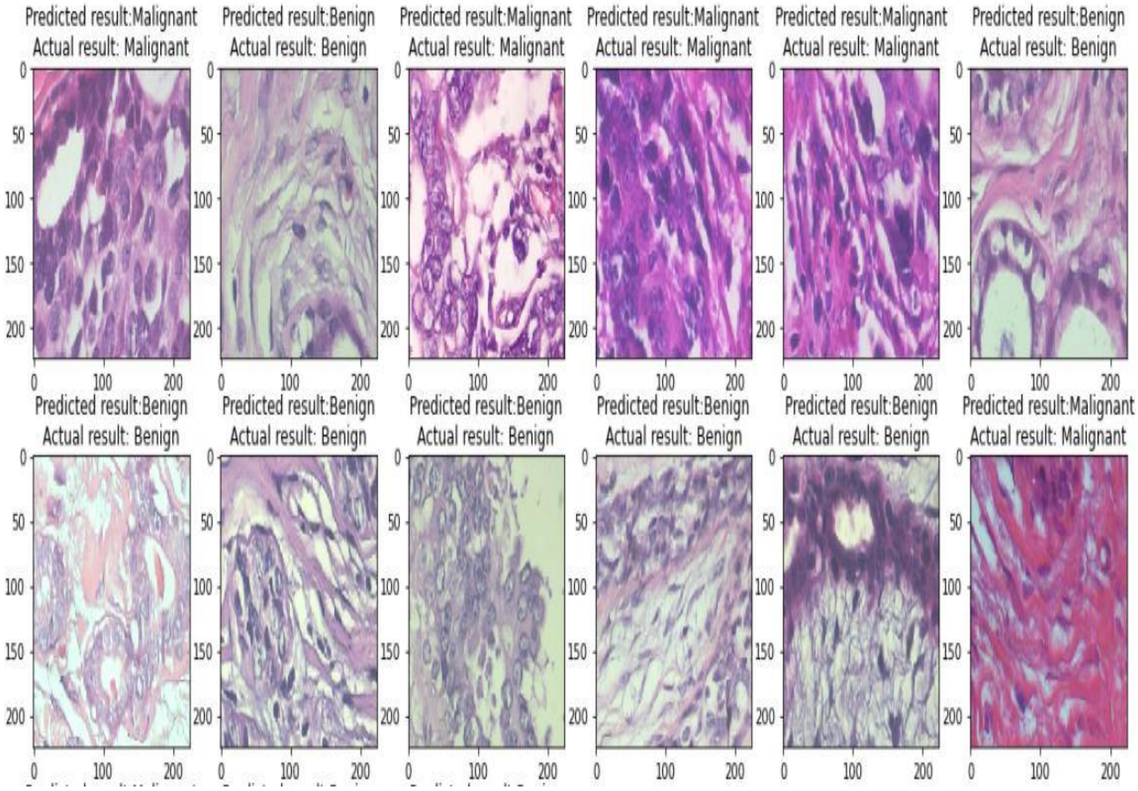

Figure 9: Output Images

CONCLUSION

In order to produce the best classification structure, a convolutional neural network is genetically constructed in this study's unique breast cancer detection method utilising Python programming. When the perfect topology of a neural network is unknown or it is challenging to establish an optimal structure by trial and error, this is a preferable option. Our method obtains a 95% classification accuracy for the Wisconsin Breast Cancer database.

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.