BREAST CANCER CLASSIFICATION USING DEEP LEARNING

[1]Dr. S P Abirami , [2]R.S.Yokesh , [3]Sanjay Pratap T K , [4]Nidin S , [5]Abuthageer

[1]Assistant Professor Department Of Computer Science And Engineering,

Coimbatore institute of Technology

[2]Student Department Of Computer Science And Engineering,

Coimbatore institute of Technology

Abstract- *Globally, breast cancer is the most common cancer among women, and the most likely cause of female cancer deaths as per statistical analysis given by WHO. High-income countries (HICs) have made the most progress in improving breast cancer outcomes. Between 1990 and 2014, breast cancer death rates dropped by 34% in the US attributable to the combination of improved earlier detection and effective adjuvant therapies. By contrast, breast cancer is an increasingly urgent problem in low- and middle-income countries (LMICs), where historically low incidence rates have been rising by up to 5% per year. In view of earlier detection of breast cancer, the research aims in developing a classification model using CNN.*

The model primarily uses Wisconsin Dataset for training the model and classifies the cancer tissues. The accuracy of the model is evaluated and found to be better compared to other existing models.

Another model that uses EfficientNetB7(CNN) model to evaluate the Breast tumour cells to classify whether the given histopathological images are benign or malignant or normal in nature.

Keywords- *Breast Cancer, WHO,CNN, EfficientNetB7, Histopathological, Benign, Malignant.*

I.INTRODUCTION

Breast Cancer is the most affected disease between women all over the world. All most 25% of all cancers with an estimated 1.67 million new cancer cases diagnosed in 2012 and its incidence is increasing day by days. It is also said that Breast cancer is the second leading cause of death for women worldwide.

Early detection of cancer can reduce the risk of deaths for cancer patients**.** To increase the survival rate the early diagnosis of breast cancer and a trustworthy detection model is required.

The goal of the research is to identify and classify Malignant and Benign patients by proposing a simple and efficient model for early detection of breast cancer with minimal error percentage.

This model uses an Artificial neural network and produces Accuracy of 98% and f1 score as 98% - Benign and 97% - Malignant

EfficientNetB7 model uses the Histopathological image dataset and process the image and classifies the image into Benign or Malignant or normal class based on the analysis of mammographic cells. the total number of layers in EfficientNet-B7 the total is 813. All these layers are divided into 5 modules. The proposed Efficient NetB7 model uses the concept of Transfer learning.

Transfer learning is a machine learning research subject that focuses on storing and transferring knowledge learned while addressing one problem to a different but related problem. For instance, skills learned when learning to recognise vehicles could be applied to recognising trucks. So first a set of layers are made constant and the last few layers that are necessary for the classification are attached at the end and then trained for particular those layers . then in the second process of training all layers trained in such a way that knowledge gained by the model in image classification is now used for classifying breast tumor cells.

II.RELATED WORKS

In recent years many machine learning and deep learning models are propped to classify the breast cancer

data the patients.

Here[1] Moh'd Rasoul Al-hadidi ,Abdulsalam Alarabeyyat and Mohannad Alhanahnah have a proposed amodel where they first preprocessed the mammography images and abstracted the information and buld an BPNN and Logitic regression model to classify the images.

In [2] they have used the Wisonsin Dataset in detection and classification of the Breast cancer based on different attributes. Applying Naïve Bayes, SVM, Logistic Regression, KNN, Random Forest, MLP and CNN classifiers, higher accuracy is obtained which is up to 98% to 99%.

In [3] they've proposed the deep learning method convolutional neural network that mostly used for classification of images dataset. It is basically divided into three parts first collection of dataset and applied pre-processing algorithm for scaled and filter data then split dataset for training and testing purpose and generate some graph for visualization data. They've concluded deep learning technology is a good way for diagnosis breast cancer with Wisconsin Breast Dataset. This database provides 569 rows and 30 features in the dataset. In this paper they've just used 11 features for classification.

Another work of Nur Syahmi Ismail and Cheab Sovuthy [4] they've used deep learning technique using VGG16 and ResNet50 network to classify between normal tumour and abnormal tumour using IRMA dataset and have implemented for normal and abnormal breast cancer detection The best result of classification accuracy was VGG16 with 94%. Compared to ResNet50 with 91.7% in term of accuracy.

Ramik Rawal [5] In this paper they have used four algorithms SVM, Logistic Regression, Random Forest, and KNN to predict the outcome the breast cancer. It broadly speaks about three domains. First domain is prediction of cancer before diagnosis, second domain is prediction of diagnosis and treatment and third domain focuses on outcome during treatment. The accuracy obtained by SVM (97.13%) is better than the accuracy obtained Naïve Bayes and k-NN that have an accuracy that varies between 95.12 % and 95.28 %.

III.METHODOLOGY

The section describes the methods adopted for classification and Various other details for the experiment.

A.Dataset

The WBCD diagnosis dataset comprises 699 samples including malignant. This dataset becomes train the deep model as it is separated into train and test subsets.

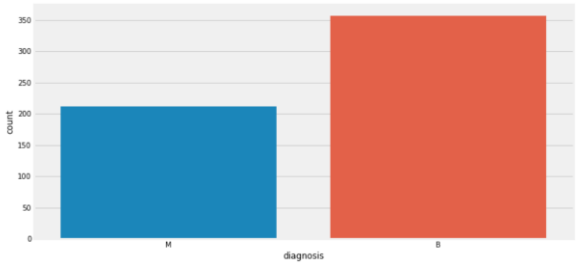


Figure 1. A counter plot for total number of malignant and benign count in the dataset.

For the CNN model the dataset that is taken is the histopathological images from Kaggle.

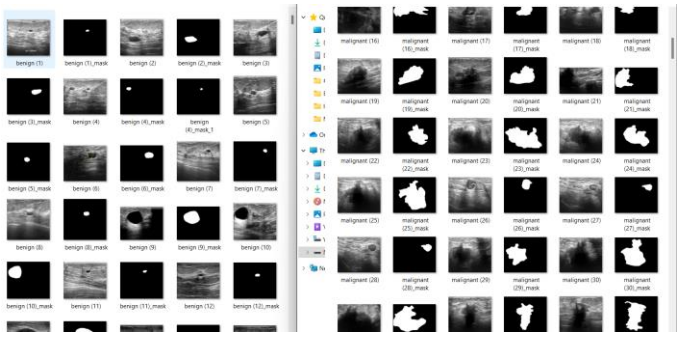


Figure 2 A screenshot of the downloaded image dataset.

B.Data visualization

The Wisconsin data is plotted against a heatmap and their correlations are identified.

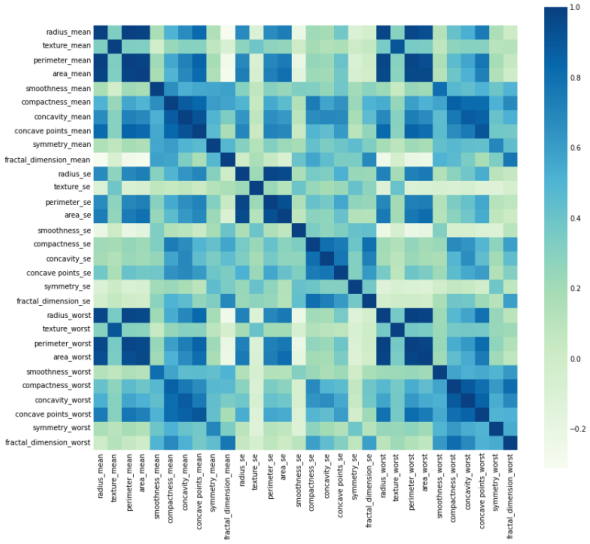


Figure 3 Heatmap to show the correlations between the features.

C.*PREPROCESSING*

The wisconsin dataset is preprocessing and scaling function to normalize the data. The so-called Min-Max scaling method is an alternative to Z-score normalisation (or standardisation) (often also simply called "normalization" - a common cause for ambiguities). The data is scaled to a fixed range - usually 0 to 1 - in this method.

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}}.$$

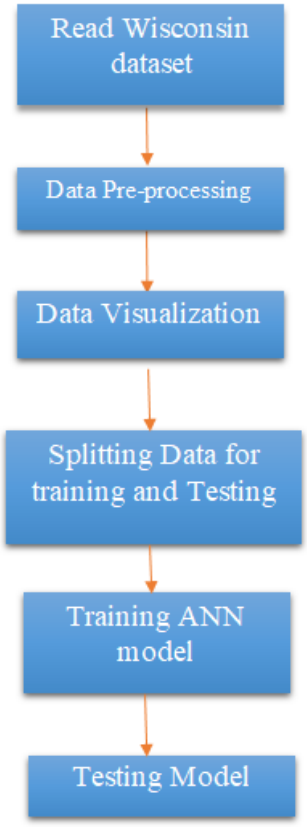
For the CNN model the image encoded to the respective class and prepared for the training of model.

IV.EXPERIMENT

A. ANN model

The data is imported and then processed and normalized using MinMax scaler.
The data is then split to train and test parts with the random state of 42 and then 33% of the data is separated for testing purpose and the remaining for training.

1.*FLOW OF THE MODEL*



2. MODEL SETUP

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 15)	465
dropout (Dropout)	(None, 15)	0
dense_1 (Dense)	(None, 30)	480
dropout_1 (Dropout)	(None, 30)	0
dense_2 (Dense)	(None, 1)	31

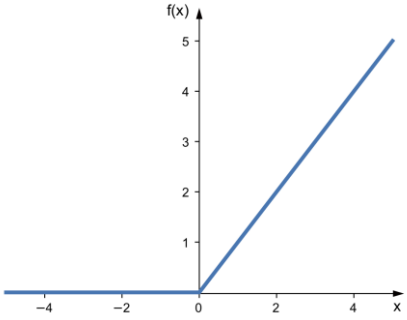
Total params: 976
Trainable params: 976
Non-trainable params: 0

The simple ANN model is trained with 30 inputs as features and then a two hidden layers with the activation function as ReLU and one output in the output which gives the value either 0/1.

a.) ReLU function

The activation function of the Rectified Linear Unit (ReLU) is defined as the positive part of its argument.

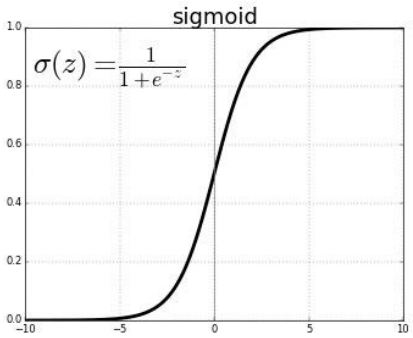
$$f(x) = \begin{cases} x & \text{if } x > 0, \\ 0.01x & \text{otherwise.} \end{cases}$$



b.) Sigmoid Function

A mathematical function with a characteristic "S"-shaped curve, also known as a sigmoid curve, is known as a sigmoid function.
The logistic function is a frequent example of a sigmoid function.

$$S(x) = \frac{1}{1 + e^{-x}}$$



3.Training the Model

The model is trained with the adam optimizer and loss function to be with binary_crossentropy. And the results are obtained.

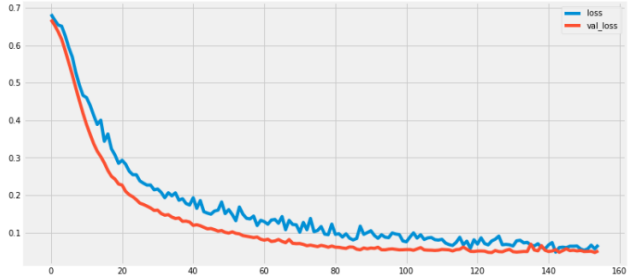


Figure 5 Training Vs Validation loss of the ANN model.

B.CNN model

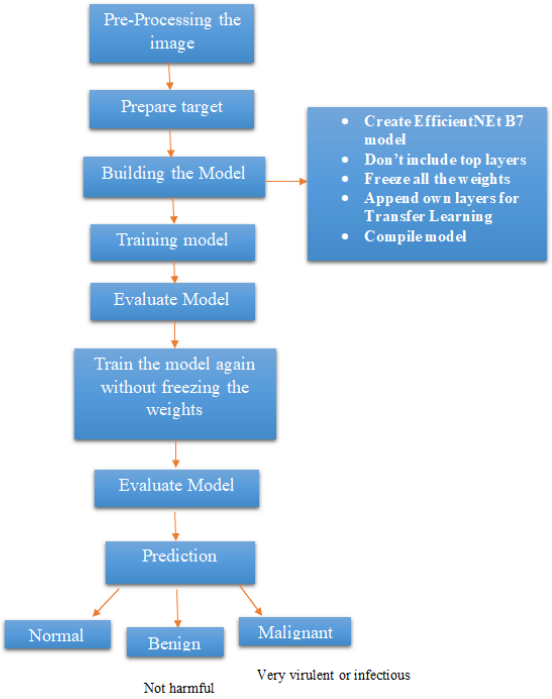
1. IMAGE PREPROCESSING

The EfficientNet B7 model is imported for the application of this model for image classification.

The images are preprocessed and the fiunction to convert image to array format is written.

Images are read from all the directory and the images are split for the testing and training purpose.

2. FLOW CHART



3. BUILD MODEL

The base model EfficientNetB7 model is imported and all the layers of the first part are set false , that is they are freezed in the first iteration.

The required number of new layers are appended at the end the model

The layers include two convolution layers , where the images datasets are converted o 2 dimensional matrices and are padded . the data is normalized and the activation function used here is that ReLU.

And two other fully connected layer , where the 2 dimensional matrices is reduced to small set of images , flattened and given to the output layer . In the output layer the image is recognised either malignant or Normal

or Benign , either of three classes is given as output , and is activated using softmax function.

With the second iteration the whole EfficientNetB7 model is used with making the top layers as well to train the model.

4. TRAINING THE MODEL

The model is thus trained and the results are thus obtained. The prediction is done with type of image given and the respective output is given output.

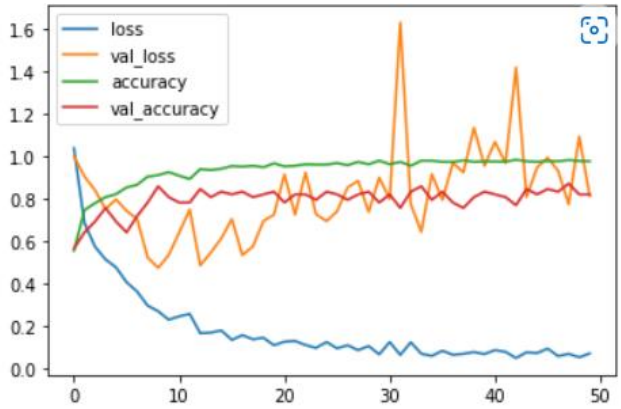


Figure 7 show the graph between training Loss and validation Loss, training Accuracy and validation Accuracy.

5. EVALUATING THE MODEL AND PREDICTING THE MODEL

The trained model is used to predict the various classes of images and validated and the prediction results to be 98%.

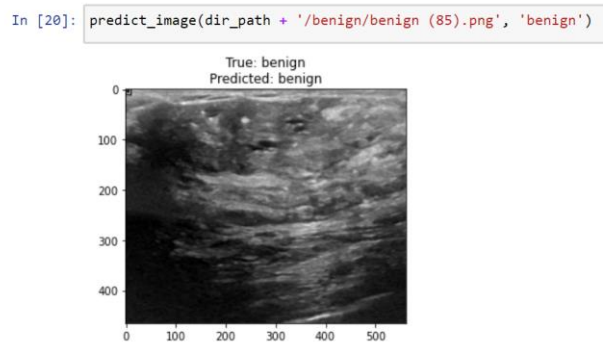


Figure 8.1) Benign Test

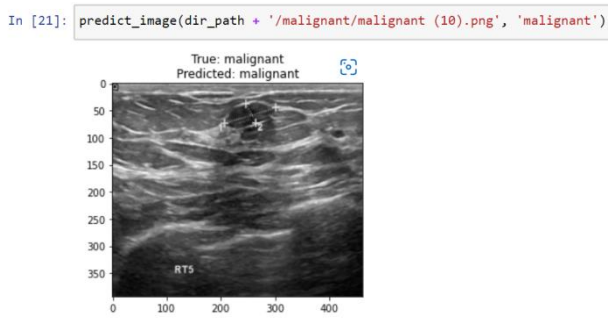


Figure 8.2) Malignant Test

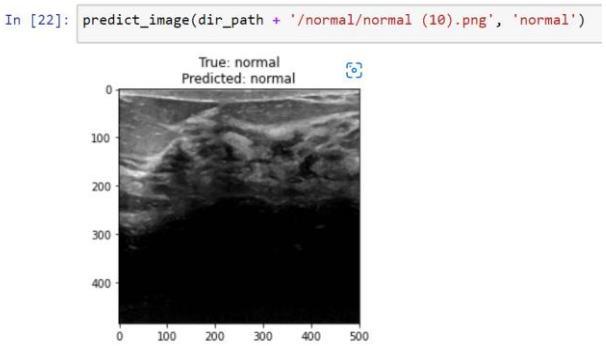


Figure 8.3) Normal Test

V.CONCLUSION

The project uses a deep learning technique convolutional neural system in that, a latest model of EfficientNet B7 which is trained of using the image datasets and another model of Artificial neural Network that is trained using the Wisconsin dataset. Both the models can be used for classification of breast cancer. The given datasets is properly pre-processed and used for training and testing the model.

The trained model can be used in the hospitals for diagnosing the patients whether they are affected with breast cancer or not. If affected what is the stage of the tumour either benign or Malignant, which will be helpful for the early diagnosis and could save lives of the people.

VI.REFERENCES

[1] Moh'd Rasoul Al-hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahnah Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm 2016 9th International Conference on Developments in eSystems Engineering.

[2] Shajib Ghosh, Jubaer Hossain , Dr. Shaikh Anowarul Fattah, Dr. Celia Shahnaz , Asir Intisar Khan Efficient Approaches for Accuracy Improvement of Breast Cancer Classification Using Wisconsin Database 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh.

[3] Naresh Khuriwal , Dr Nidhi Mishra **Breast Cancer Diagnosis Using Deep Learning Algorithm** International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018).

[4] Nur Syahmi Ismail , Cheab Sovuthy Breast Cancer Detection Based on Deep Learning Technique *Universiti Teknologi PETRONAS* Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia.

[5] Ramik Rawal , BREAST CANCER PREDICTION USING MACHINE LEARNING 2020 JETIR May 2020, Volume 7, Issue 5.