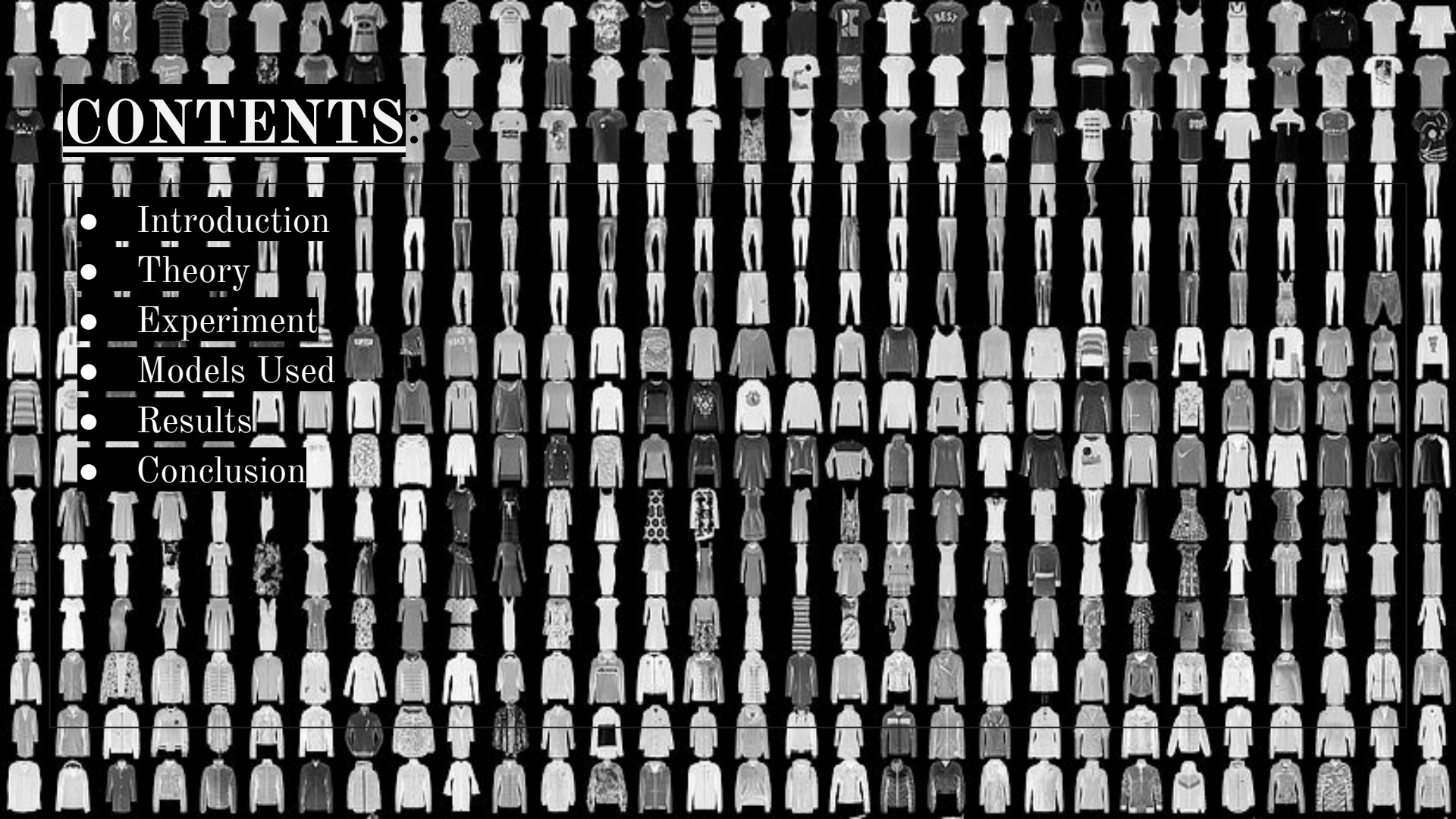
DETECTION OF INVASIVE DUCTAL CARCINOMA FROM BREAST HISTOPATHOLOGY IMAGES USING DEEP ENSEMBLE NEURAL NETWORKS

by

Shreya Sahay (1704219)
Richik Ghosh (1704285)
Kaushan Manna(1704271)

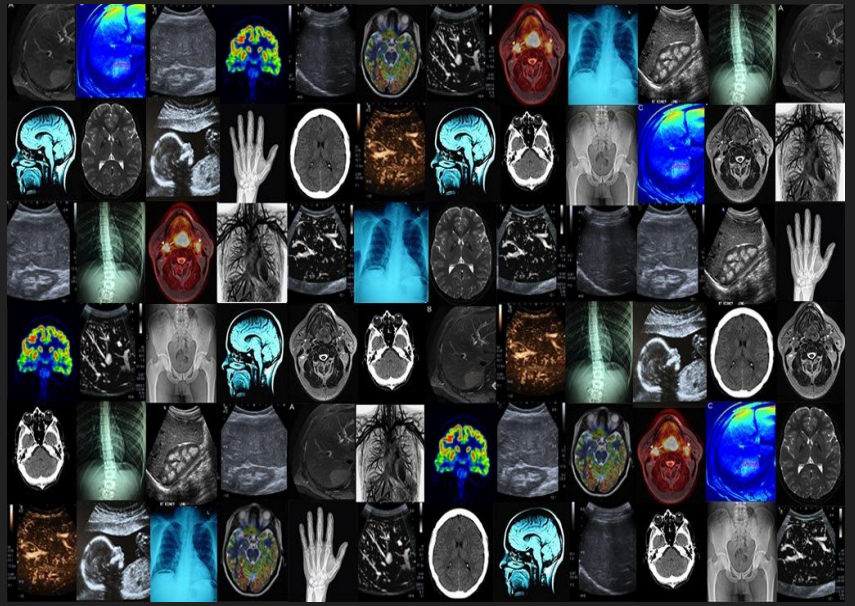


CONTENTS:

- Introduction
- Theory
- Experiment
- Models Used
- Results
- Conclusion

INTRODUCTION

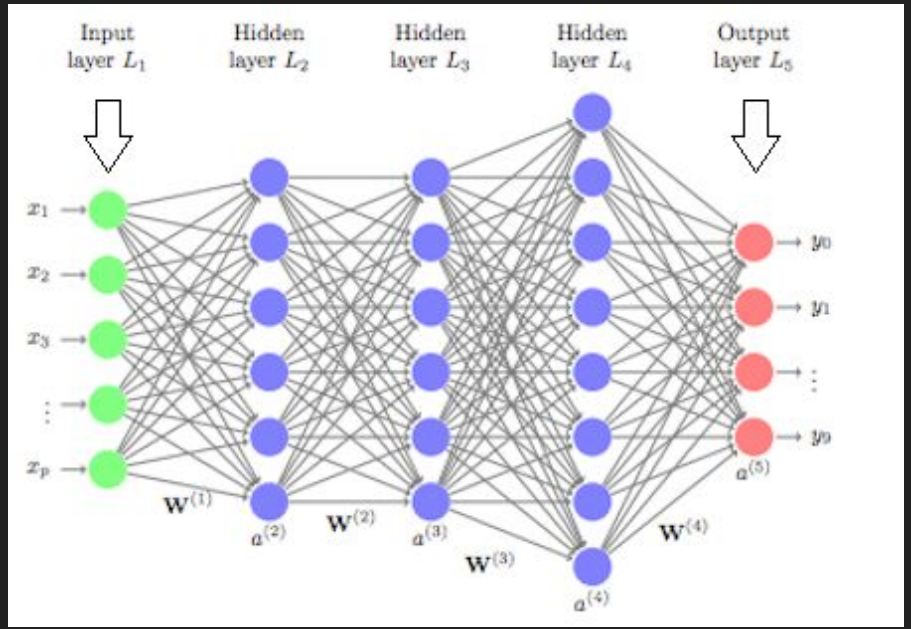
Medical imaging is the technique and process of creating visual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues (physiology). Medical imaging seeks to reveal internal structures hidden by the skin and bones, as well as to diagnose and treat disease. Medical imaging also establishes a database of normal anatomy and physiology to make it possible to identify abnormalities. Although imaging of removed organs and tissues can be performed for medical reasons, such procedures are usually considered part of pathology instead of medical imaging.



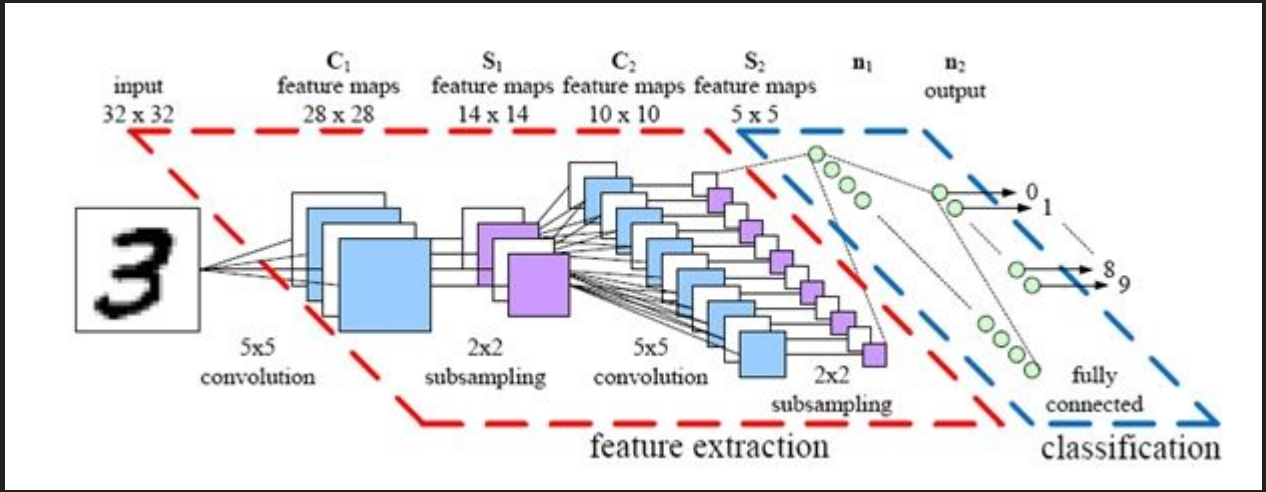
THEORY

A **deep neural network (DNN)** is an artificial neural network (ANN) with multiple layers between the input and output layers.There are different types of neural networks but they always consist of the same components: neurons, synapses, weights, biases, and functions.DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is expressed as a layered composition of primitives.·The extra layers enable composition of features from lower layers, potentially modeling complex data with fewer units than a similarly performing shallow network.

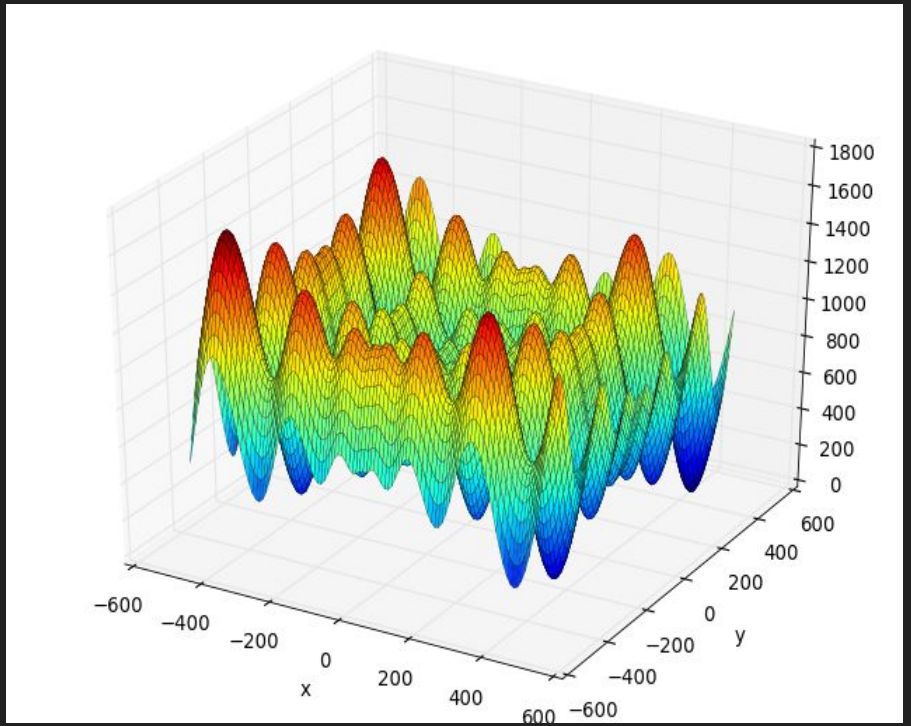
Deep architectures include many variants of a few basic approaches. Each architecture has found success in specific domains. It is not always possible to compare the performance of multiple architectures, unless they have been evaluated on the same data sets.



In convolutional neural networks image data is used to make classification models. As CNNs have certain feature extraction mechanism it can be used to enhance features in the image samples. These feature enhanced pixel values are then fed to multiple arbitrary number of fully connected layers and passed onto a softmax activation layer to make classification.

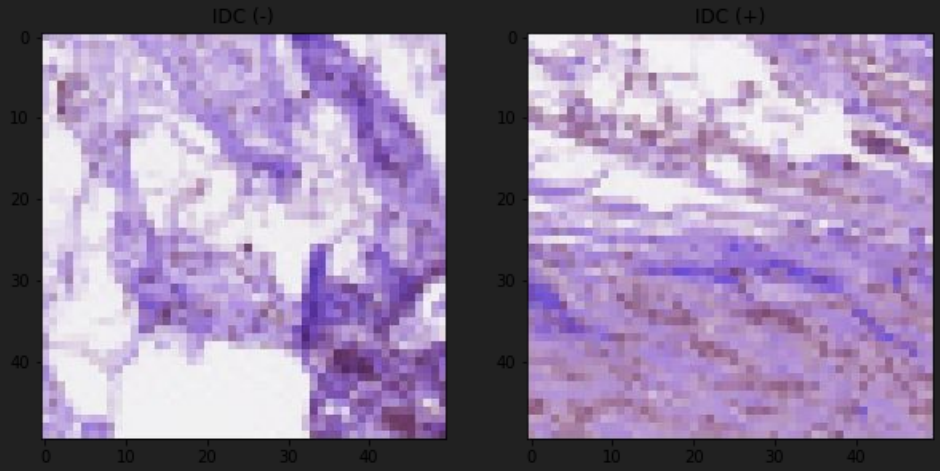


The predicted pixel values are subtracted from real values in the test set. This subtracted values of all the instances are then plotted against each other. This is known as the loss function plot. Now the optimization of this loss function means minimizing the loss function values. The lower the better. Now reaching the global minima is the main goal here, avoiding the local minimas. Stucking in these local minimas can give corrupted results.



EXPERIMENT

The data initially consisted of **277,524 patches**, with an individual class count of **198,738 Non-IDC** and **78,786 IDC patches**. The dataset is very refined with high feature-containing patches that were extracted to make best-suited features for individual model training. No features were changed or color channels were altered to prevent damaging of feature maps. The dataset was reduced to 160,000 images and was dissected into training, test, and validation directories, each of size **80%, 10%, and 10%** of the original respectively.



IDC positive samples tend to have a denser **colour intensity** as compared to IDC negative samples, according to their respective graphs. This feature is used by the models in differentiating between each sample.

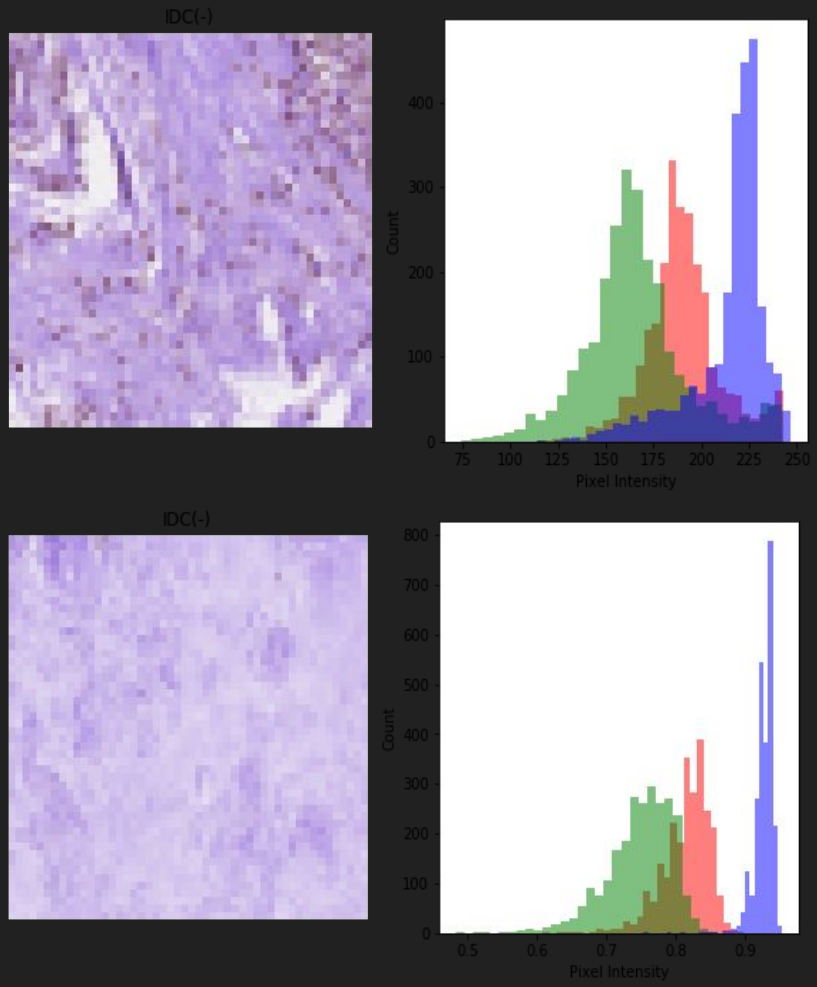
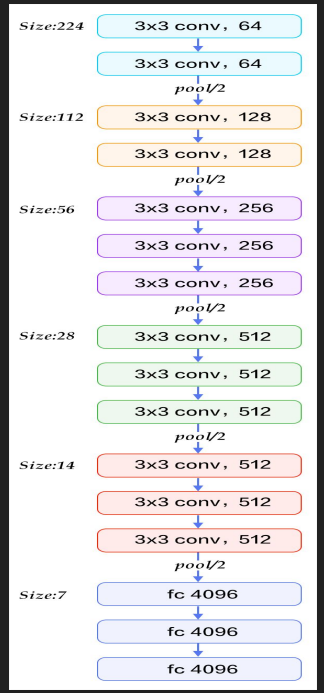


FIG: IDC positive sample (TOP) with its colour graph and IDC negative sample with its colour graph.

MODELS USED



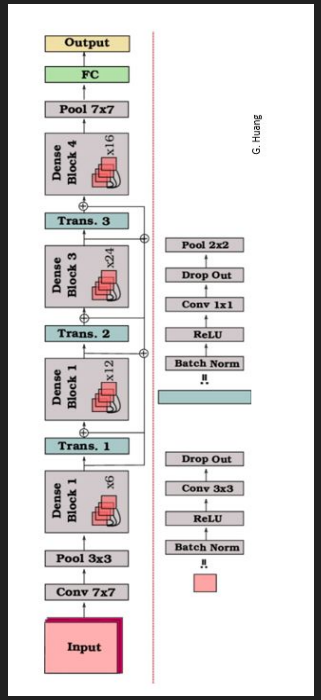
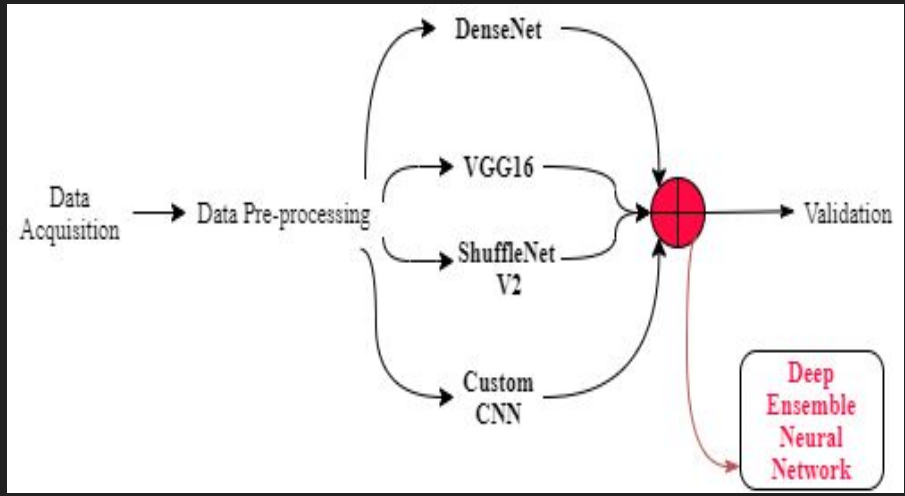


FIG: The three models used in our experiment : **VGG16, DenseNet and Inception V2.**

All the models are now averaged together after each and every one of the models are fitted on the data. The best results are acquired by Ensemble learning, as it uses the above technique of merging the results of transfer learning models.



RESULTS

Here the accuracy and precision scores have been determined to be near about **95%.** Along with which recall scores are also high and which is a good indicator of the model as it depicts the number of false negatives which is less, implying that our model is performing well on the above labels.

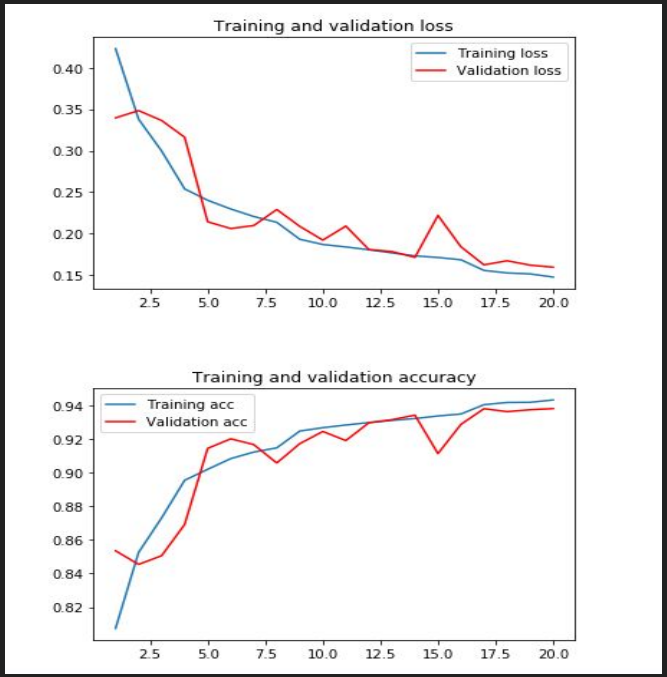


FIG: Performance graphs

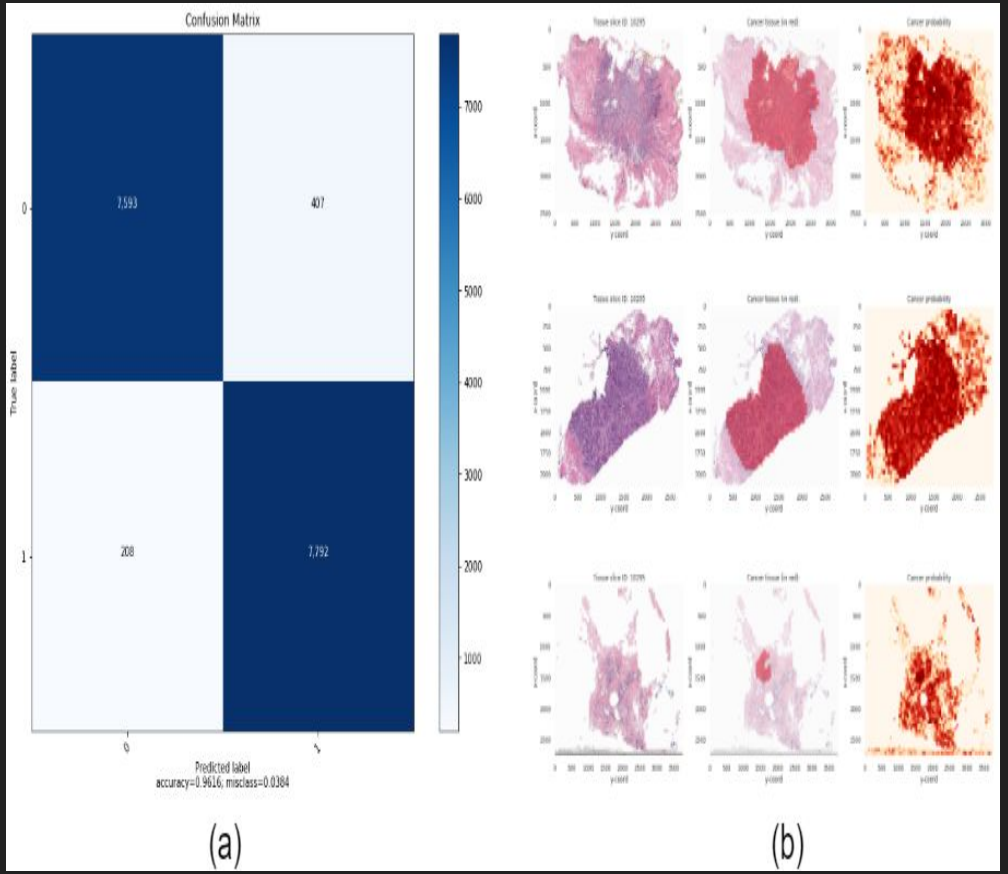
The model-specific scores in terms of Sensitivity, Precision and F1-score are listed in the adjoining table. The overall sensitivity and precision were valued to be 0.973 and 0.949 respectively. Using sensitivity and precision scores, F1-score was determined to be 0.961

Table 2: Performance Evaluation of individual architectures

Metrics Used	DenseNet	Custom CNN	VGG16	Inception V2
Accuracy	0.93	0.943	0.95	0.95
Precision	0.94	0.95	0.95	0.95
Recall	0.92	0.94	0.94	0.96
F1-Score	0.93	0.94	0.95	0.95

The ensemble model is accurately able to classify 15,385 histopathological samples out of a total of 16,000 samples as depicted in the confusion matrix in fig. (a). 615 samples out of a total of 16,000 patches were mismeasured or misclassified.

Fig.(b) shows the probability of the zone being affected by cancer in varying colour gradients of red (light red for low and dark red for high probability). These colour based visualisations that have been adopted and implemented could further emphasize the performance of the model in terms of predicting the cancer affected areas. The ensemble framework achieves a balanced accuracy of 96.2%.



CONCLUSION

In this project the detection of invasive ductal carcinoma from breast histopathology tissue sample images was successfully carried out using deep artificial neural networks with a resultant accuracy of ~95%. Three high performance DNN models were employed for the same. The power of each of the four individual networks are combined together into an ensemble framework which achieved a detection accuracy of 96.2%.

The further scope that can be pondered upon is the use of **deep ensemble neural networks** for even more improved performance in terms of detection accuracy.