Breast Cancer Detection - A Comparative Study of different kinds of Convolutional Deep Learning Models

Anonymous CVPR submission

Paper ID *****

Abstract

Among the various types of cancer, breast cancer is one of the most common types of cancer, and for this many people lose their lives. Various impactful advancements have been done in classifying lethal cells with Artificial Intelligence(AI) more specifically, computer vision technologies over the last decade. Implementing these Machine Learning (ML) algorithms has made the breast cancer detection process convenient. In this study, we have used various deep-learning techniques to detect breast cancer such as LeNet, VGG16, inceptionV3, Xception, and Transformer. We showed experiments by changing 3 important hyperparameters which are batch size, learning rate, and optimizers, and found how they performed with respect to various models. After implementing the models, we found that the highest accuracy 98% was achieved by the MobileNet model. Apart from that Xception model also performed well by acquiring 98% when the optimizer is "Adam" and the Learning rate is.001.

1. Introduction

Breast Cancer is nowadays a very egregious disease among both men and women. According to CDC (Centers for Disease Control and Prevention), there are about 2,64000 cases observed. Each year about 42000 women and 500 men in the United States lost their lives [10]. The dataset we are using for this project is basically the images of histopathological reports, which contains the images of specimen containing cancerous elements. It is generally efficient to analyze the malignant and benign cancer cells of patients who are affected by Cancerous tumors. Generally, there are three types of cells, that are categorized as Invasive Ductal Carcinoma (IDC), "Ductal Carcinoma", and "Invasive Lobular Carcinoma" [7]. But the fatal cell that is the significant cause of breast cancer is IDC.

Patients diagnosed with this IDC type are mostly in critical condition and the number of affected people is not small [1]. For detecting these types of cells there are various techniques among them histopathological images are commonly used. The classification output is generated to assist the medical personnel in understanding the situations of patients who have malignant tissues that can cause great loss in their lives. Although there are some cons to these histopathological images sometimes the resolution of the images is not good enough [14]. But for analyzing breast cancer's malignant or benign, its impact is inevitable. That's why the importance of early prediction of breast cancer plays a vital role.

For image classification, there are various convolutional methods in Machine Learning (ML). We implemented the basic Convolutional Neural Network Model(CNN) [11], Visual Geometry Group(VGG) model [9], Inception [13], Xception [2], MobileNet and Transformer model. We analyzed the impact of changes in train and test size, by changing the parameters. We used the traditional CNN model as our baseline model and compared it with other models. Vision Transformer(ViT) is a very recent development that has been done in computer vision research. It uses pre-trained and fine-tuned datasets to analyze a new model.

In this report, we showed variations of the models on different aspects. For example, we examined the effect of the amount of data on the overall accuracy of a specific model. Also, we changed the number of epochs and other parameters and hyperparameters such as learning rate, batch size, and optimizer of the model.

In the project proposal, we mentioned that the dataset would be UCI Machine Learning Repository. But as there were no images in that dataset and all of our applied models would be deep learning models, we have changed our dataset to Breast Histopathology Images from UCI Machine Learning Repository. As the dataset was changed, we changed the idea to implement the general regression models and applied image processing on ML. Here we processed the dataset and on this dataset applied various Vision methods and among them analyzed different models and showed the model performed better among these models.

2. Related Works

2.1. LeNet:

LeNet (Large Fully Connected Multi-Layer Neural Network) is first introduced in a paper by Yann LeCun. It has 7 layers where each cell relates to a 5*5 layer and is also connected to 16 feature graphs that are also connected to the previous layer. It takes 32*32 images as input and produces a 28 * 28 * 6 feature map as output. Yann LeCun et el. tried different versions of LeNet on a dataset of handwritten digits and found success [8].

2.2. VGG16:

VGG 16 is a deep convolutional neural network with 16 layers, and the input size is 224*224*3. In VGG16 architecture, the structure of the model consists of convolution layers of size 3*3 filters with a stride size of 1. The padding size is the same all the time. The max pool operation is applied using a 2*2 filter with a stride of size 2. Lastly, it has two fully connected layers, and to predict the label it uses the Softmax function. Karen Simonyan and Andrew Zisserman applied very deep convolutional neural networks for large-scale image classification and VGG with 2 nets, multi-crop, and dense evaluation came out as the best model in their case [12].

2.3. InceptionV3:

The InceptionV3 model will be applied to the dataset. It is 48 layers deep convolutional neural network. We can load the pretrained version of this model trained on the ImageNet dataset. The input size of the images for this network is 299*299*3. It is an image recognition model which has an accuracy rate of 78.1% applied to the ImageNet dataset. There are three types of building blocks in this model. 1*1, 3*3, and 7*7 convolutions are applied to build this model [13].

2.4. Xception:

Xception Model is a 71 layers-deep convolutional neural network model. This model also takes inputs of images of size 299*299*3 [2]. There are three different sections of this model – entry flow, middle flow, and exit flow. This model is built upon the knowledge of the Inception model. The name Xception comes from the abbreviation – The Extreme Inception. The architecture of the model is based on the depth-wise separable convolution layers. That is Xception is an extension of the Inception mode where the modules are replaced by the depth-wise separable convolutions. For a very high number of classes of 17000, Francois Chollet devised a new convolutional network based on Inception and named the model as Xception (stands for Extreme Inception), for a large image dataset, and his model was proved to be effective [2].

2.5. MobileNet:

MobileNet model also has depth-wise separable convolution. The input of the MobileNet model has 224*224*3 dimensions. The depth-wise separable convolution layer has two distinct parts – depth-wise convolution and point-wise convolution. MobileNet is built for mobile devices and embedded vision applications as it has less number of parameters compared to other deep convolutional neural network models. It is also faster to run compared to other deep convolutional models. Previous research demonstrated its effectiveness for different types of image tasks including face detection, object detection, fine-grained recognition of dogs, etc. [6].

2.6. Vision Transformer:

In their paper published in 2022, Behnaz Gheflati and Hassan Rivaz showed the usage of Vision Transformers (ViT) to classify breast US (ultrasound) images for the first time1. To address the imbalanced dataset issue, they used a weighted cross-entropy loss function. The researchers applied a linear classifier instead of an MLP head in ViT model. They found that the model was equivalent to or better than SOTA (state-of-the-art) CNNs in terms of classification accuracy and AUC [4].

Z. He et el. successfully implemented a novel Deconv-Transformer (DecT) network model by incorporating color deconvolution in the form of convolutional layers2. They applied the model to 3 different datasets such as UC, BreakHis, BACH to check the reliability of their work. They fused both RGB and HED color space information in DecT model. Result showed that DecT model was more robust to color differences images than the other models such as RGB-ViT, HED-ViT, DecT-HED, DecT-conv [5].

3. Dataset

To implement the models, we will use the Breast Cancer Histopathology Dataset. The dataset is a collection of total of 277,524 images. Among these images, 198,738 images IDC are negative and 78,786 images are IDC positive. The size of each image is 50*50 pixels. The images are derived from 162 H&E-stained breast histopathology samples. Usually, the breast tissue contains many cells but only a few of them are prone to cancer. In this dataset, the images labeled with ”1” indicate the characteristics of invasive ductal carcinoma.

4. Method

We will explain our project for detecting and classifying Breast Cancer Histology images. The total projects consist of three different phases-

4.1 Input image processing

4.2 Training and testing the model

4.1. Input processing

Before we feed the dataset into the model as input, we need to preprocess the dataset. For the training purpose, we took a small amount of image data from the dataset. For most of the model, 1500 images are used to train the dataset. It is because, in reality, getting the labeled medical images are not easy. In many practical scenarios, it is often seen that large data are available but small fractions of it are labeled. These types of datasets require a higher level of human expertise to manually label the data *e.g.* X-ray images, MRI images, etc. Also, these labeling processes are time-consuming and expensive. In these cases, the semi-supervised approach addresses these problems with more economically viable methods. So, we need to maximize the label gain from a low-sampled or partially sampled dataset using these methods.

Deep learning algorithms are gaining a lot of interest in recent times for various tasks involving classification, segmentation, detection, and recognition of certain patterns in different fields. But, to get better results in terms of accuracy, deep learning algorithms usually need a lot of data. Without the availability of large labeled datasets, it would be difficult to achieve better results. Also, creating a large labeled dataset is expensive and sometimes it is not possible to gather such an amount of data for a specific task *e.g.* medical, finance, and government sectors data where privacy and security of real data is a prime concern.

We have resized our input images according to the model input structure. We took 1500 images to train the models. But before the training procedure, we normalized the pixel values of the images

4.2. Training and Testing the Models

We have used different variations of Deep Learning models o classify. We have used different variations. First, we run the models keeping a learning rate of 0.001 with a batch size of 32. Then we vary the learning rate by 0.01 and 0.1 keeping the batch size the same. Then we measured the best learning rate in terms of accuracy. Taking the best learning rate then we vary the batch sizes. As already a train is completed with batch size 32, we again vary the batch size to 16 and then 64. Then taking the best learning rate and batch size we vary the optimizer into two different categories, one Adam optimizer and the other is RMSProp optimizer.

For all the models we run up to 20 epochs to get the final training and validation accuracy.

The steps are shown in Figure 1.

Deep Convolutional Models - We will apply several deep convolutional models on this dataset with certain variations mentioned in the 'Introduction' section. The framework of the models are given below-

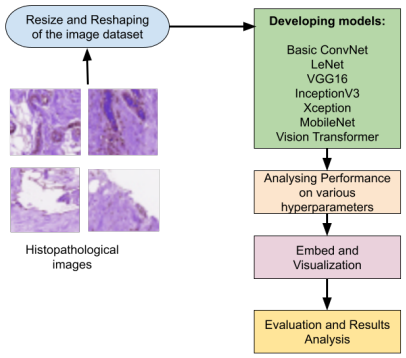

Figure 1. The model of the overall system architecture explaining our workflow for breast cancer classification

(i) **Basic ConvNet** - First we trained with very basic CNN model. That consists of just one convolutional layer and two dense layer. From these two dense layers, one layer is the output layer. As we classifying the images either in cancerous or non-cancerous cell, so our classification is a binary classification. Hence, sigmoid activation function is used instead of softmax function in the output dense layer. It has total 7,684,398 trainable parameters. The structure of the basic CNN model is depicted in Figure 2.

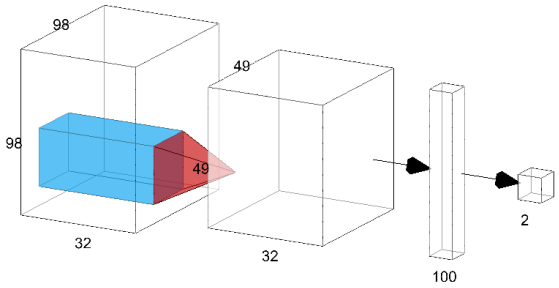

Figure 2. Basic Convolutional Neural Network Model Architecture

(ii) **LeNet** [8] - LeNet is more sophisticated model compared to basic CNN model. It has three convolutional layers and two dense layers. Between these two dense layers, one layer is the output layer. In LeNet, there are 3,690,126 trainable parameters. Also average pooling operation is applied in LeNet. The structure of the

LeNet is depicted in Figure 3.

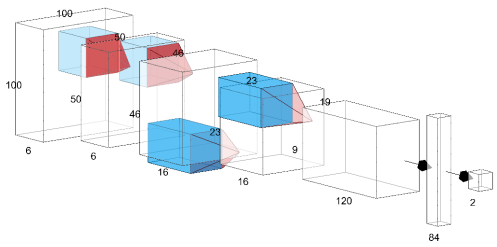

Figure 3. LeNet Model Architecture

(iii) **VGG16** [9] - VGG16 has 14 convolutional layers and three dense layers among them one layer is the output layer. Apart from the output layer, this model has 16 layers. Also, there are five max pooling operations. In the output layer, we keep 2 units in the dense layer as our task is a binary classification of images. The structure of the VGG16 model is depicted in Figure 4.

(iv) **InceptionV3** [13] - Inception is a very deep convoutional neural network network. It takes the input as 299*299*3 dimension. Inception model factorized convolution with large filter size. Three different types of convolution was used to build up InceptionV3 model, *i.e.* 1*1, 3*3, 5*5 convolution. Also, auxiliary classifier is used to deal with dropout. The InceptionV3 model is depicted in Figure 5. It has 23,885,392 trainable parameters and 311 number of layers.

(v) **Xception** [2] - Xception model used depthwise convolution to deal with classification. It has three flows - Entry flow, Middle flow, and exit flow. It has 22,855,952 trainable parameters in total and 71 layers in total. The structure of Xception model depicted in Figure 6.

The MobileNet model has only 13 million parameters with the usual 3 million for the body and 10 million for the final layer and 0.58 Million mult-adds

(vi) **MobileNet** [6] - MobileNet model has approximately 13 Million trainable parameters, hence its easy to run on mobile devices. It has 28 layers. The input size of the MobileNet is 224*224*3. The architecture of the MobileNet is depicted in Figure 7.

(vii) **Vision Transformer** [3] - Vision Transformer revolutionize the classification. It consists of transformer encoder that includes Multi Head Attention layer which concatenates all attention outputs linearly. Also the Multi-layer perception layer contains the Gaussian Linear Unit (GeLU). Lastly, it includes Layer Norm, which is added to improve training time and overall performance.

In transformer architecture, residual connections are included to flow through the network without passing through no-linear activation.

The steps of vision transformer is as follows-

(i) Split the input image into patches or fixed sizes.

(ii) Flatten the image patches.

(iii) Generate linear embedding from the flatten image patches.

(iv) Create positional embedding.

(v) Feed the sequence as input to transformer encoder.

(vi) Pretrain the Vision Transformer with image labels.

(vii) Tune for image classification.

The structure and procedure of Vision Transformer is depicted in Figure 8.

Comparison - Based upon the results for the abovementioned model, we will compare the result in terms of accuracy and loss for different dataset sizes.

5. Results

We have trained and tested our model on the Breast Cancer Histopathology image dataset with different parameter variations. To train the model we have used a small portion of the images. In all the cases, we have taken 1500 images for training and testing split. The training and testing split portions are divided into 80% and 20% of selected images. We used the random seed option to get the same portions of images on a different run. Also to verify the parameters of learning rates, batch sizes, and optimizes we always restarted the kernel in the Kaggle notebook so that the previously trained model won't affect the later versions of the results.

We run our model on the Kaggle notebook. Kaggle provided us with glob technique to fetch the dataset efficiently and we also used GPU P100 to accelerate our running time. We run all the variations up to 20 epochs to come up with the results. To draw the model, we have used a web-based tool.

The results of the models in terms of validation accuracy are shown in Table 1.

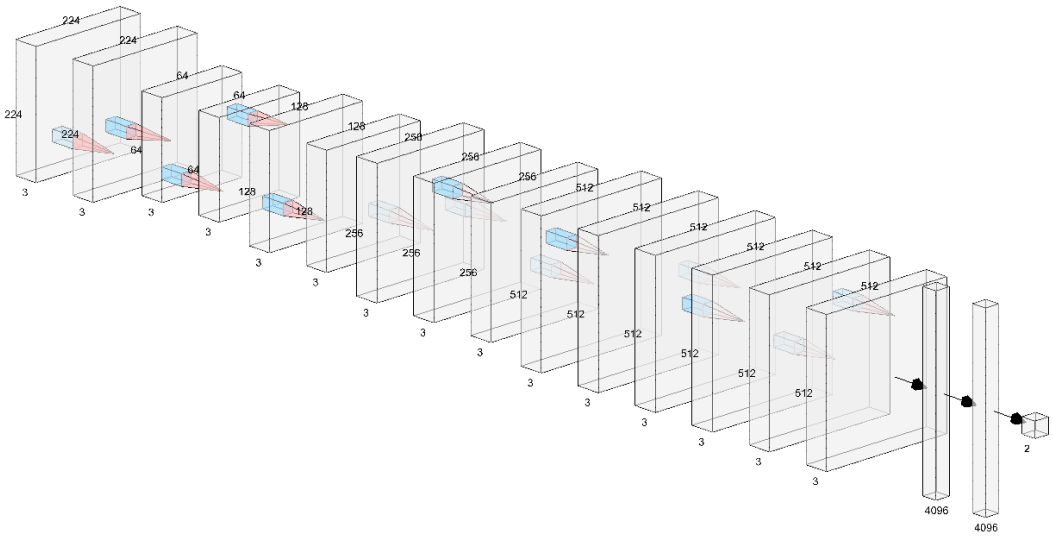

Figure 4. The structure of the VGG16 model. [9]

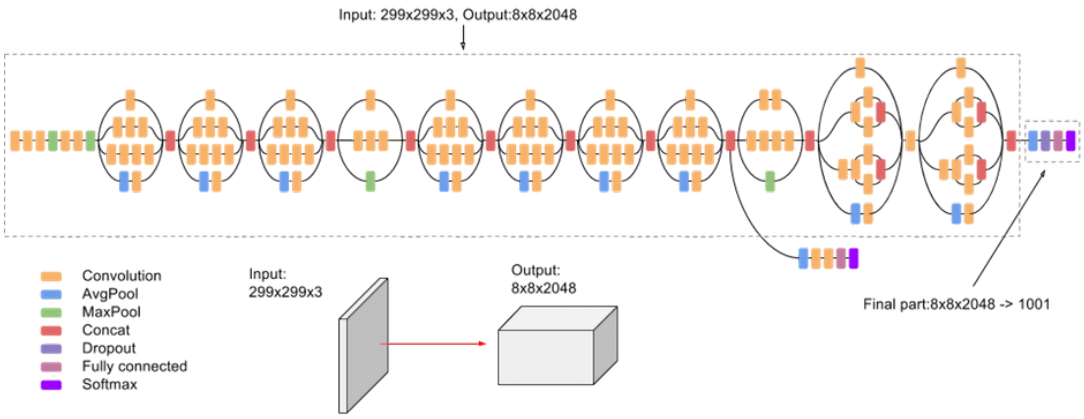

Figure 5. The structure of the InceptionV3 model. [13]

Table 1. Performance Evaluation for Models

Models	Learning Rate Variations			Batch Size Variations			Optimizers Variations	
	0.001	**0.01**	**0.1**	**16**	**32**	**64**	**Adam**	**RMSProp**
Basic CNN	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.92
LeNet	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.95
VGG16	0.89	0.87	0.87	0.88	0.88	0.89	0.88	0.89
InceptionV3	0.94	0.94	0.94	0.95	0.95	0.95	0.95	0.95
Xception	0.98	0.97	0.97	0.97	0.98	0.98	0.98	0.97
MobileNet	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Vision Transformer	0.62	0.63	0.63	0.64	0.63	0.63	0.62	0.63

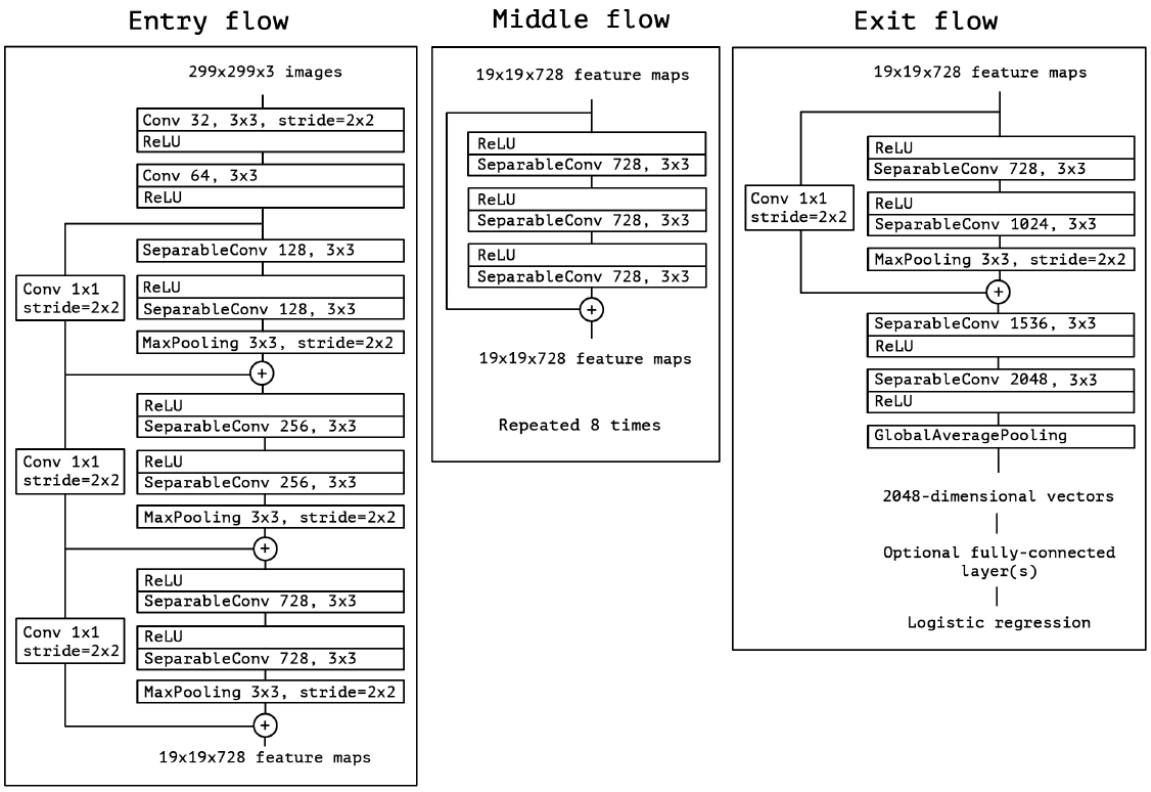

Figure 6. The structure of the Xception model. [2]

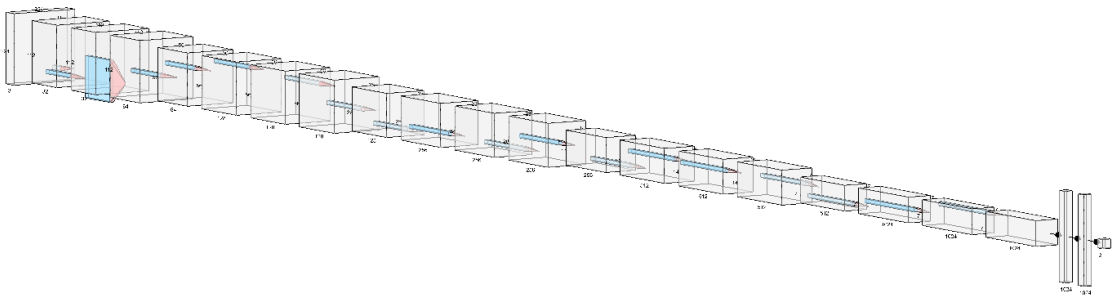

Figure 7. The structure of the MobileNet model. [6]

6. Conclusion

In this research, we applied 7 different deep learning models including a basic CNN and transformer to detect breast cancer. We fine-tuned our model by changing the hyperparameters such as learning rate, batch size, and optimizers. We have found that the MobileNet outperformed the other models in terms of prediction accuracy and other metrics. However, there are more ways to extend this work and explore this problem in the future. To start with, image embedding can be done before training the final model. Then it will be easier to build models with large datasets. Secondly, data augmentation methods can be applied. It will help to manage the overfitting of the model. Thirdly, we have tried 3 different learning rates and batch sizes. Exploring more learning rates and batch sizes might help to get better accuracy. In the future, this will be an interesting area to explore.

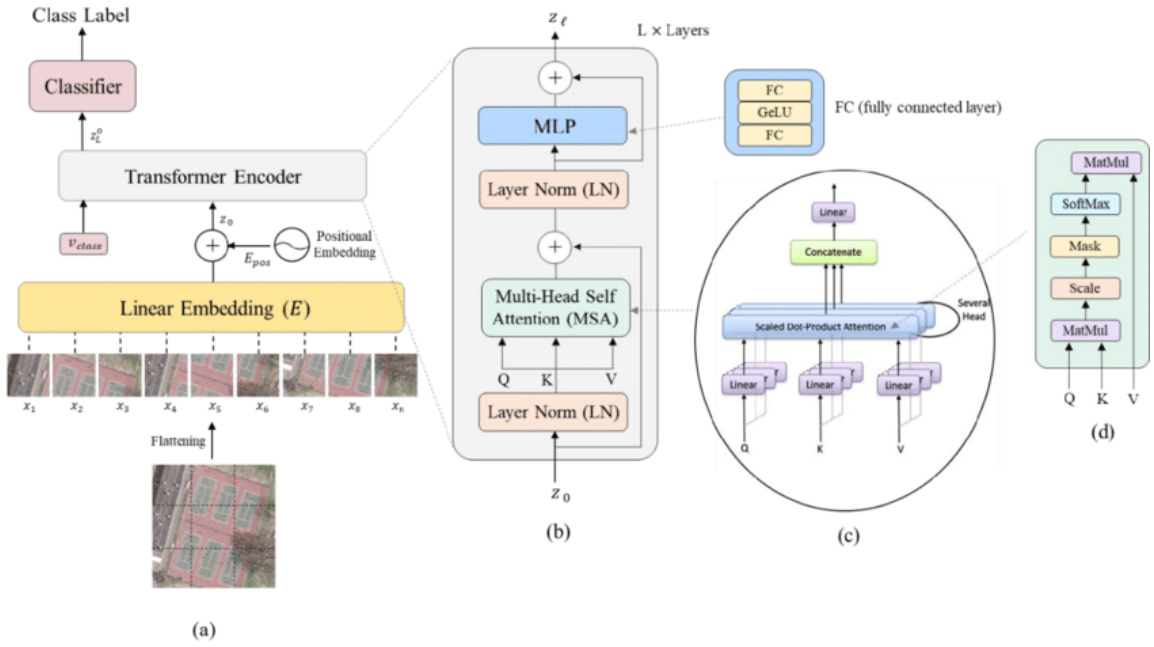

Figure 8. The structure of the Vision Transformer model. [3]

7. Division of Labor

Task distribution among group members for this project is mentioned below- The division of labor to build models:

Basic CNN: Rafae Abdullah - Code recreated from Keras Documentation.

LeNet: Rafae Abdullah - Code recreated from Keras Documentation.

VGG16: S M Rafiuddin - Code recreated from Keras Documentation.

InceptionV3: S M Rafiuddin - Code recreated from Keras Documentation.

Xception: S M Rafiuddin - Code recreated from Keras Documentation.

MobileNet: Farzana Islam Adiba - Code recreated from Keras Documentation.

Transformer: Farzana Islam Adiba - Code recreated from Keras Documentation.

The division of labor to write the report:

Abstract: Farzana Islam Adiba

Introduction: Farzana Islam Adiba

Related Works: Rafae Abdullah

Dataset: S M Rafiuddin

Methods: S M Rafiuddin

Results: S M Rafiuddin

Conclusion: Rafae Abdullah

References: Farzana Islam Adiba, Rafae Abdullah, S M Rafiuddin

References

[1] Saad Awadh Alanazi, M. M. Kamruzzaman, Md Nazirul Sarker, Madallah Alruwaili, Yousef Alhwaiti, Nasser Alshammari, and Hameed Muhammad Siddiqi. Boosting breast cancer detection using convolutional neural network. *Journal of Healthcare Engineering*, 2021:11, 2021. 1

[2] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017. 1, 2, 4, 6

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Trans-

formers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 4, 7

[4] Behnaz Gheflati and Hassan Rivaz. Vision transformers for classification of breast ultrasound images. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 480–483. IEEE, 2022. 2

[5] Zhu He, Mingwei Lin, Zeshui Xu, Zhiqiang Yao, Hong Chen, Adi Alhudhaif, and Fayadh Alenezi. Deconv-transformer (dect): A histopathological image classification model for breast cancer based on color deconvolution and transformer architecture. *Information Sciences*, 608:1093–1112, 2022. 2

[6] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 2, 4, 6

[7] Abhinav Kumar, Sonal Saxena, Sameer Shrivastava, Vandana Bharti, and Sanjay Kumar Singh. Chapter 20 - internet of things and other emerging technologies in digital pathology. In Sanjay Kumar Singh, Ravi Shankar Singh, Anil Kumar Pandey, Sandeep S. Udmale, and Ankit Chaudhary, editors, *IoT-Based Data Analytics for the Healthcare Industry*, Intelligent Data-Centric Systems, pages 301–312. Academic Press, 2021. 1

[8] Yann LeCun, Larry Jackel, Leon Bottou, A Brunot, Corinna Cortes, John Denker, Harris Drucker, Isabelle Guyon, UA Muller, Eduard Sackinger, et al. Comparison of learning algorithms for handwritten digit recognition. In *International conference on artificial neural networks*, volume 60, pages 53–60. Perth, Australia, 1995. 2, 3

[9] Sidratul Montaha, Sami Azam, Abul Kalam Muhammad Rakibul Haque Rafid, Pronab Ghosh, Md. Zahid Hasan, Mirjam Jonkman, and Friso De Boer. Breastnet18: A high accuracy fine-tuned vgg16 model evaluated using ablation study for diagnosing breast cancer from enhanced mammography images. *Biology*, 10(12), 2021. 1, 4, 5

[10] Division of Cancer Prevention, Centers for Disease Control Control, and Prevention. Basic information about breast cancer, 2022. 1

[11] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015. 1

[12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2

[13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 1, 2, 4, 5

[14] Nusrat Mohi ud din, Rayees Ahmad Dar, Muzafar Rasool, and Assif Assad. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in Biology and Medicine*, 149:106073, 2022. 1