

ZEWAIL CITY OF SCIENCE AND TECHNOLOGY

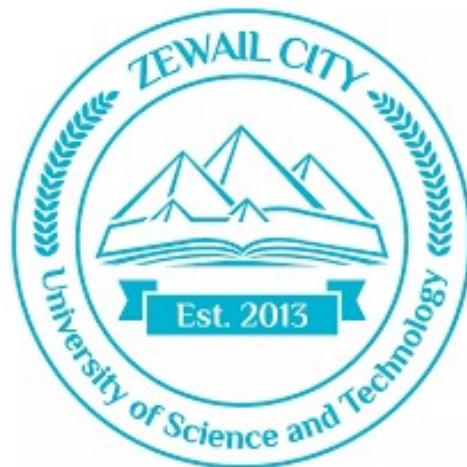
DEEP LEARNING
CIE 555

Invasive Ductal cancers (IDC), Diagnosed with Conventional Neural Networks (CNN)

Ahmed Adel 201901464

Omar El-Sakka 201900773

June 23, 2022



Contents

1 Abstract	2
2 Introduction	2
3 Objectives	3
4 Transfer learning	3
4.1 EfficientNet family	3
5 Methods	4
5.1 WSI and H&E	4
5.2 Invasive ductal carcinoma(IDC)	4
6 Dataset description of IDC dataset	5
7 Literature review	6
7.1 J. Zhang, X. Guo, B. Wang and W. Cui, "Automatic Detection of Invasive Ductal Carcinoma Based on the Fusion of Multi-Scale Residual Convolutional Neural Network and SVM [1]	6
7.2 Ranjan, R <i>et al</i> , Comparative assessment of CNN architectures for classification of breast FNAC images [2]	9
7.3 F.A. Spanhol, L.S. Oliveira, P.R. Cavalin, C. Petitjean, L. Heutte Deep features for breast cancer histopathological image classification	10
7.4 Kaggle competition code	11
8 Proposed models	11
8.1 IDC	11
9 Results of proposed models	13
9.1 IDC	13
10 Fine tuning some models	16
11 Using SVM as classifier	16
12 Deployment using MATLAB	17
13 Ethical considerations	19
13.1 Data Privacy	19
13.2 Ensure AI transparency	19
13.3 Create opportunities for employees	19
14 Time line	19
15 Team members contribution	20
16 Conclusion	20
17 References	20

1 Abstract

Invasive Ductal carcinoma(IDC) is the most popular cancer type in breast, and one of the most killing causes, however, if it is being diagnosed at early stages, there is high potential of survival. Though, IDC has been studied for decades, and pathologists have been trained to detect it from images, their accuracy is still relatively poor; on average it's 75%. In this report We try to provide the society with some new trained models to help them deciding the malignancy state of the patient, and overcome the accuracy of pathologists in classifying the cancer.

The challenging part in IDC, is that it has no clear symptoms; Consequently women don't recognize the cancer until very late stage of the cancer, where survival rate is very poor. In this report we will present some new models; that we have trained on IDC dataset. We will provide Literature review for IDC to see where the medical society stopped at, and deploy our models in MATLAB project, followed by discussing some ethical considerations relevant to our work.

2 Introduction

Breast Cancer is one of the most common causes of deaths right now, it is a type of cancer that may start in just one breast or both breasts. Although benign breast tumors increase abnormally in breasts, tumors will not spread out, and will not affect their neighbours. So, they are not very dangerous. Malignant tumor is much more dangerous; and may result in infecting the neighbouring cells, moreover; it may return back to the patient after being cured. There are some types of benign breast lumps that may increase the probability of getting a malignant one if not diagnosed at early stage.

In 2020, It was estimated that 2, 261, 419 new cases were diagnosed in women across the world with breast cancer, and 684, 996 women all over the world died due to it in that year. If those cases have been detected early, the number of death could have been decreased.

Classifying the tumor type is a bit challenging for pathologists, as the scanned image varies a lot from one patient to another, moreover, from one tissue to another; As the scanned image has many factors that might affect doctor's decision. like tissue thickness, staining protocol,..etc.

Radiogenomics is one of the emerging fields, that states how we can study the human at cell resolution with the aid of only digital pictures. Pictures can be gotten in vivo way, in other words, noninvasive, Consequently there is no pain or fear. The main challenge exist in how to extract some features from just a normal image and to predict cell activity, or cancer.

The overall sensitivity of doctors at clinical diagnosis for breast cancer is 50.5%, and average accuracy is 75%. So there was a demand to use another more complicated tool to represent such a higher dimensional problem. Convolution neural network(CNN) is one of the most popular architectures provided firstly by Yann Lecun, to extract features from images. CNN is very powerful, it can understand very deep complex features that naked human eye can never detect. Meanwhile, it's also a disadvantage, as doctors can't understand details of the decision making process.

We tend in this project to use different pretrained CNN models in transfer learning to predict whether the tissue suffers from IDC or not. We will compare the result and show the best one and deploy them. Our main objective is to get an accuracy higher than usual doctors.

3 Objectives

Our object was to get a good model in terms of Accuracy, Specificity, Sensitivity, Precision, recall, and f1 score. and explain how dangerous this cancer is, and highlight how it's important to detect the tumor in early stage via literature review, followed by publishing all our model online to be publicly available to be used by anyone.

specificity here measures proportion of true negatives that is correctly predicted, its equation is provided below.

$$\text{specificity} = \frac{TN}{TN + FP}$$

Where

- **TN**: True Negatives, Negative examples that have been predicted correctly.
- **TP**: True Positives, Positive examples that have been predicted correctly.
- **FN**: False Negatives, Negative examples that have not been predicted correctly.
- **FP**: False Positives, Positive examples that have not been predicted correctly.

Whereas, Sensitivity measures how well we can capture positive examples, and its equation if provided below.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Moving to Precision, it's one of the very popular metrics, its equation if provided in the following equation, where recall mean simply, how many of true positives are recalled. F1-score, is metric that capture both precision, and recall.

$$\begin{aligned}\text{Precision} &= \frac{\text{TruePositives}}{\text{ActualResults}} = \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{\text{TruePositives}}{\text{PredictedResults}} = \frac{TP}{TP + FN} \\ \text{F1-score} &= 2x \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\end{aligned}$$

We see that Sensitivity is our most important metric, as we are afraid more about cases that are positive and not get diagnosed correctly, in other words, we need to minimize FN.

4 Transfer learning

Transfer learning is widely used technique to develop new CNN models. Producing a new specific architecture for certain tasks usually takes about 3-6 months of trial and error, additionally it needs huge amount of data to converge successfully.

Transfer learning basically depends on pretrained network, trained usually on large dataset, and we benefit from their months of development, and trial and error to find the best fine tuned architecture, which usually could be generalized to many problems with only slight modifications, by re-training the model for only small number of epochs with weights initialized as same as original one at first.

4.1 EfficientNet family

The main idea of EfficientNet networks is that how to scale well-defined model architecture up to level where it holds for a larger dataset and new images higher in resolution.

EfficientNet is based on neural network developed by AutoML MNAS framework, And scale it up, there are some certain assumptions they did to balance the dimensions which are

$$\begin{aligned} \text{Depth} &= \alpha^\phi, \text{Width} = \beta^\phi, \text{Resolution} = \gamma^\phi \\ \text{Such that } \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned}$$

It was found out it performed far better than many competitors in many field, and proved itself in many Transfer-learning tasks

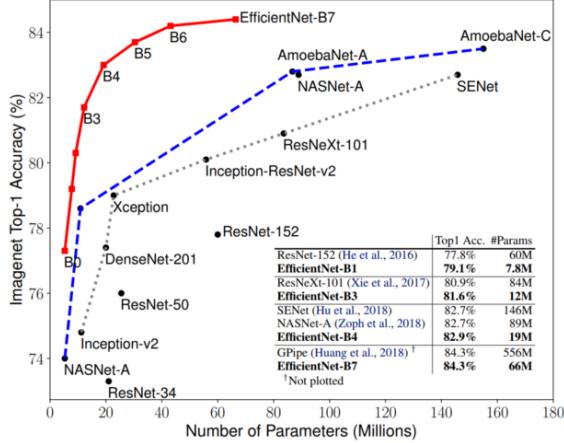


Figure 1: EfficientNet Performances

5 Methods

5.1 WSI and H&E

Whole slide image or as referred as (WSI) indicates scanning a whole microscopic slide, and producing high resolution digital file(Digitalization of the image). It's a very common technique when comes to biological labs; producing the high resolution image has many advantages as it decrease the possibility of the slide getting broken or lost, and is also very helpful in tele-diagnoses, and second opinion, for example someone might send an image of his scanned tissue to a well-known expert or large hospital and get diagnosed without travelling.



Figure 2: microscopic slide

Hematoxylin and Eosin (HE) is one of the most popular stains ever used in medical diagnosis. It shows broad range of cytoplasmic, nuclear, and extracellular matrix features that is so essential for decision making.

5.2 Invasive ductal carcinoma(IDC)

Duct is simply just a thin tube that carries milk from lobules to the nipple, And carcinoma indicates to any malignant tumor found on either skin(Epithilium tissum) or internal tissue of organs

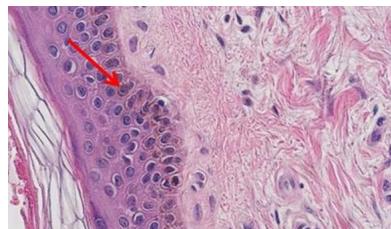


Figure 3: Tissue stained with HE

Invasive Ductal Carcinoma (IDC) of the Breast

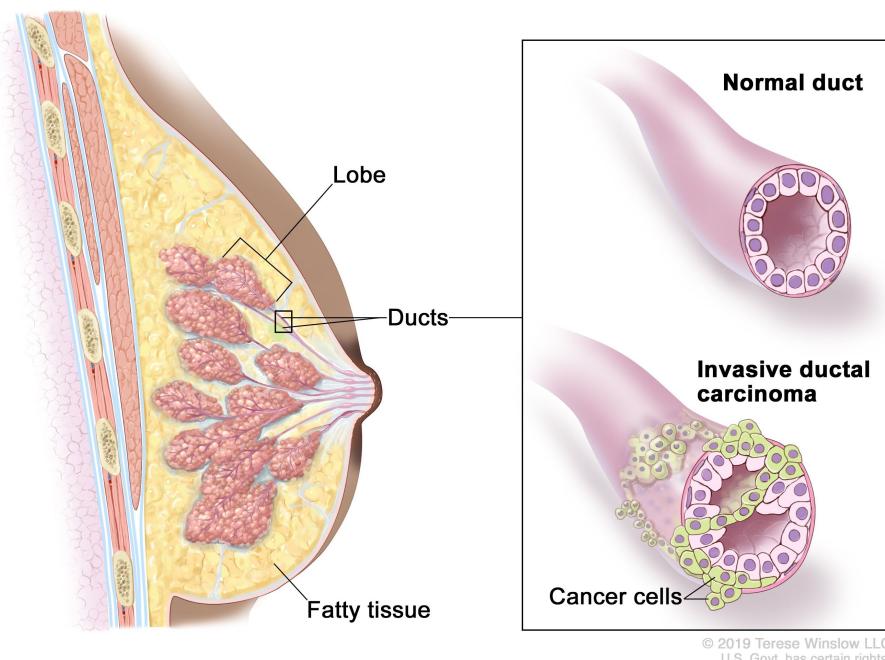


Figure 4: IDC example

According to (Sheng), IDC probably does not cause any symptoms, probably it's found by scans. However, these are some symptoms that might appear, or noticed by doctor

1. Swelling of all parts of the breast.
2. Skin irritation.
3. Skin dimpling.
4. Nipple discharge something other than milk.
5. Redness or thickening of nipple area.

6 Dataset description of IDC dataset

Our dataset is described in detail in *Cruz-Roa et al, Automatic detection of Invasive Ductal Carcinoma in whole slide images with Convolutional Neural Networks* paper, additionally they proposed a CNN architecture.

Our main focus in this report is to classify Invasive Ductal carcinoma(IDC) from normal tissue. Hence our dataset compromises of 162 women diagnosed with IDC at the Hospital of the University of Pennsylvania and The Cancer Institute of New Jersey. A biopsy is extracted from Each patient; which is tissue section that is probably suffering from cancer, and then scanned by whole slide scanner at 40X magnification.

Whole slide image generated by the scanner suffers from large size, it might reach to 60MB, which will affect training procedure of our model, as we would then have smaller batch size. To solve such problem, a common approach is to split the image in non-overlapping images patches, in our case an image was split into 100x100 pixels via grid sampling. However not all images hold valuable information for our model, and to save time, an expert excluded patches that contain fatty tissue or slide background.

Image patches were then annotated by expert pathologists, and they generated binary classification mask; consequently, an image patch is considered to be positive, if at least 80% of the patch fall into annotation mask. An example of image and how it's sampled is shown in figure (5), where red square indicates positive example of IDC, and green indicate indicates Negative path of IDC

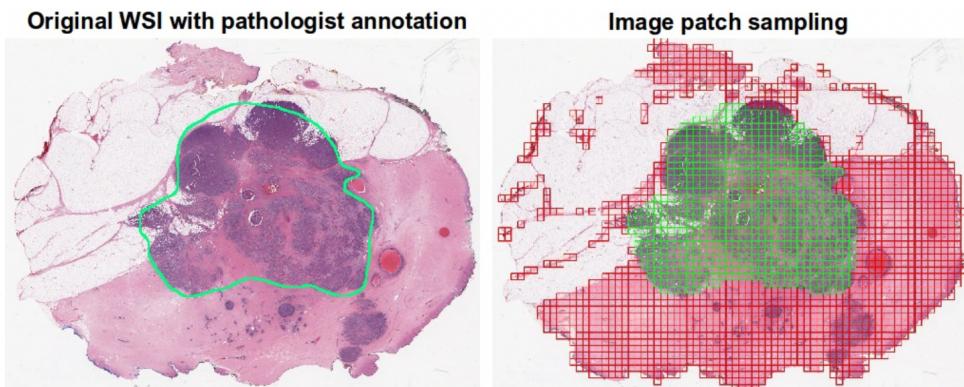


Figure 5: Image Path vs Patch sampling

The manual annotation raised some apprehension for us, how can we sure about their labeling?, that gave us an idea to use the probability of a patch being positive to be 0.9 and let other possibility to be 0.1 not 0, and vice versa for the negative examples.

7 Literature review

7.1 J. Zhang, X. Guo, B. Wang and W. Cui, "Automatic Detection of Invasive Ductal Carcinoma Based on the Fusion of Multi-Scale Residual Convolutional Neural Network and SVM [1]

Zhang et al [1] worked on a very similar problem to us, he worked on 162 women diagnosed with breast cancer; specifically Invasive Ductal Cancer(IDC). They proposed new block type they called it MSRC block; it's little similar to residual networks, and they used CNN feature extractor followed by either Fully connected layer (FC) classifier, or using SVM classifier.

They stated the breast cancer is one the most common type of breast cancer all over the world, only in 2018; about 2.1 million cases appeared having breast cancer, moreover, about 70% of these cases are suffering from IDC. However, there is a very promising survival rate when diagnosed in early stages, then they have 80% survival rate. The diagnosis is still a challenging problem, the average accuracy of pathologists is only 75%, furthermore, it's more challenging problem as the biopsy after being extracted it's stained with Hematoxylin and Eosin(HE), where the nucleus is stained purple, and cytoplasm is stained red, so it's little difficult for pathologists to diagnose correctly when there are multiple factors affecting the final decision, including the biopsy, stain, thickness of tissue, area of tissue,..etc, The lack of experienced pathologist in developed countries, and small hospitals in developed countries decrease the chance of early diagnosis.

in this paper they worked on large scale image of high resolution, so of course it was hard to feed them all at once on Graphics Card Unit (GPU), so what they did was that they cut the image, and labeled it, based on the origin of cut and patient ID, and patient class. They also split the data into 3 categories, training, validation, and testing in ratio 0.6:0.2:0.2.

MSRC block was originally proposed by this paper, it's based on He Kaiming *et al*, where he proposed theory of residual to solve the problem of degradation. The MSRC block is showing in figure (6). One key feature in their block is that they used three different kernels 1*1, 3*3, 5*5, to improve feature extraction ability. They justified their proposal by stating that MSRC merges multiple-scale features followed by max-pooling to reduce dimensionality, and that possibility of overfitting.

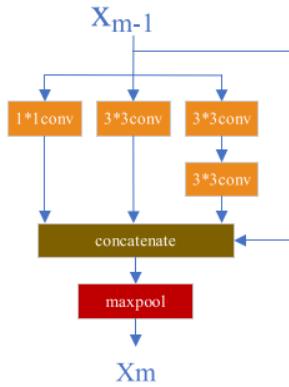


Figure 6: MSRC block

Their methodology was quite unique for us, they are as follows

1. Data processing.
2. Establish MSRC block with softmax output activation function.
3. Train the model and save it.
4. used the model without FC classifier, they only used feature extractor, and fed the data to SVM classifier for training
5. testing stage, by getting feature extracted using CNN and classify it using SVM

the structure and hyperparamters is as showing in figure 7, and 8. and their result is as showing in figure 9. Noting that they have experimented X-MSRC block, for each for in figure 4, represents results of using X MSRC block in the CNN model.

Variable	Setting
Epochs	100
Batch size	128
Learning rate schedule	Adam
Learning rate	0.001
Weight decay	0.0001
Loss function	Cross-entropy cost function

Figure 7: hyperparamters settings

Layer	Number	Activation
Fully connected layer	512	Relu
Fully connected layer	128	Relu
Dropout layer	—	—
Output layer	2	SOFTMAX

Figure 8: FC layers

This paper has inspired us to the possibility of using SVM rather than normal FC layers, and considering to try ResNet in transfer learning, or trying their proposed block.

	Acc	Bac	F1
X=1	83.33	81.15	73.90
X=2	84.70	82.67	75.98
X=3	85.20	83.56	77.00
X=4	86.99	84.82	79.2
X=5	83.81	81.07	74.70

Figure 9: Accuracies results

7.2 Ranjan, R *et al*, Comparative assessment of CNN architectures for classification of breast FNAC images [2]

In this paper they did an comparison between different common architectures performances used usually in transfer learning, they compared between VGG16, VGG19, ResNet-50, GoogleNet-V3. They worked on FNAC dataset, patients were diagnosed with Breast cancer. their goal was to find the optimal CNN architecture for such a task.

They started with stating that is the second most type of malignancy after lung cancer, and it's the fifth common cause of death. The probability of death is much higher in developing countries due to late diagnosis of cancer.

Fine Needle aspiration cytology (FNAC), is a method to collect samples from organs with minimal feeling of pain. They collected the biopsies from patients in Ayursundra Healthcare Pvt. Ltd, Guwahati, India, and good doctors prepared the samples and slides. about 212 images(99 benign, and 113 malignant). Due to the lack of slides, they used Data augmentation to provide 2120 image of benign tumor and 1130 image of malignant.

A new trick for us was proposed in this paper, as known, the image has three channels Red, Green, and Blue (RGB). and looked at every channel individually, and found out that the red channel was the best in terms of lack of noise (from Blood cells) and clarity of nucleus and their boundary. Their sample of images in showing in figure 10 .They also performed thresholding on the image using otsu thresholding, followed by Histogram equalization on red channel pictures, it completely removed cytoplasm. Their hyperparamter setting is as in the following table.

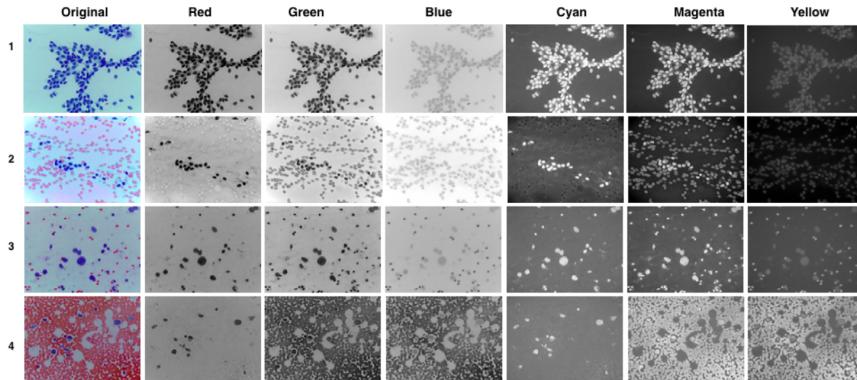


Figure 10: Output of six different channels

Model	Batch size	Epoch	Verbose	SGD-LR-rate	momentum
VGG16	32	12	1		
VGG19	32	12	1		
ResNet50	32	12	1		
GoogleNet-V3	32	12		1e-4	0.9

They found out that GoogleNet V3 fine tuned, was the best one in terms of accuracy and loss, according to table of results in figure 11.

Model	Accuracy (%)	Loss
VGG 16	63.2	0.6395
VGG 16 fine tuned	88.67	0.2967
VGG 19	60.84	0.8606
VGG 19 fine tuned	88.2	0.2875
ResNet 50	85.61	0.3414
ResNet 50 fine tuned	90.56	0.2571
GoogLeNet V3	71.88	0.3295
GoogLeNet V3 fine tuned	96.25	0.0828

Figure 11: Output of six different channels

This paper has gave us some ideas to consider only individual channels especially red in image classification. and using histogram equalization after thresholding the image. It's possible also that GoogleNet-V3 architecture is pretty suitable for this kind of problem.

7.3 F.A. Spanhol, L.S. Oliveira, P.R. Cavalin, C. Petitjean, L. Heutte Deep features for breast cancer histopathological image classification

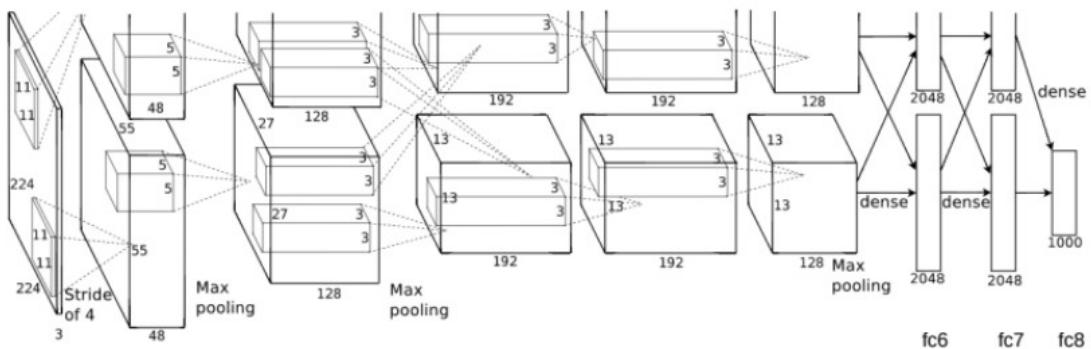


Figure 12: AlexNet model used as a base model

Deep Features for Breast Cancer Histopathological Image Classification: is a paper that provides all methods used and their accuracy and made a comparison with the literature to give finally the best accuracy to be 90 plus or minus 6.7

What this paper provides is using the AlexNet model as a base model, and also modifying or adding some more layers to enhance the overall accuracy of the model[5]. So, our methodology to make a better performance is to use other transformations, tune the set of numbers of parameters in the model, and if increasing data would give a better performance may we use different methods for increasing the data like GANs.

7.4 Kaggle competition code

Kaggle's users provide their code and model to be publicly available. It was quite rich source of information for us so we just scammed their rest and most important points in each's notebook.

ID	Author	Transferred?	Transferred from	Result accuracy	AOC	Notes
1	DANIELH CARRANZA	No	—	86.9%	0.948	—
2	SAYANTAN DAS	No	—	89%	—	—
3	SACHIN SHARMA	Yes	EfficientnetB0	75%	—	—
4	A MERII	No	—	58%	—	SVC
5	A MERII	No	—	69%	—	SVC
6	ANGIE ASHRAF	Yes	MobileNetV2	84.79%	—	—
7	A MERII	No	—	80%	—	SVC
8	ZEAD OMAR	No	—	84%	—	Custom CNN
9	TH DUY	No	—	82%	—	Custom CNN
10	PRATYUSH PATNAIK	No	—	86%	—	Custom CNN
11	SHIRISH	Yes	efficientnet-b5	87%	—	—
12	GERRY	Yes	InceptionResNetV2	86%	—	—
13	AYUSH VERMA	Yes	VGG16	91%	—	—

8 Proposed models

8.1 IDC

In this section, we present all models we have worked on, to make a comparison between them, and details of each model will be discussed as well. It's worth noting that we have splitted our data in all models into training, validation, and test test, in ratio 0.6;0.2:0.2.

First we provide summary of classifiers that we have used in different models on in the following tables. Note they are to be referred to when discussing different models proposed.

Classifier 1

Layer name	Number of neurons	Activation	Dropout Value
Fully connected layer	512	Relu	—
Dropout layer	—	—	0.5
Fully connected layer	256	Relu	—
output layer	2	Softmax	—

Classifier 2

Layer name	Number of neurons	Activation	Dropout Value
Fully connected layer	256	tanh	
Fully connected layer	256	tanh	
Dropout layer	—	—	0.25
Fully connected layer	256	tanh	
output layer	2	Softmax	

Classifier 3

Layer name	Number of neurons	Activation	Dropout Value
Fully connected layer	256	tanh	
Dropout layer	—	—	0.25
Fully connected layer	256	tanh	
Dropout layer	—	—	0.25
Fully connected layer	256	tanh	
output layer	2	Softmax	

Classifier 4

Layer name	Number of neurons	Activation	Dropout Value
Fully connected layer	512	tanh	
Dropout layer	—	—	0.25
Fully connected layer	256	tanh	
Fully connected layer	128	tanh	
output layer	2	Softmax	

Classifier 5

Layer name	Number of neurons	Activation	Dropout Value
Fully connected layer	256	tanh	
Fully connected layer	256	tanh	
Dropout layer	—	—	0.25
Fully connected layer	256	tanh	
output layer	2	Softmax	

Classifier 6

Layer name	Number of neurons	Activation	Dropout Value
GlobalAvgPooling	—	—	
Fully connected layer	4096	Relu	
Fully connected layer	4096	Relu	
Dropout layer	—	—	0.2
Fully connected layer	2096	Relu	
output layer	2	Softmax	

Second we present all models that were trained in the following table with its details. Note that we are referring to classifier from the previous tables architectures. Retrain column indicates whether if we retrain all layers or just the classifier, or half of layers. It's worth noting also that batch size of all model was 64.

All models

ID	Transferred from	Re-train	Classifier	Optimizer	Learning rate	Epochs
1	EfficientNetV2L	The whole model	classifier 2	RMSprop	1e-3	10
2	ResNet50	The whole model	classifier 2	RMSprop	5.5e-4	10
3	ResNet50	The whole model	classifier 2	RMSprop	2e-5	10
4	ResNet50	The whole model	classifier 3	RMSprop	2e-5	10
5	ResNet50	The whole model	classifier 3	RMSprop	2e-5	10
6	ResNet50	The whole model	classifier 4	RMSprop	2e-5	10
7	EfficientNetB3	The whole model	classifier 2	RMSprop	1e-3	10
8	InceptionV3	The whole model	classifier 5	RMSprop	5.5e-5	10
8	InceptionV3	The whole model	classifier 5	RMSprop	5.5e-5	10

Note the following

- Model 3 is same as model 2, but with less learning rate.
- Model 4 is same as model 5, but with more regularization.
- Model 5 is same as model 4, but without early stopping.
- Model 6 is same as model 5, but with different classifier.

Note we have actually tried number of other models with different optimizers, but some of them were missed, we couldn't download the model before colab got disconnected, hence we couldn't calculated our metrics, and some were highly overfitting; so we didn't report any them.

9 Results of proposed models

9.1 IDC

Here we provide summary of all models that have been discussed in *Proposed model section*

model number	Accuracy	Sensitivity	Specificity	Precision	Recall	F1 score
1	86.33%	0.83	0.85	0.68	1.24	0.443
2	88.4%	0.78	0.90	0.75	1.38	0.489
3	86.8%	0.76	0.905	0.75	1.457	0.4957
4	87%	0.75	0.91	0.75	1.457	0.4957
5	86.4%	0.78	0.89	0.73	1.36	0.477
6	86%	0.72	0.911	0.754	1.62	0.51
7	86.9%	0.04	0.99	0.85	-0.05	-0.05
8	87.29%	—	—	—	—	—

We provide here graphs for accuracy and cross entropy vs Epochs for each model

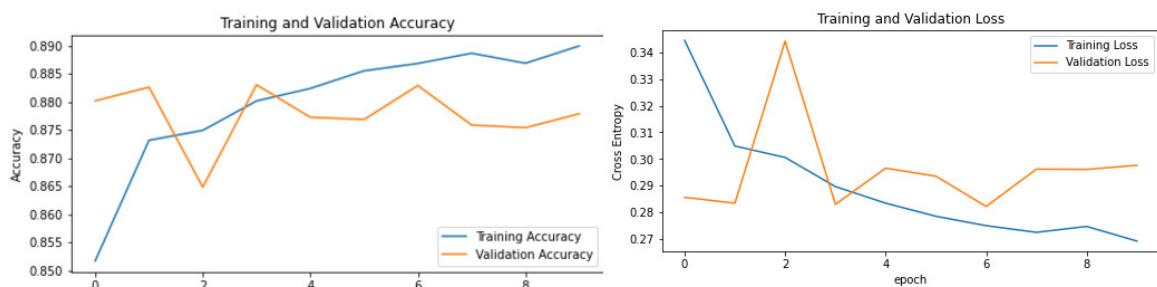


Figure 13: Model 1

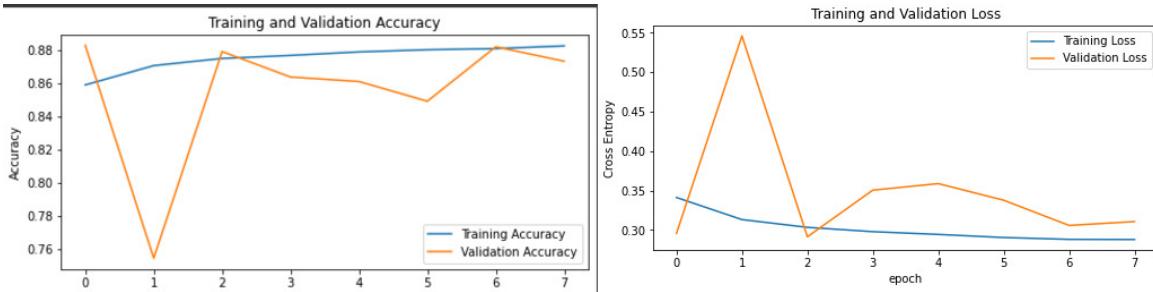


Figure 14: Model 2

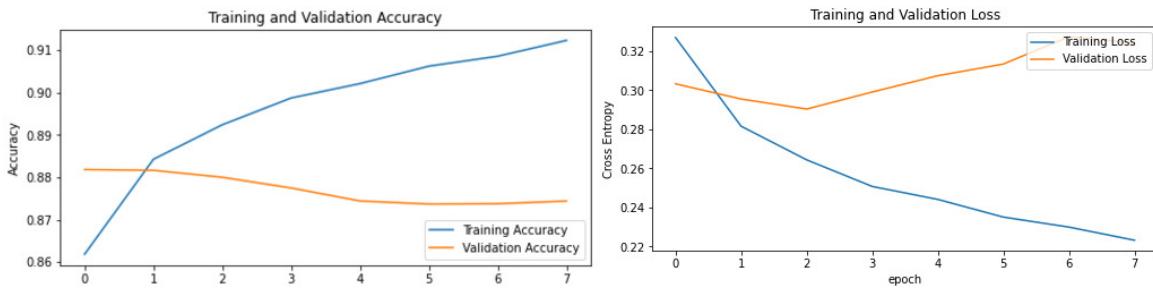


Figure 15: Model 3

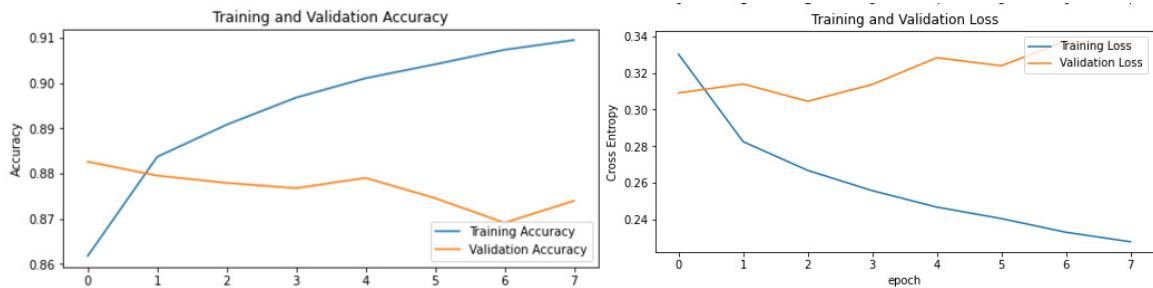


Figure 16: Model 4

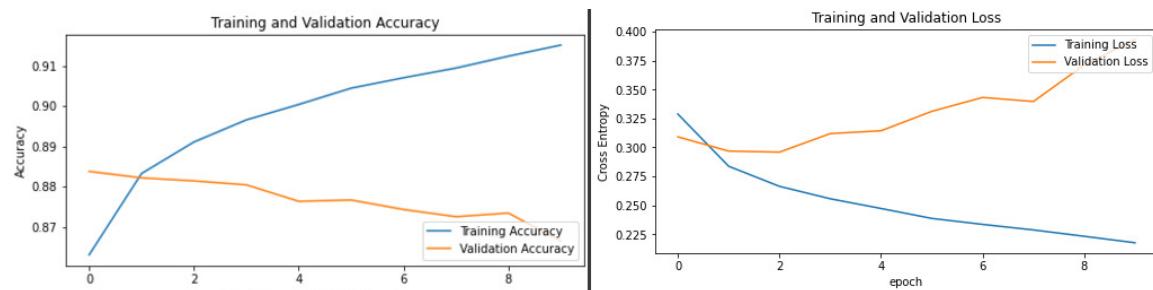


Figure 17: Model 5

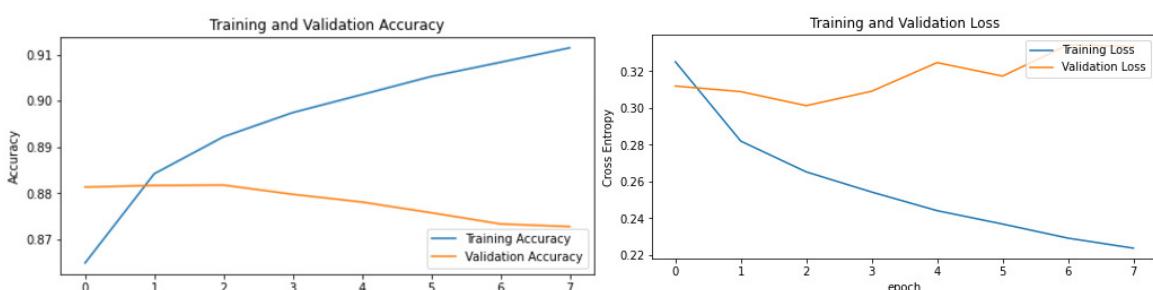


Figure 18: Model 6

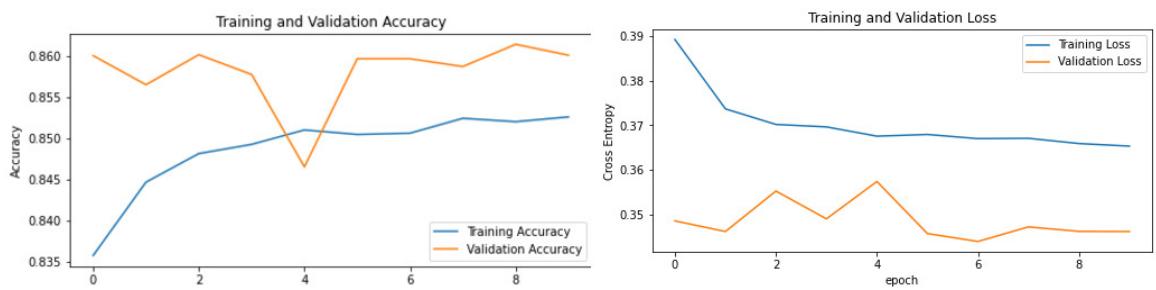


Figure 19: Model 7

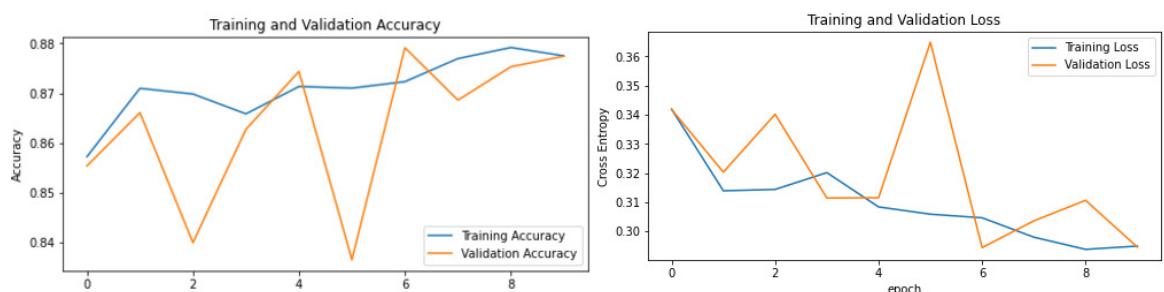
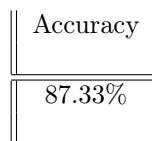


Figure 20: Model 8

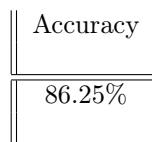
10 Fine tuning some models

We tried to fine tune some models, The search space was all about classifier, how many layers, number of neurons per layer, and activation function in each layer, and whether there should Dropout layer or not. We present in the following figure their results. It's worth noting that google colab kept disconnecting before finishing the search, and we noticed that the accuracy curves in previous models saturated after first two epoch, we made max trials to be 10, and epochs per trial to be 1.

Layer name	Number of neurons	Activation function.
InceptionV3	—	—
Flatten	—	—
Dense	768	tanh
Dense	768	tanh
Dropout	—	—
Dense	768	tanh
Output	2	softmax



Layer name	Number of neurons	Activation function.
ResNet50	—	—
Flatten	—	—
Dense	256	tanh
Dense	768	tanh
Output	2	softmax



11 Using SVM as classifier

We tried to use SVM as classifier, many papers in different field we have read have used SVM, so we decided to give it a try, and in this section we report our procedure and results.

We used model 2, which was based on ResNet50, and used extracted features from this model, by creating a new model, its input is the base model 2, and output is just the layer after flattening. its dimensions was 256. we predicted then features for all the training examples, and fed it to SVM and calculated our basic metrics. We got an accuracy **87%**, and in the following table we report some of the metrics calculated by Sikit-Learn.

	-	Precision		Recall		F1-score	
Positive class		0.89		0.93		0.91	
Negative class		0.80		0.70		0.75	

12 Deployment using MATLAB

we extracted some models in keras and used them in MATLAB program. MATLAB is really easy in making apps very fast.

First you select your image by Clicking "Select an image" Button

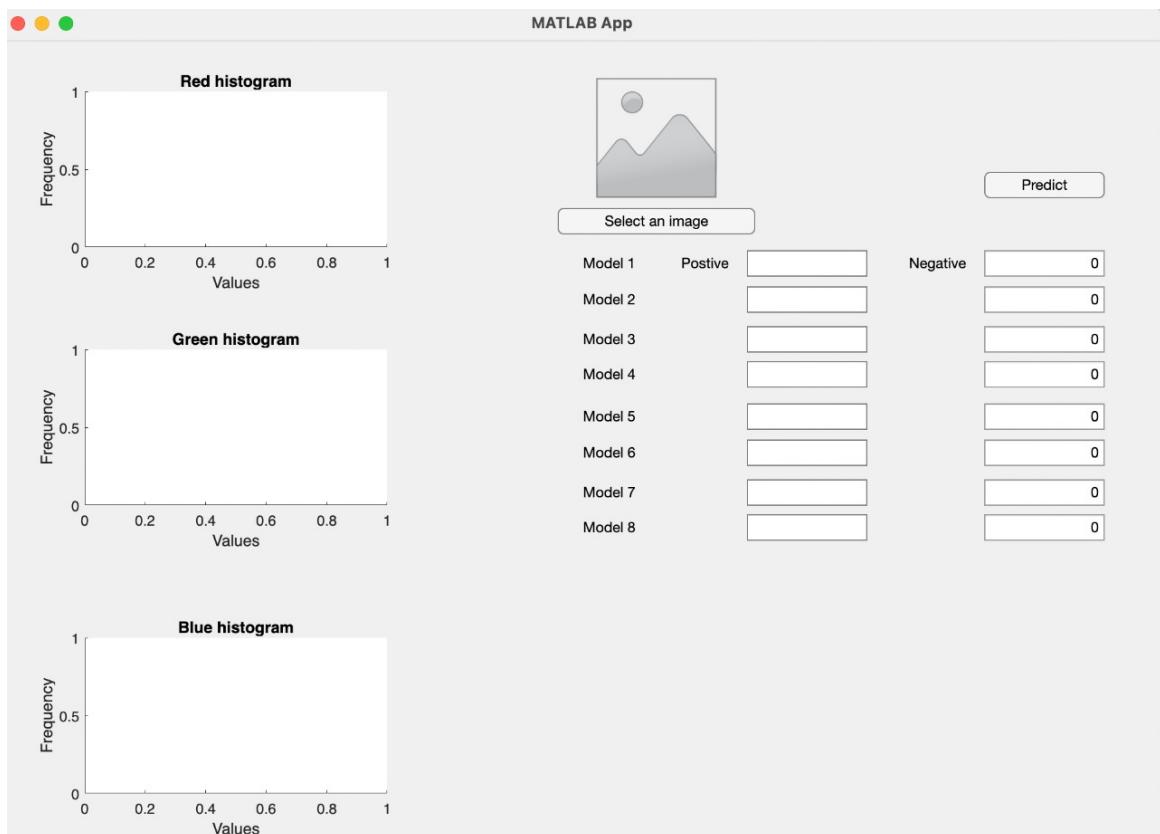


Figure 21: Model 6

then we are able to see the histogram of the image
after clicking "Predict" you will have to wait a little for the models to be loaded and output.

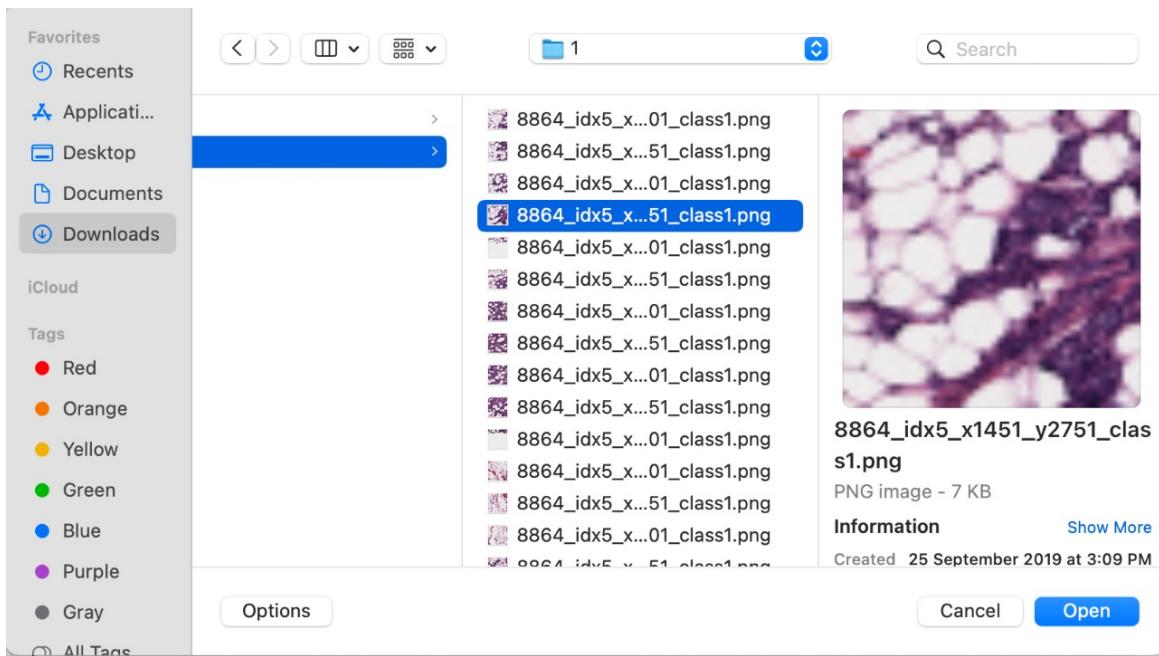


Figure 22: Model 6



Figure 23: Model 6

13 Ethical considerations

AI still has many challenges, and some ethical considerations must be taken, and we provide here some insights that is relevant to our project.

13.1 Data Privacy

According to (Badr, 2021) is one of the main challenges in medical imaging analytic. Any thing that might let any patient identifiable shall be hidden, Hence medical data is not publicly available as much as other fields. Some work have been done to mask the data while meeting some anonymity criteria by Fund el al.

13.2 Ensure AI transparency

Deep neural network is capable of identifying so many hidden deep features, might be able to extract gender and be biased to it. To avoid any misunderstanding, it's always better to build an explainable AI system.

13.3 Create opportunities for employees

AI has the potential to overcome the whole human system, consequently, there would be lack of provided jobs. To avoid such a problem, there must some humans involved in AI systems, for example to be second opinion in medical imaging.

14 Time line

We did some quiet plan for our project as not to rushed in some tasks. We gave enough time section for each task respective to its importance, we have cared a lot to literature review to understand our situation and what accuracies it has gotten till now. in the following table we present our plan for the project. starting from May 7.

Starting from	Duration	What to do.
May 7	1	Getting familiar with our problem by reading online and watch Youtube videos
May 8	2	Literature review
May 10	3	Kaggle code reading, and getting ideas from scientific papers
May 14	3	Start coding
May 17	2	Report writing
May 21	—	Getting feedback for our project
May 24	—	Reflecting the feedback on our code and analysis
May 26	4	Revising the final report

15 Team members contribution

We mostly divided our work in halves between team members, everyone focused on certain base models and tried to work on different classifiers.

16 Conclusion

in this report we have used trial-and-error technique, for many times, one of the models provided us with the best accuracy 88.8%. But most importantly, we have achieved sensitivity of 87%, which is way higher than average doctors.

We couldn't reach the best accuracy published in papers now, as we were limited by computational power, we couldn't search in large space with sufficient number of epochs, the image size was still too small for the model to extract features from; as it was only 100x100. We encourage in further work, to have better computational power for search in larger space with more epochs, and try stacking independent models that we have been developed; to outperform our best results. We also highly advise women all over the world to regularly check their breast for any cancer, especially if any weird mass was felt in breast area, for higher survival rate.

17 References

- [1] Zhang, Jianfei Guo, Xiaoyan Wang, Bo Cui, Wensheng. (2021). Automatic Detection of Invasive Ductal Carcinoma Based on The Fusion of Multi-Scale Residual Convolutional Neural Network and SVM. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3063803.
- [2] Saikia, Amartya Bora, Kangkana Mahanta, Lipi Das, Anup. (2019). Comparative assessment of CNN architectures for Classification of Breast FNAC images. *Tissue and Cell.* 57. 8-14. 10.1016/j.tice.2019.02.001.
- .
- [3] Badr, E. (2021). Images in Space and Time. *ACM Computing Surveys*, 54(6), 1–38. <https://doi.org/10.1145/3453657>.
- [4] Angel, C.-R., Mi, W., Pan, B., Guo, Y., Li, J., Xu, R., Zheng, J., Zou, C., Zhang, T., Liang, Z., Zou, J., Zou, H. (2021). Automatic Pancreatic Ductal Adenocarcinoma Detection in Whole Slide Images Using Deep Convolutional Neural Networks. *Frontiers in Oncology*, 11. <https://doi.org/10.3389/fonc.2021.665929>.
- [5] F.A. Spanhol, L.S. Oliveira, P.R. Cavalin, C. Petitjean, L. Heutte Deep features for breast cancer histopathological image classification 2017 IEEE international conference on systems, man, and cybernetics, SMC 2017, banff, AB, Canada, october 5-8, 2017 (2017), pp. 1868-1873, <https://doi.org/10.1109/SMC.2017.8122889>