Queensland University of Technology

13/11/2022

IFN646 Biomedical Data Science

Project Title: Breast Cancer Mammography Classification

Supervisors: Professor Dimitri Perrin, Dr Jake Bradford

Student names:

Ayush Raj (N10844821)

Chak Ho Chan (N10954856)

Li Jen Shao (N11096837)

Table of Contents

ABSTRACT 2

INTRODUCTION 3

What Is a Tumour? 3

Brief explanation about AI 4

Objective of this project 5

Literature Review 6

Motivation for the approaches 6

Gradient-weighted Class Activation Mapping (Grad-Cam) 7

Gradient based saliency maps 7

Methodology 8

Pre-processing 9

Deep learning model 11

Model Results 13

Interpretation and understand the decision making 13

Discussion 15

Future Works 17

Conclusion 17

References 19

Appendices 21

ABSTRACT

Deep learning-based neural network advances made recently in biological image processing might be used to increase the efficiency of Computer Aided Diagnosis (CAD) systems. An overview of the most current cutting-edge deep learning-based CAD systems created for mammography and breast histopathology pictures is provided, considering the significance of breast cancer globally and the promising outcomes reported by VGG back propagation-based approach in breast imaging. The study describes how well a mammographic imaging can forecast a positive patient for breast cancer while considering the breast tissue textural characteristics such as energy, contrast, correlation, and other texture descriptors at each pixel. The Convolutional Neural Network model, which we suggest as a machine learning-based method to modelling breast cancer, classifies the image as either normal tissue, Benign or Malignant tumour. We reached an accuracy of nearly 64% for the convolutional neural network model on the given test set and the model was found to have ability to detect abnormalities but have difficulties on classifying a cancerous tissue. All the source code and results were uploaded: https://github.com/chakhoho/IFN646project.

INTRODUCTION

Breast cancer continues to be the second largest cause of cancer death worldwide and is the most frequent cancer in women. The abnormal development of the cells lining the breast lobules or ducts is breast cancer. These cells have the capacity to spread to many bodily areas and multiply uncontrolled. The most common symptom found was breast thickening or new lumps, especially if they are present in only one breast.

What Is a Tumour?

A tumour is an unnaturally large lump of tissue within the body that has no discernible function. Cells that expand and divide too fast might give rise to it. They grow and behave differently depending on whether they are benign (noncancerous) or malignant (cancerous).

Benign vs Malignant Tumours:
A benign tumour is made up of cells that are not a threat to invade other tissues, and the cells are contained within the tumour.

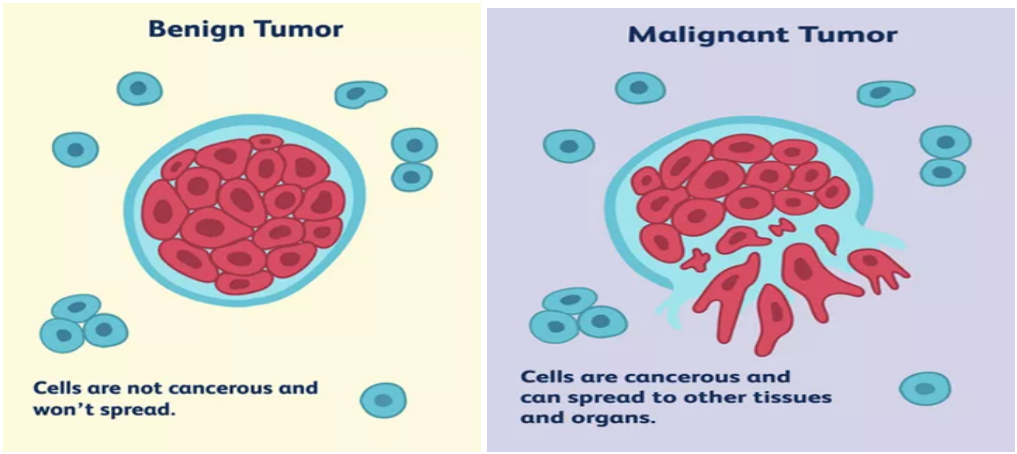


Figure 1: A demonstration of Benign and Malignant tumour.

Malignant means that the tumour is made of cancer cells that can grow uncontrollably and invade nearby tissues.
A common screening test for early breast cancer diagnosis is digital mammography. Breast density, which is mostly determined in modern clinical practice using the qualitative Breast Imaging and Reporting Data System (BI-RADS) density categories, assesses the amount of fibro glandular (i.e., dense) tissue observed on a mammogram.

The value of mammography screening has long been understood, and earlier breast cancer identification has been linked to improved results. Overdiagnosis, overtreatment, false-positive rates with accompanying psychological harm, excessive expenses, and biopsies are only a few of the drawbacks of mammography screening. Due to the varied ways that breast cancer presents itself as well as the masking

effect created by thick breast tissue, mammography interpretation varies with experience, is subjective, and prone to inaccuracy.

Several studies have revealed that breast tissue textural characteristics are also associated with the risk of breast cancer. Energy, contrast, correlation, and other texture descriptors calculate the local characteristics at each pixel and create a collection of statistics data from the distribution of the local attributes. This report discusses in-depth analysis of digital images of mammograms and associated imaging features that are potentially associated with breast cancer risk. Although two-dimensional (2D) mammography reading times are between 30 and 60 seconds, the massive quantities of mammograms and double reading of each mammogram lead to resource and manpower problems. The newly emerged deep learning techniques mitigates the above problems to some extent.

Brief explanation about AI

Deep learning is a subset of machine learning, which is based on deep neural networks, which oversimplify yet imitate human brain neurons. To extract features from basic to sophisticated abstractions from the input, Deep learning relies on deep layered architecture, which enables hierarchical learning. A mathematical model is learned by using an observation-based "training" set of data. The supervised learning system can "learn" the characteristics of malignancy from the tagged annotations in mammography since a tumour is accurately drawn on the picture (as the labels). A neural network is made up of a series of layers: an input layer (for example, the raw pixels from a mammogram), one or more hidden layers, and an output layer (for example, the model's prediction of the label "benign" or "malignant").

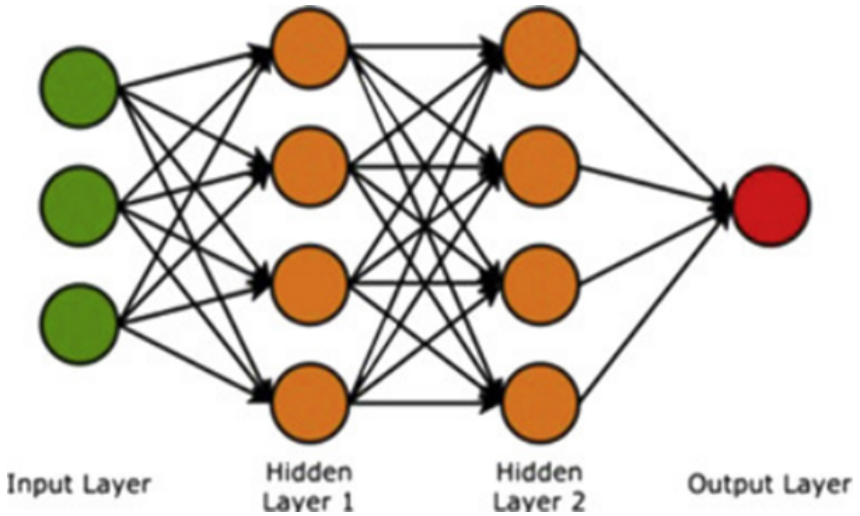


Figure 2: A demonstration of neural network.

Objective of this project

The objective is to investigate the learned convolutional neural network model to later generate predictions for new "test" data. In a machine learning regime, "learning" or "training" often refers to the iterative process of comparing current predictions (such as a malignant tumour) against a benign mass or normal tissue using an assessment measure known as the "objective function" (sometimes referred to as the "cost" or "loss" function). The model adjusts with additional training to close this difference in a clinically meaningful way, that is, until the predicted label matches the actual label.

Literature Review

Motivation for the approaches

The tools used for investigating and analysing the mammographic imaging for breast cancer are VGG Neural Networks. In most cases, it alludes to a deep convolutional network for object identification that performed exceptionally well on the ImageNet dataset. Currently it is the most capable model for object detection. Key features include using ReLU activation function in-place of tanh function, optimization for multiple GPU's and overlapping pooling. Also, it does address overfitting by using data augmentation. It also improved the traditional CNN model on training image data.

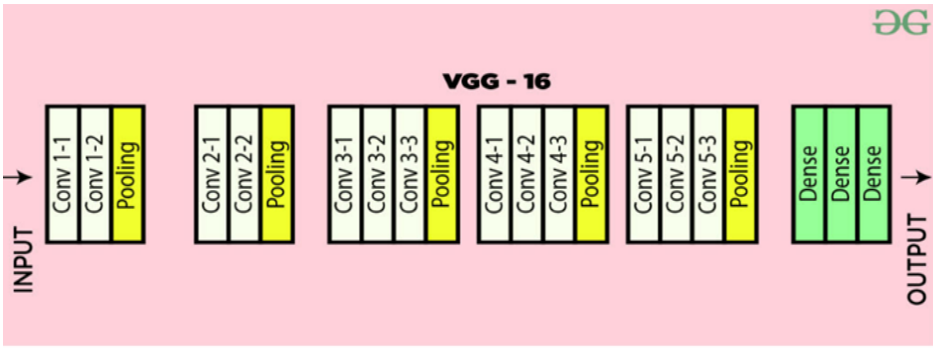


Figure 3: A demonstration of VGG neural networks style structure.

A dimensioned picture serves as the network's input (224, 224, 3). The same padding is present on the first two layers' 64 channels with a 3*3 filter size. Following a max pool layer of stride (2, 2), two layers have convolution layers with a 128 filter size and a filter size (3, 3). The next layer is a stride (2, 2) max-pooling layer and is identical to the one before it. After that, there are 256 filters spread over 2 convolutional layers with a 3 by 3 filter size. The next layers are a max pool layer, followed by two sets of three convolution layers. They both have the same padding and 512 filters of size (3, 3). A stack of two convolution layers receives this picture next.

Gradient-weighted Class Activation Mapping (Grad-Cam)

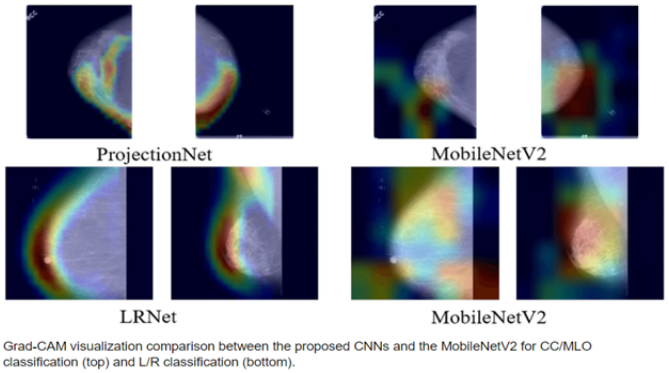


Figure 4: A demonstration of how Grad-Cam works.

The method has greater adaptability and accuracy than earlier methods. Although complicated, the result is fortunately clear. In general, we start with a picture as our input and build a model that is stopped at the layer for which we wish to build a Grad-CAM heat-map. For prediction, we affix the completely linked layers. The model is then applied to the input before the layer output and loss are collected. The gradient of the output of our chosen model layer with respect to the model loss is then determined. To overlay the heat-map with the original picture, we next take portions of the gradient that contribute to the prediction and decrease, resize, and rescale them.

Gradient based saliency maps

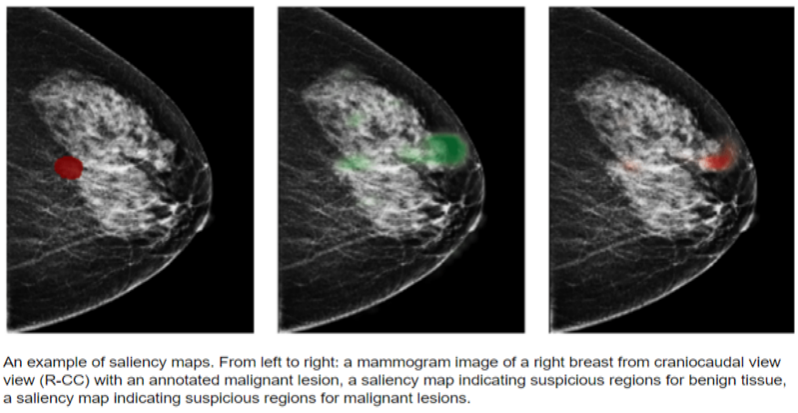


Figure 5: A demonstration of how Gradient base saliency maps work.

They are a well-liked visualisation technique for understanding why a deep learning network chose a particular action, like categorising an image. The gradient expresses how much a variable may influence

the outcome of another variable. Saliency maps are typically shown as heatmaps, with hotness corresponding to regions with a significant influence on the model's choice.

Methodology

The mammograms are obtained from CBIS-DDSM (Curated Breast Imaging Subset of DDSM), it is an updated and standardised version of the Digital Database for Screening Mammography (DDSM), it contains normal, benign, and malignant cases with verified pathology information. Initially, the DDSM is a database of 2,620 scanned film mammography studies with 9684 images. Initially, the data size is over 160GB, due to the hardware limitation we used the processed version of CBIS-DDSM from Kaggle (https://www.kaggle.com/datasets/cheddad/miniddsm). This MINI-DDSM dataset contains all the mammograms from the original CBIS-DDSM database, and it has been converted from (.ddsm) file to (.png) file. There are 3 kinds of images which are normal mammograms, mammograms with benign tissue and mammograms with malignant tissue, here we want to focus on right breast CC view images only due to hardware limitation. The goal of this analysis is to train a deep learning model and classify the mammography into 3 classes (Normal, Benign and Malignant).

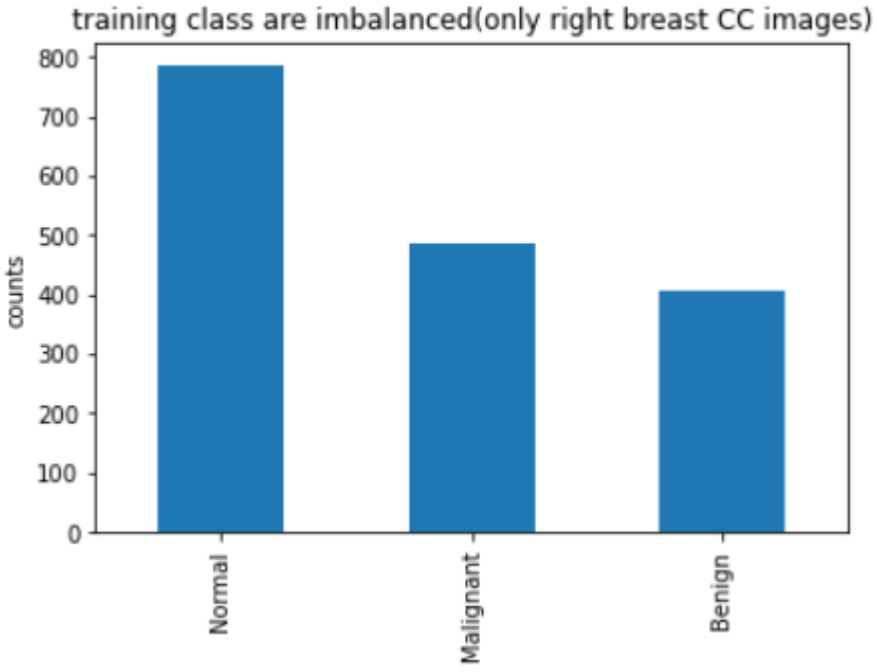


Figure 6: Number of samples in the right breast CC view subset. Normal class has 785 images, Benign class has 407 images, Malignant class has 487 images.

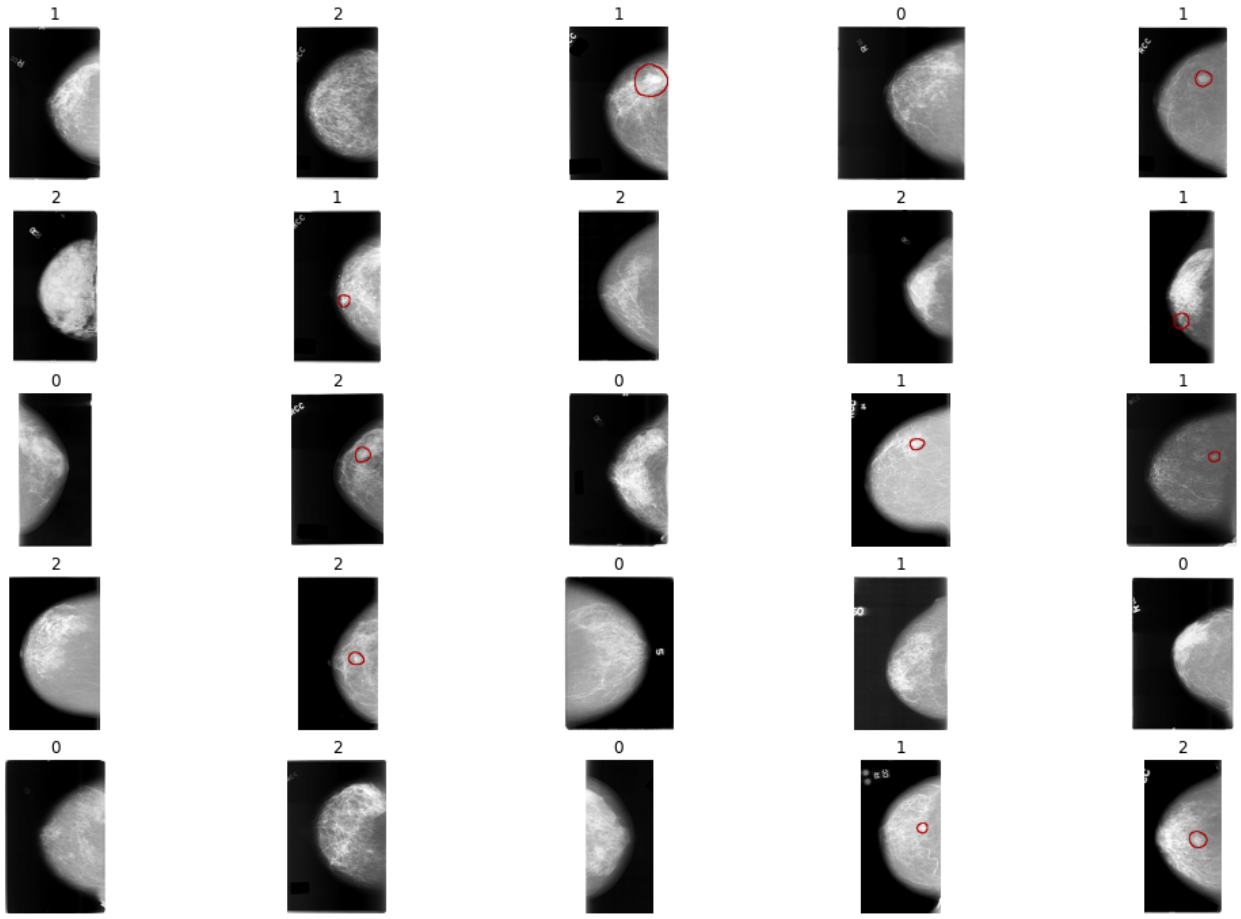


Figure 7: A sanity check of the initial right breast CC mammography. In terms of the label, 0 represents Normal class (normal breast tissue), 1 represents Benign class (abnormal but no cancer), 2 represents Malignant class (abnormal with cancer). Some of the images has the red annotation representing the Region of Interest (ROI), which indicates a general location of the abnormalities. Here we really cannot tell the difference by eyes.

Pre-processing

From figure 6 we can understand that the number of samples in each class are imbalanced, therefore we need to resample the data for the sack of fairness in the analysis. Here we under sample the images and randomly pick 400 images from each class (1200 images totally) to solve the class imbalanced issue. From the previous literature review we understand that there are some special pre-processing techniques that can help to enhance the detail of images, they are image-cropping, image resizing and normalisation, and image enhancement by CLAHE. From Figure 7 we can see some black spaces in the image background, these black spaces can potentially add noise into the data, and image-cropping can help to reduce the noise by cutting off some of the black space.

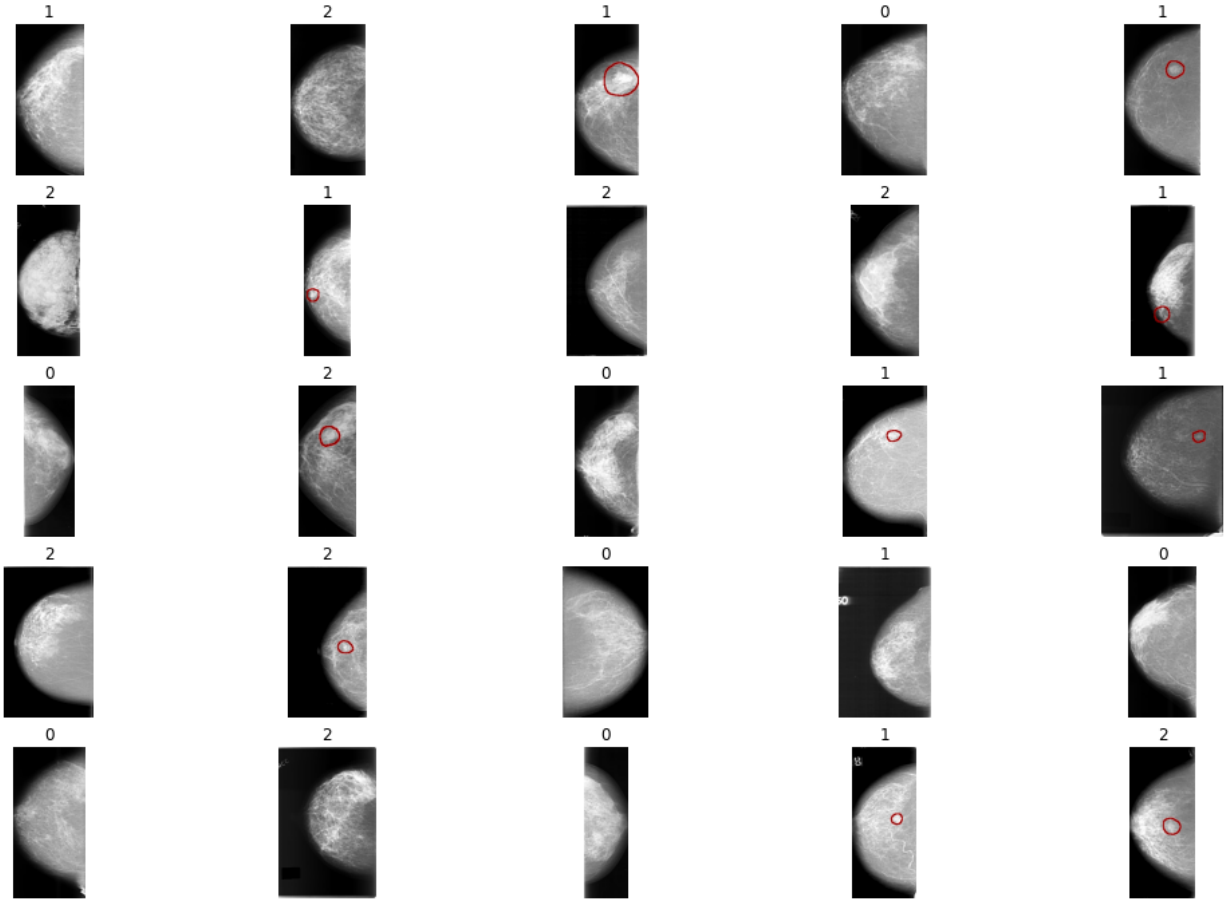


Figure 8: A sanity check of the cropped mammography, as you can see most of the empty background has been cropped.

Next is image resizing and normalisation. Initially, each of these mammograms has slightly different image sizes, and all the images need to be in the same scale in order to train a deep learning model. Therefore, we resize all the mammograms to 256, 256(width, height). Another reason for resizing is because 256 by 256 is the closest square shape to the image's original size, since most of the neural network models assume a square shape input images, it would be better if we resize all the images into square shape. Theoretically, we should get a better result if we increased the size of images without distorting the image, because that can reveal more details of the breast tissue for the model. However, training a model with a bigger size of images also requires better hardware. Normalisation is a typical process that can put images into a common statistical distribution in terms of size and pixel values, this can further reduce the noise in the data and help the model to converge faster while training the network.

Finally, contrast limited adaptive histogram equalisation (CLAHE) can help to enhance the contrast and reveal more details in the breast tissue, it is especially helpful when we deal with black and white colour

images (Pisano et al., 1998). It computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the luminance values of the image.

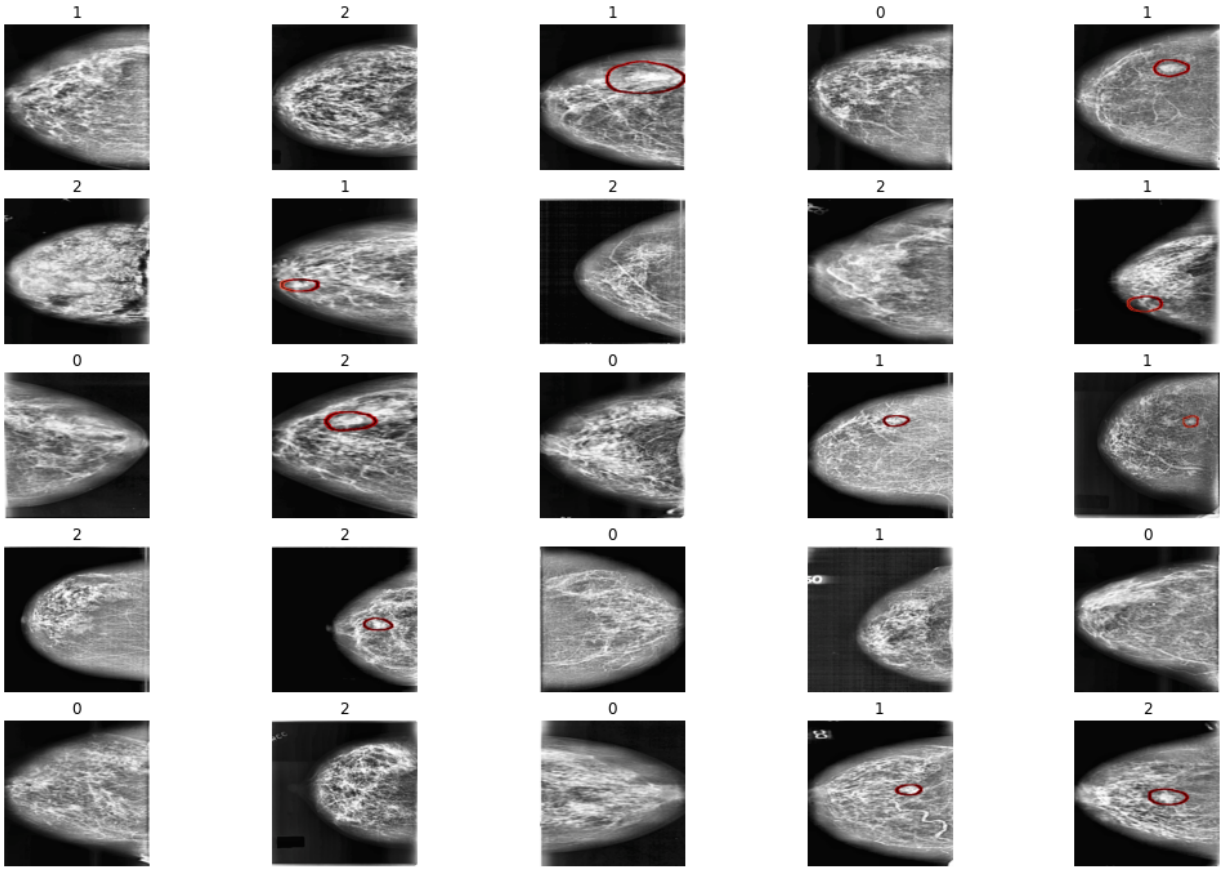


Figure 9: A sanity check of the final version of the mammography. As you can see after pre-processing, the details and contrast in the breast tissue has been enhanced.

Deep learning model

We trained a VGG-architecture backpropagation neural network based on pre-processed mammography, from previous literature review we understand that using deep learning algorithms on image classification can generate a better result than using traditional methods such as Random Forest or Support Vector Machine. We use 840 images as the training set, and 360 images as the testing set to make the prediction and assess the model performance, the response will be 0(Normal), 1(Benign) or 2(Malignant). We also produced a classification report and a confusion matrix to measure the model performance. VGG is a classical convolutional neural network architecture and was introduced by Karen Simonyan and Andrew Zisserman (2014). A VGG style network means we use one or several pairs of convolutional layers to build the network structure, one of the advantages of VGG comparing to normal neural network is that it has deeper depth structure but with almost the same or less number of

parameters in the model, by building a deeper network to get more accurate prediction on image classification. In machine learning, backpropagation is a widely used algorithm for training feedforward neural networks, it is the practice of fine-tuning the weights of a neural net based on the loss obtained in the previous epoch (iteration). Thus, we can see in our classification result, the training loss and validation loss was reduced in each iteration by this gradient descent method, meanwhile it also indicates that the model is able to do the generalisation if we observe the reduction of loss. In the model structure we also used Batch Normalization layer and Dropout layer to prevent overfitting because we assume it should have a lot of noise in the data by considering the nature of the mammogram itself. If we increase the number of images to train, we should be able to reduce the overfitting problem due to more training samples for the model to generalise the result.

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
images (InputLayer)          [(None, 256, 256, 3)]     0

conv2d_10 (Conv2D)           (None, 256, 256, 8)       224

conv2d_11 (Conv2D)           (None, 256, 256, 8)       584

batch_normalization_5 (Batc  (None, 256, 256, 8)       32
hNormalization)

spatial_dropout2d_5 (Spatia  (None, 256, 256, 8)       0
lDropout2D)

max_pooling2d_4 (MaxPooling  (None, 128, 128, 8)       0
2D)

conv2d_12 (Conv2D)           (None, 128, 128, 16)      1168

conv2d_13 (Conv2D)           (None, 128, 128, 16)      2320

batch_normalization_6 (Batc  (None, 128, 128, 16)      64
hNormalization)

spatial_dropout2d_6 (Spatia  (None, 128, 128, 16)      0
lDropout2D)

max_pooling2d_5 (MaxPooling  (None, 64, 64, 16)        0
2D)

conv2d_14 (Conv2D)           (None, 64, 64, 32)        4640

conv2d_15 (Conv2D)           (None, 64, 64, 32)        9248

batch_normalization_7 (Batc  (None, 64, 64, 32)        128
hNormalization)

spatial_dropout2d_7 (Spatia  (None, 64, 64, 32)        0
lDropout2D)

max_pooling2d_6 (MaxPooling  (None, 32, 32, 32)        0
2D)

conv2d_16 (Conv2D)           (None, 32, 32, 64)        18496

conv2d_17 (Conv2D)           (None, 32, 32, 64)        36928

batch_normalization_8 (Batc  (None, 32, 32, 64)        256
hNormalization)

spatial_dropout2d_8 (Spatia  (None, 32, 32, 64)        0
lDropout2D)

max_pooling2d_7 (MaxPooling  (None, 16, 16, 64)        0
2D)

conv2d_18 (Conv2D)           (None, 16, 16, 128)       73856

conv2d_19 (Conv2D)           (None, 16, 16, 128)       147584

batch_normalization_9 (Batc  (None, 16, 16, 128)       512
hNormalization)

spatial_dropout2d_9 (Spatia  (None, 16, 16, 128)       0
lDropout2D)

flatten_1 (Flatten)          (None, 32768)             0

dense_2 (Dense)              (None, 256)               8388864

dropout_1 (Dropout)          (None, 256)               0

dense_3 (Dense)              (None, 3)                 771

=================================================================
Total params: 8,685,675
Trainable params: 8,685,179
Non-trainable params: 496
```

Figure 10: A visualisation of the model structure. Note the input size is 256, 256, 3(width, height, channel) and the output size is 3 because we are predicting 0, 1 or 2. We also used Batch Normalisation layer and Dropout layer after each pair of convolutional layers.

Model Results

	precision	recall	f1-score	support
Normal	0.85	0.82	0.83	120
Benign	0.55	0.53	0.54	120
Malignant	0.52	0.56	0.54	120
accuracy			0.64	360
macro avg	0.64	0.64	0.64	360
weighted avg	0.64	0.64	0.64	360

```
12/12 [==============================] - 0s 22ms/step
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2583cc225e0>
```

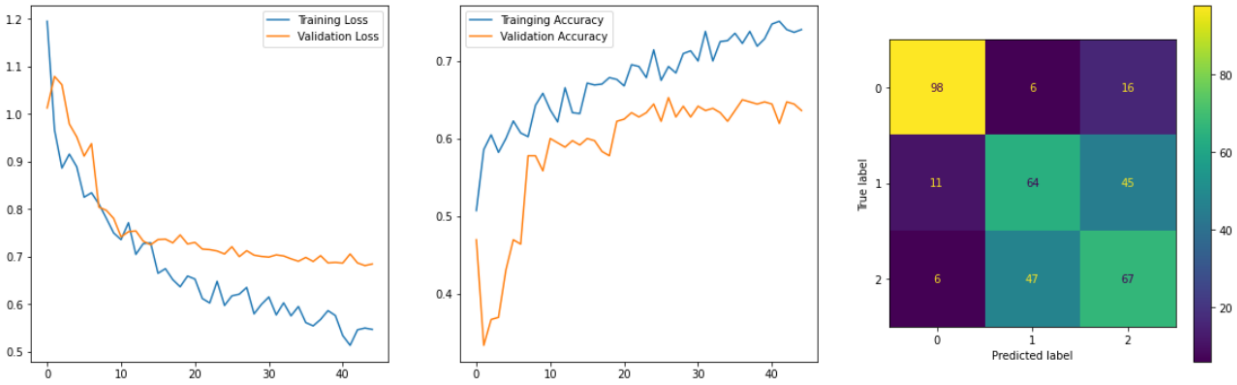


Figure 11: A classification result and visualisation of model performance after 45 iterations.

By observing the result from figure 11 we can understand that the model was just able to classify 3 classes of mammography, but the result is not that impressed. Ideally, we want to see the reduction of loss in each iteration and the increase of accuracy in each iteration due to backpropagation structure, both training curve and validation curve should be as close to each other as possible. Here the validation loss reduced from 1.01 to 0.68 after 45 iterations, the validation accuracy started from 0.46 to about 0.65 after 45 iterations. From the f1-score and the confusion matrix of three classes we can understand that the model seems can easily recognize the Normal class images but had difficulty on classifying Benign and Malignant class images relatively. We reached an overall accuracy of 0.64, we decided to stop the training at 45 iterations because it seems the model will start to get overfitting after 45 epochs.

Interpretation and understand the decision making

We would like to know how model did the decision making after model training. Gradient-weighted Class Activation Mapping (Grad-Cam) is one of the techniques that can help us to understand the feature importance in image classification problem. Grad-Cam can generate heatmaps by using the information from the final convolutional layer of the model, then overlap the heatmap with our original images to see whether the important region that the model recognised are in the same location as the ROI or not. This process can help us to uncover some of the information from neural network's black box, also can help to verify the prediction result.

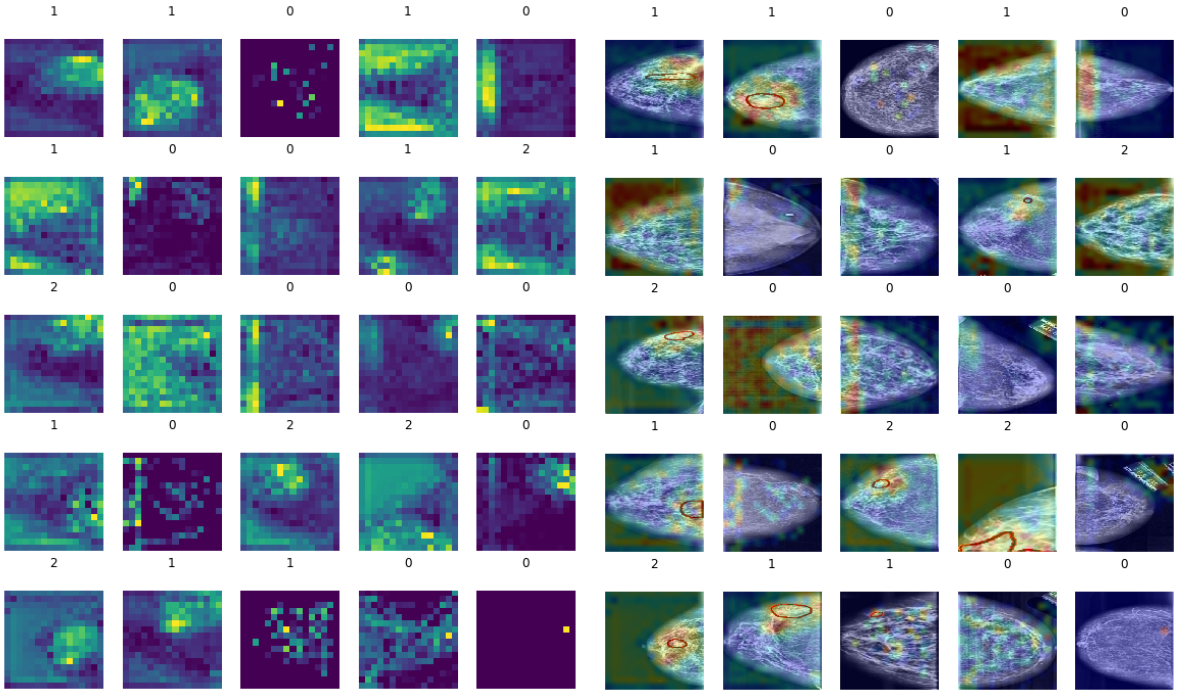


Figure 12: A heatmap that was generated by Grad-Cam, note that in this stage we are using testing set images.

From figure 12 we can see after combining the heatmap and our testing set images, some interesting patterns were revealed. The red colour region represents the important area for the prediction that were recognised by the model, the red annotation is the ROI that were recognised by the pathologist. Even though we only achieved overall accuracy of 0.64, we can still see that the red area is generally close to or in the red annotation, and some of the dense tissue had been wrongly recognised as important region. Combine the result from figure 11 and 12, it seems the model are good at classifying normal breast images and abnormal breast images but hard to classify whether the abnormality is cancerous or not.

Discussion

Female breast cancer is a deadly disease worldwide. Women's lives can be saved if the cancer is detected at the earliest stage. Classifying mammogram images based on features is difficult because extracting the best features from mammogram images is a challenging task. We used a publicly available mammogram image datasets, CBIS-DDSM to extract features and perform classification. Data augmentation was performed to increase visibility of the data. Deep learning models can achieve relatively better results when classifying image datasets.

Convolutional Neural Networks (CNN) were applied in this project. This model was developed to increase the depth of CNNs for image classification. Although VGG architecture has been considered one of the best methods in image classification, it is interesting to ask why this model is better compared to different models. To estimate the performance of the system, different measures such as accuracy, precision, recall, F1-score were used. Unfortunately, our hardware specification is not sufficient in running different models, it takes quite a long time to train the image dataset. The final model achieves up to 64% training accuracy. It can be further improved by using more image enhancement techniques to strengthen the image signal and obtain higher accuracy. Comparing CNN with different types of neural network models is also needed.

Furthermore, it is hard to differentiate a single type of mammogram difference just by observation, as benign or malignant can be based on different factors. Our classification can provide a preliminary result to classify normal breast or abnormal breast. It needs further experiments such as breast biopsy and immunoassay for tumour markers to investigate if the breast tissues belong to benign or malignant. In the feature selection process, we only applied Right and CC mammograms. It would be interesting to use different types of mammograms (MLO). Since we cannot tell which types of mammograms will provide better features to the prediction. Then with the use of a CNN model might be able to differentiate mammograms from some clues of different types of mammograms. The result might be more convincing if it can use the whole dataset with different image types to train the model, or ensemble a few referent models to the optimal model.

Regarding the region of interest (ROI) of our model to classify mammograms. The Grad-Cam was generally able to show the abnormal location within the ROI. Some of the dense tissue had been wrongly detected as the important region to the prediction, this suggests the model had difficulties on classifying dense tissues to benign or malignant. Ideally, it can mean some areas may have more than one abnormal tissue, the model can spot them as early as possible so it can do further testing and prevent it getting worse, it could simplify work and easily get a preliminary result, but we cannot tell since we did not achieve high accuracy.

Figure 13: Before using contrast limited adaptive histogram equalization (CLAHE).

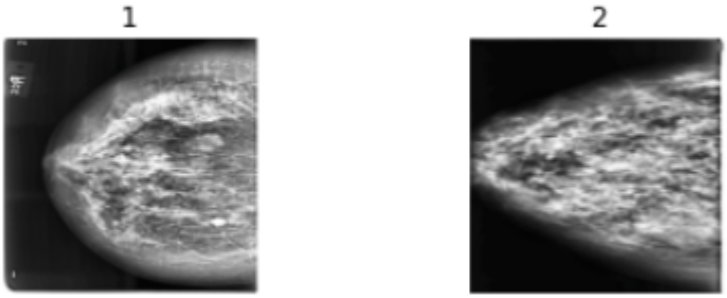


Figure 14: After contrast limited adaptive histogram equalization (CLAHE).

One of the limitations is the quality standard of the mammography, the size of each mammograph is different, it could be caused by the mammography capture by different practitioners or the situation of different patients and machines. We have resized the breast sample images into a smaller image since it will be easier for the low capability processor to process a large number of images. Image resizing is a critical step as it can help our machine to train models more efficiently. As the CNN model is relatively slow to train and a high memory task. However, the downside of downsizing image could lead to information lost. The result demonstrated the basic idea of using resized and contrast images to do the classification. Perhaps different parameters in model such as number of iterations, different sizes, input of higher resolution images, other algorithms (eg. VGG16, VGG19, ResNet, Inception and Xception with Keras) and content layer or better computational capability would increase the accuracy and reduce the effect of image downsizing.

In the image pre-processing stage, we applied 3 main steps which are cropping the unnecessary background, image normalization and contrast limited adaptive histogram equalization (CLAHE) which enhance black and white image contrast. The model performance was assumed to be improved. As in our first experiment by purely using cropped images (see Figure 7), the model cannot learn very well because the algorithm was unable to find a good pattern. Then, with the help of normalizing and contrasting images (see Figure 8 and 9), the model eventually produces relatively good results since the mass lesions appear much more clearly. So, we assume applying more image enhancement techniques will be favourable to the model learning process. There is some image synthesising composed of the truncated or normalised image or other image processing techniques that could be done so the prediction result can be more accurate.

In the ethical perspective, there are reliability issue when we were developing the artificial intelligence model in clinical (Mendelson, 2019). It will lead to liability issues when evaluating complex medical cases. Furthermore, the use of artificial intelligence model needs to consider ethical concerns to facilitate transparency and reliability. As Geis et al (2019) mentioned the simple production and AI models development, which emphasises the importance of understanding complexity on radiology. There are only a few studies that have evaluated how artificial intelligence can be used in a responsible manner to maximise its impact ethically. So further evaluation of the accuracy and reliability of artificial intelligence is also needed to ensure deep learning models to be used confidently for clinical application.

Future Works

Firstly, there are different approaches or solutions that can be applied to do the classification. It is still an open possibility to try different machine learning algorithms with a different mixture of pre-processing and filtering techniques. For the step of training and validation are also essential when we are developing a clinically related model. Since public citizens might have to worry about the reliability and fairness of AI practices in terms of equity in diagnosis and treatments. Therefore, data scientists or developers have responsibility to prove the model reliability and ethically responsible.

Secondly, there are image quality issues when we are dealing with the mammography data. For example, the image quality is not standardised and some of the images can add difficulties to training. The Full Field Digital Mammography (FFDM) is commonly used in current practices, and it needs to enhance the screening technologies for better quality image. In the near future, some emerging screening technologies such as Digital breast tomosynthesis (DBT) are getting popular. As a result, the machine learning algorithm or AI must be able to handle 3-dimensional data with the use of greater computational power.

Thirdly, it is crucial for the model developers and radiologists to sit and work together to lessen the knowledge gaps from both sides. It will improve the reliability and more easily be approved if the AI model can be used as a clinical tool. Additionally, there can be different types of classes for breast tumours (normal, benign and malignant case and stage 1 to stage 4), models can be trained to be more complex and able to classify all classes of breast tumour. Even if AI tools have been approved and implemented, AI algorithms need to be continuously monitored for possible improvements and security risks in order to reduce the probability of imaging data corruption.

Conclusion

In summary, we have conducted a machine learning approach for classifying breast cancer through the CBIS-DDSM mammography dataset. Given the complexities and limitations such as data quality, computational limitations, our model reached 64% accuracy. During the literature review process, we learnt different pre-processing techniques to enhance the signal of image data. In our model, we have applied cropping, resizing, normalisation as well as contrast improvement through contrast limited adaptive histogram equalization (CLAHE). Lastly, we use grad cam to compare the ROI with the classification result. It shows our model can identify the problematic tissues. However, it still needs further investigation for better understanding of the breast tumour.

We have demonstrated honesty by citing all the used references and have applied critical analysis skill in evaluating the pre-process steps, results, and possible limitations. It is ensured that the model is reproducible as we make sure the notebook and required data are zip and stored properly for the reader(https://github.com/chakhoho/IFN646project).

It is recommended that the use of mammogram data sources must have better quality mammography and sufficient for training. Since a clear image can be able to let the algorithm learn much easier. Also, different image pre-processing, image signal enhancement techniques and machine learning algorithms can be deployed and tested. It is because different combinations of techniques might perform differently and learn efficiently.

References

[1] What Is a Benign vs. Malignant Tumor? (2022, October 17). Verywell Health. https://www.verywellhealth.com/what-does-malignant-and-benign-mean-514240

[2] Le, E., Wang, Y., Huang, Y., Hickman, S., & Gilbert, F. (2019). Artificial intelligence in breast imaging. Clinical Radiology, 74(5), 357–366. https://doi.org/10.1016/j.crad.2019.02.006

[3] GeeksforGeeks. (2022, August 24). VGG-16 | CNN model. https://www.geeksforgeeks.org/vgg-16-cnn-model/

[4] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights Into Imaging, 9(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

[5] Nan Wu, Zhe Huang, Yiqiu Shen, Jungkyu Park, Jason Phang, Taro Makino, S. Gene Kim, Kyunghyun Cho, Laura Heacock, Linda Moy, & Krzysztof J. Geras. (2020). Reducing false-positive biopsies with deep neural networks that utilize local and global information in screening mammograms. ArXiv: Image and Video Processing.

[6] Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., ... & Kohli, M. (2019). Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Radiology*, *293*(2), 436-440.

[7] Le, E., Wang, Y., Huang, Y., Hickman, S., & Gilbert, F. (2019, May). Artificial intelligence in breast imaging. *Clinical Radiology*, *74*(5), 357–366. https://doi.org/10.1016/j.crad.2019.02.006

[8] Mendelson, E. B. (2019). Artificial intelligence in breast imaging: potentials and limitations. *American Journal of Roentgenology*, *212*(2), 293-299.

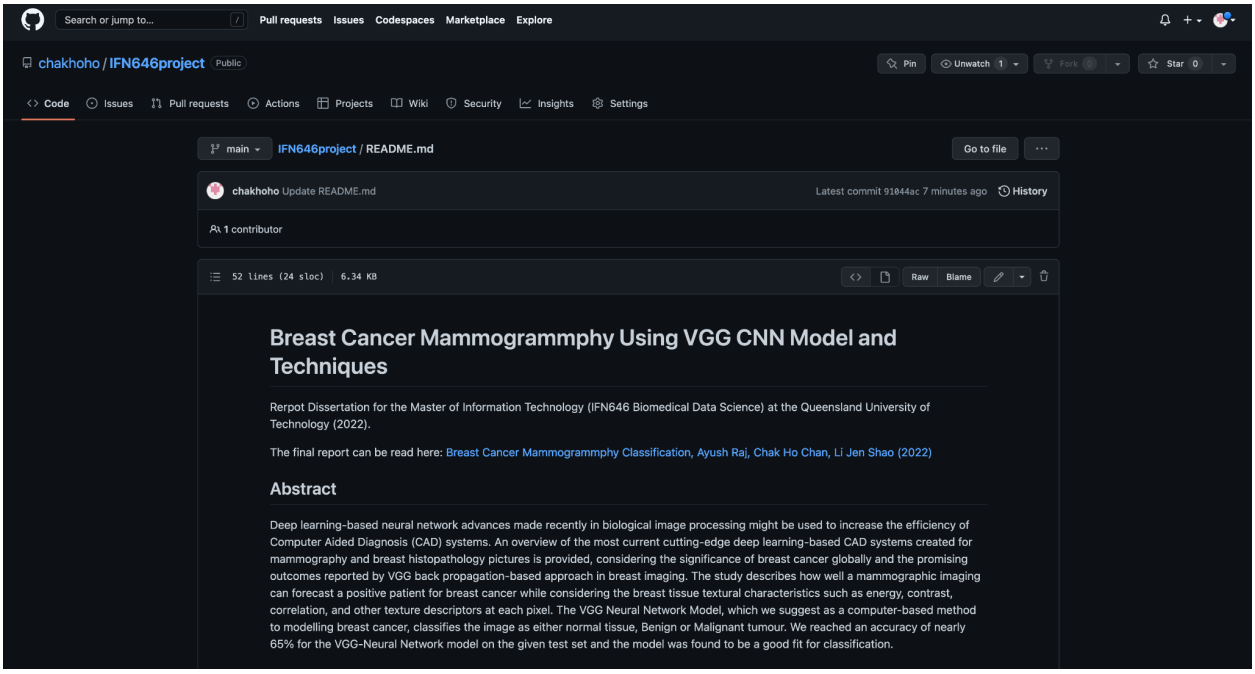
[9] Mohamed, A. A., Berg, W. A., Zuley, M. L., Sumkin, J. H., & Wu, S. (2019, November 19). Deep learning modeling using normal mammograms for predicting breast cancer risk. *Medical Physics*, *47*(1), 110–118. https://doi.org/10.1002/mp.13886

[10] Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., ... & Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, *11*(4), 193-200.

[11] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Appendices

Appendix 1: All the resources can be found in the GitHub (https://github.com/chakhoho/IFN646project).



Appendix 2: 48 hours extension