

Image Registration and Machine Learning for Mammography



Final Year Project
By Lucas Meader

Department of Electrical & Electronic Engineering
Faculty of Engineering and Physical Sciences
University of Surrey

Abstract

There is a strong need for a reduction of unnecessary biopsy referrals in the area of X-ray mammography screening. One way to achieve this is for an improved tumour classification system so that radiologists have greater confidence in non-referral of some patients for biopsy. Deep convolutional neural networks have seen huge improvements in recent years and although their application to mammography tumour classification has seen some progress it is still an unsolved problem. Lack of labeled data is seen to be one of the main issues hindering progress. This report covers research into whether a new image registration technique can be used to locate tumours in full-slide mammograms using crops from spot compression images. Although the full automation of tumour location and extraction process was shown to be of little value to the machine learning pipeline, a semi-automated approach is proposed, tested and shows to be significantly more effective. Furthermore, a deep learning model using transfer learning is proposed and tested with results of: AUC 82%, Sensitivity 0.79%, Specificity 65% and PPV 70%.

Table of Contents

ABSTRACT	2
1. INTRODUCTION.....	4
1.1 BREAST CANCER.....	4
1.2 MAMMOGRAPHY.....	4
1.3 SCREENING	8
1.4 COMPUTER AIDED DIAGNOSIS WITHIN MAMMOGRAPHY	9
1.5 AIMS AND OBJECTIVES.....	10
2. DEEP LEARNING	11
2.1 INPUTS AND OUTPUTS.....	11
2.2 NEURAL NETWORKS	12
2.3 PERCEPTRON.....	12
2.4 CONVOLUTION	13
2.5 ACTIVATION LAYERS.....	14
2.6 POOLING	14
2.7 SOFTMAX	15
2.8 BACKPROPAGATION.....	15
2.9 DEEP NEURAL NETWORKS.....	16
2.10 IMAGE AUGMENTATION.....	16
2.11 TRANSFER LEARNING	17
2.12 EVALUATION METRICS	18
3. DEEP LEARNING IN MAMMOGRAPHY	19
3.1 TUMOUR CLASSIFICATION	19
4 PREPROCESSING AND IMAGE REGISTRATION	20
4.1 DATA SOURCE.....	20
4.2 BENIGN/MALIGNANT SORT	21
4.3 SPOT CLASSIFICATION SORT.....	22
4.4 MOVING OF JSON FILES	22
4.5 IMAGE PAIRING	23
4.6 EXTRACT PATCH FROM SPOT COMPRESSION VIEW.....	24
4.7 MOVE ‘LARGE AREA’ FILES.....	25
4.8 EXTRACT COORDINATES FROM IMAGEDB FILES	25
4.9 IMAGE REGISTRATION	26
4.10 SEMIAUTOMATED PATCH EXTRACTION	27
4.11 IMAGE AUGMENTATION TO INCREASE IMAGE COUNT	27
4.12 CONVERTING TO TIF AND ADDING A 2 ND AND 3 RD LAYER TO ALL IMAGES	28
5 DEEP LEARNING	29
6 RESULTS	31
6.1 IMAGE REGISTRATION RESULTS	31
6.2 DEEP CONVOLUTIONAL NEURAL NETWORK RESULTS	33
7 DISCUSSION.....	33
8 CONCLUSION AND PROPOSED FUTURE WORK	34
REFERENCES	35

1. Introduction

1.1 Breast Cancer

In 2015 there were 55,122 new cases of invasive breast cancer in the United Kingdom and in 2016 11,563 people died from breast cancer in the UK alone. (Cancer Research UK, 2018)

Breast cancer causes cells in the breast to grow in an irregular way which may lead to form a malignant tumour. The stage and grade of the cancer are metrics that are used to determine the size of the tumour and whether it has spread to other parts of the body.

Surviving breast cancer depends largely on what stage the cancer is determined to be at the time of diagnosis. Female breast cancer survival rates past five years decreases rapidly from stages 1 to 5, as can be seen in figure 1. (Cancer Research UK, 2018)

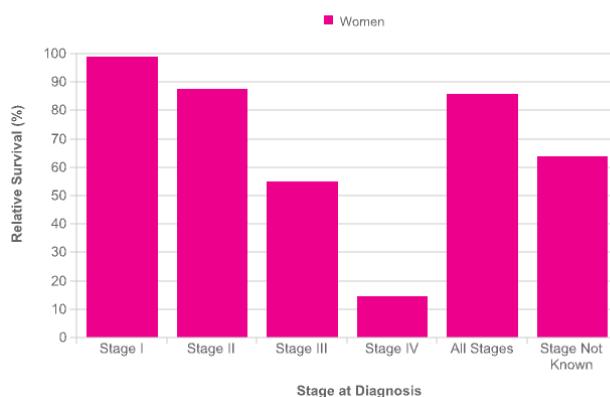


Figure 1 – Bar chart detailing the percentage of women who survive past five years depending on stage of diagnosis. The higher the stage the lower the relative survival. (Cancer Research UK, 2018)

Since the stage of a cancer can worsen over time it is important to diagnose patients as early as possible to improve their likelihood of survival. In the United Kingdom and much of the western world this is achieved by regularly screening patients over the age of 50 using an x-ray imaging technique called mammography.

1.2 Mammography

Mammography is the de-facto gold standard for detecting breast cancer during screening due to its relative low cost and high sensitivity.

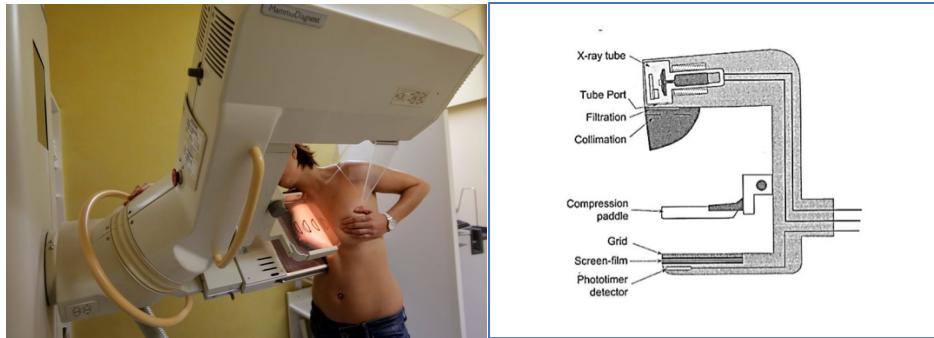


Figure 2 – Left: Mediolateral oblique mammography scan. Right: Components of a mammography machine. (**Health, 2012 (Mammographic Equipment Schematics)**)

During a mammography screening, typically two scans are taken of each breast. A mediolateral oblique (MLO) scan and a cranio-caudal (CC) scan. MLO is a side view of the breast taken at an angle (figure 2, left) and CC is a top down view. The breast is compressed between two horizontal or vertical paddles depending on the scan orientation. Pectoral muscle can appear in MLO scans showing as large areas of lighter gray (figure 3, right)

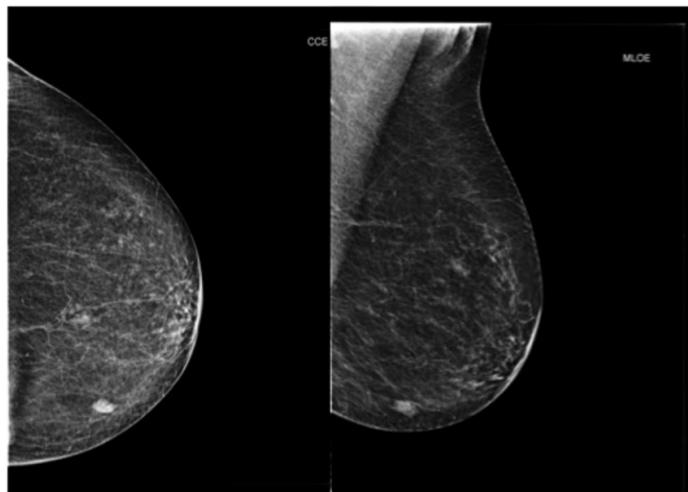


Figure 3 – Left: Cranio-caudal (CC) view mammogram. Right: Mediolateral Oblique (MLO) view mammogram. (**Graziano, 2012**)

The compression spreads the breast tissue out to as wide an area as is comfortable for the patient, allowing for as little ‘overlap’ of tissue as possible. Since the result of a standard mammography exam (mammogram) is the sum of the breasts depth, regions of interest (ROI) could be obscured by breast tissue from areas above and below the ROI.

Once the patient is in place, low-dose x-ray technology is used to penetrate the breast tissue to create an image of the inside of the breast. The image is then examined by a trained radiologist to decide if there are any suspicious areas. If called back, typically a second screening mammography

exam is performed and/or Ultrasound investigation. If both of these prove positive, the woman will be referred for a fine needle biopsy.

Since the original mammography machine, there have been three main areas of improvement; digital mammography; computer-aided detection (CAD) and breast tomosynthesis.

Digital mammography from a patient point of view is very similar to an analogue mammogram. However, instead of producing an image on a film the result is a digital image which can be viewed and manipulated on a computer by the radiologist. Digital mammogram images are stored in Digital Imaging and Communications in Medicine (DICOM) file format. In addition to detailed image data the DICOM file format contains scan and patient information within each image. The information section of DICOM file is referred to as the ‘header information’ and each bit of information is stored next to standardized reference numbers, making it easier to find using coding languages.

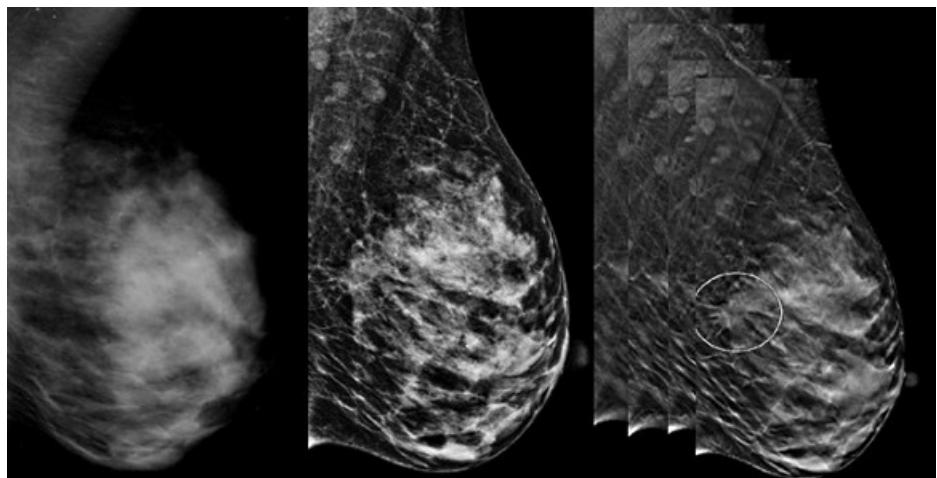


Figure 4 – Left: Screen film mammogram. Middle: Standard digital mammogram. Right: Digital breast tomosynthesis, representation of multiple image that can be traversed through. (**UW Health**)

Computer-Aided Detection (CAD) is software that assists radiologist by highlighting regions of interest within a digital mammogram slide. Regions of interest would be areas of dense breast tissue, abnormal masses, or calcifications.

Digital Breast Tomosynthesis (DBT) is a technique whereby three different x-rays are taken at three different angles. The resulting digital images are synthesized together to create a three dimensional image that can be traversed through on the z, as well as x and y planes. This results in a higher likelihood of smaller calcifications and masses to be found with very little, if any, extra radiation for the patient. Figure 5 depicts the tomosynthesis x-ray tube swing.

In addition to these improvements there are two ways of gaining a clearer picture of a ROI during a mammogram, spot view and magnification view. There are three different techniques to spot view: focal spot view; coned-down spot view and spot compression view. When spot views are combined with magnification views, they are called spot magnification views or paddle magnification view (Dronkers, Hendriks, Holland, & Rosenbusch, 2002)

To achieve a magnification view scan, a couple of adaptations to the screening setup are required. A micro foil x-ray tube is used to compensate for the geometric un-sharpness from the magnification and a special magnification table that raises the breast away from the film plate and closer to the x-ray source. (figure 6) (Dronkers, Hendriks, Holland, & Rosenbusch, 2002) A bright white parameter can be seen on a magnification view scan. This is the outer parameter of the paddle used to hold the breast in place during the scan. (figure 6, right)

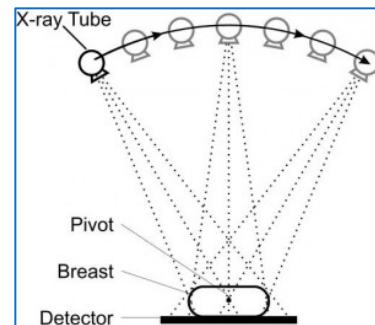


Figure 5 – DBT - The dotted lines represent the three different angles the x-rays are taken at, pivoting on the center of the breast. (Comprehensive Diagnostic Imaging)

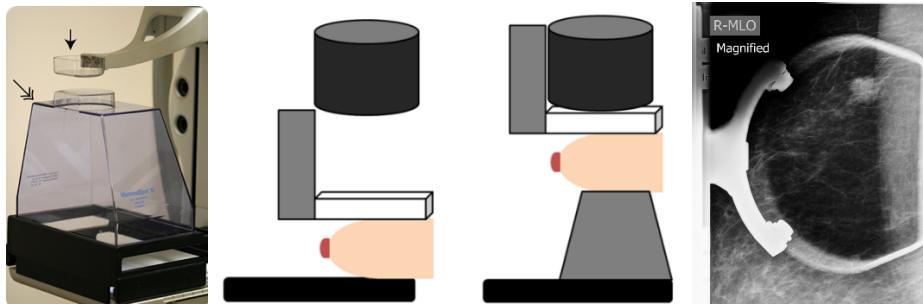


Figure 6 – Left to right: The setup for a magnification view, bringing the breast closer to the x-ray source. The difference between a normal mammogram setup and a magnification view setup. An example of a magnification view scan with the white parameter present. (**mammoguide**) (breastimaging.vcu.edu, 2017) ([Dr Henry Knipe, 2018](#))

The mammography imaging chain:

1. Generation of the x-rays
2. Generation of the radiation image
3. Image improvement through the anti-scatter grid
4. Absorption of the radiation
5. Transformation of the image information in the detector
6. Film / image processing
7. Image presentation

1.3 Screening

Breast cancer screening every three years for women aged 50 - 64 was introduced in the United Kingdom in 1988 following the Forrest Report (Forrest, 1986, p. 68). This was the first screening program of its kind, and now women aged 50-70 are invited for screening. The screening program was introduced as a result of a report commissioned by the Health Ministers of England, Wales, Scotland & Northern Ireland and delivered by a working group chaired by Professor Sir Patrick Forrest (Forrest, 1986). The report concluded that deaths from breast cancer could be reduced by a third or more in women aged 50-64 if screening by mammography was offered (Forrest, 1986, p. 68). It is estimated that the UK's national breast screening program prevents 1300 deaths a year, "This corresponds to one breast cancer death averted for every 235 women invited to screening for 20 years, and one death averted for every 180 women who attend screening" (Screening, 2012). Although these statistics are positive there are a few drawbacks to regular screening which has led to research into whether regular screening has a net benefit to women or not. There are three associated potential harms to women who attend a screening: diagnosis of slow growing cancerous tumour that would not have posed a threat otherwise (overdiagnosed); exposure to small amounts of radiation which leads to 3 in 10,000 women getting cancer from radiation exposure; and unnecessary anxiety for women called back for more tests which end up being clear (Screening, 2012). Overdiagnosis is the main argument against regular breast screening, this is because for every 1 in 2000 women that are saved from dying there are 10 in 2000 that have to go through unnecessary treatment due to overdiagnosis (Gøtzsche & Jørgensen, 2013). The difficulty at the moment is that we cannot identify which case group an individual might fall into.

An investigation into whether a woman has breast cancer can start in one of two ways; when no abnormalities are found during a scheduled screening they are referred to as *asymptomatic*, as no lump or abnormality was detected before the screening. Conversely, an abnormality such as a lump may be detected at screening referred to as 'screen-detected' or alternatively detected following referral by GP and then get referred for further tests which is called a clinical presentation. If cancer is then found these are referred to as interval cancers if the woman is of screening age and the cancer is detected outside of the usual screening sessions.

Once a mammogram has been performed there are several investigations that can take place depending on the result of the mammogram and patient physiology (table 1).

Investigations	Further Investigations
Magnification View Mammography	MRI / PET Scan
Core-needle Biopsy	CT Scan
Excision (surgical) Biopsy	Bone Scan
FNA (Fine Needle Aspiration)	Sentinel Lymph Node Biopsy
Ultrasound	Stereotactic Biopsy

Table 1 – List of investigation and further investigations.

A combination of one or more investigations can be used to determine the diagnosis. Depending on the size, location and appearance of the suspicious area, one of a range of biopsies would be required. These are very invasive procedures that require a needle to be passed through the skin or an incision made, and a sample taken of the tumour. The sample is then examined for abnormalities to determine whether the tumour is cancerous. If the tumour is cancerous the stage and grade of the cancer is then decided upon. It is only at this stage that a tumour can be categorically classified as cancerous or benign (National Breast Cancer Foundation Inc, 2016).

There are five main treatment options that may follow, depending on the stage and grade of the cancer; Surgery, Chemotherapy, Radiotherapy, Hormone therapy and Biological therapy.

1.4 Computer Aided Diagnosis within Mammography

Computer Aided Detection (CADe) systems are widely used to assist radiologists in detecting and locating abnormalities within mammography images. Although these systems have proven useful, they are by no means ready to do the job of a radiologist.

The current standard diagnostic pathway consists of two radiologists independently assessing the same mammogram, referred to as ‘double reading’, and a decision rule or third party is used when the radiologists disagree [Thurfjell 1994]. Double reading has a significant increase in sensitivity over single reading [Thurfjell 1994]. However, there is still no consensus on whether a single reader using CADe can perform better in terms of sensitivity and cancer detection rate than double readers without a CADe system [Henriksen EL 2019].

Computer Aided Diagnosis (CADi) systems are less developed due to the complex nature of the problem. The development of deep learning classification models (described in section 2) will be the most likely catalyst for advancement in this area.

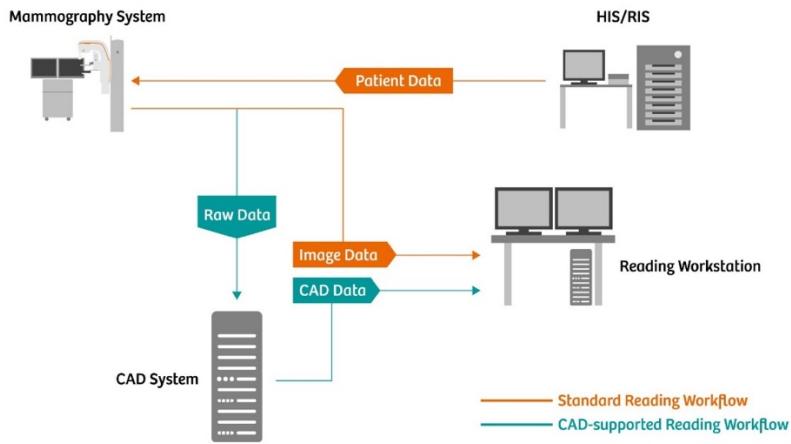


Figure 7 – The use of CAD in the mammography workflow. (Siemens, 2018)

In July 2018, The Royal College of Radiologists (RCR) released their position statement on the use of artificial intelligence within radiology as a whole. One of the paragraphs is worth citing here, “The RCR believes that AI potentially represents one of the most fundamental changes in medical care since the inception of the NHS, and strongly welcomes the introduction of appropriately regulated and governed uses of AI related technologies to enhance clinical practice.” This is a noteworthy statement from a well-respected body. It shows how far AI has come and that research and development in this area is necessary, anticipated and, when it is ready it will be revolutionary.

1.5 Aims and objectives

To identify the maximum amount of cancer cases, a threshold criterion for whether to proceed with a biopsy is set high into the true negative region, meaning many women (80% in America) have biopsies that prove to be false positives (National Breast Cancer Foundation Inc, 2016). If radiologists had more confidence in a decision not to proceed with a biopsy, thousands of women would be spared the associated trauma and anxiety. The only way radiologists can have more confidence in a decision not to proceed with a biopsy is for them to have more research to base their decision on.

The purpose of this project is to create a machine learning model that can offer a prediction as to whether a region labelled abnormal in a mammogram is cancerous or benign, based on a region of interest patch taken from a full mammography image. Both benign and cancerous patches are required for the model to ‘learn’ the differences between the many cases. The prediction could then be used in the decision-making process and lead to a decrease in false positive cases.

To extract a region of interest patch from a full mammography scan, the coordinates of the region of interest are required. Currently coordinates are only applied to mammograms that are discovered to be cancerous. Therefore, the first part of this project will be the creation of a dataset containing image patches of benign tumours. To achieve this, an image patch taken from the center of the bright white circle of a magnification view image will be used to locate the tumour in a full breast scan taken from the same angle CC or MLO using image registration techniques. A larger patch will then be extracted from the full breast scan for use in the second part of this project.

The second part of the project will be using the newly created dataset in conjunction with a dataset of patches containing cancerous tumours, to train a deep learning model to classify a tumour as benign or malignant.

2. Deep Learning

2.1 Inputs and Outputs

Grayscale (colourless) images are interpreted by computers as a grid (or matrix) of numbers. Typically each picture cell (pixel) has a value in the range 0 to 255, 0 being the blackest and 255 being the whitest. This is known as a ‘byte’ image because each pixel is stored as an 8-bit integer (2^8), an illustration of this can be seen in figure 8.

Since medical imaging requires the maximum amount of image clarity, DICOM images support all 65,536 (2^{16}) shades of gray. The mammography images used in this project will be DICOM images and as such will be used as the input to the deep learning model.

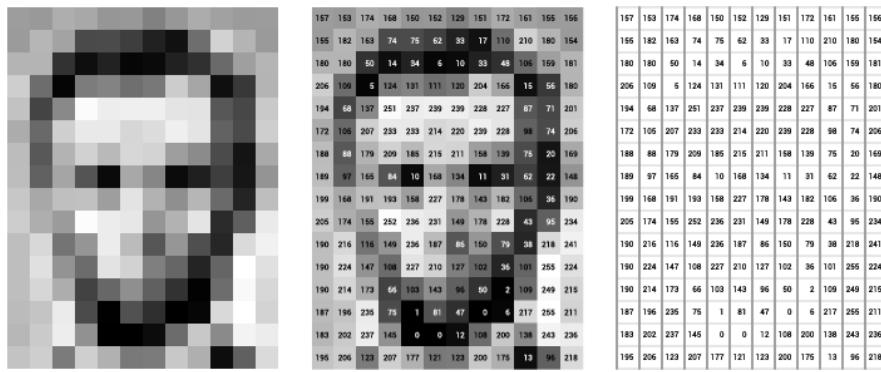


Figure 8 - An illustration of how machines interpret 8-bit grayscale images. Left: how humans see images. Right: how computers interpret images. Center: a combination of both, highlighting how the different shades of gray correspond to numbers. (Lim, 2018).

As discussed in section 1.5, the aim of this project is to give a prediction on whether a ROI is cancerous or benign. Since there are only two outcomes, this is referred to as a binary classifier.

Therefore, the ideal output of the deep learning model would be either 100% prediction of cancer or 100% prediction of benign. Realistically the output will initially be a percentage split between cancerous and benign.

2.2 Neural Networks

Neural networks are the framework that deep learning models are built on. A neuron is a node that holds a value and a network of neurons or ‘neural network’ is characterised by many neurons connected together. The idea for neural networks is loosely based on our understanding of how the human brain works [wizel]. Certain areas light up (“fire”) when they are active (figure 9, left). This is translated to deep learning by giving a neuron a higher number when it is more relevant and lower number when it is less relevant. A visualization of a deep neural network can be seen in figure 9 (right). Each white circle represents a neuron that is ‘firing’ and each arrow represents the output of one neuron connecting to the input of another.

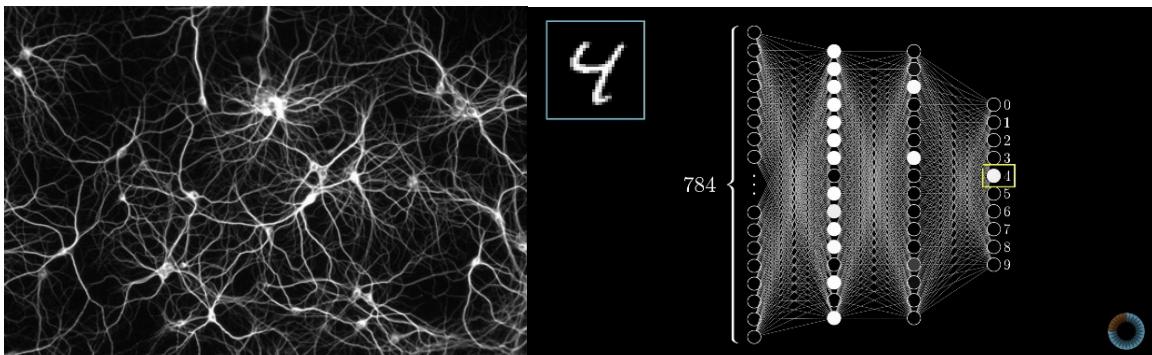


Figure 9 – Left: A visualization of brain neurons firing. Right: A visualization of a deep neural networks neurons ‘firing’ while learning to recognize numbers (Tse & Esposito, 2017) (3Blue1Brown, 2017)

2.3 Perceptron

A perceptron is a single layer neural network. Each node ($x_1, x_2\dots$) of the input layer represents a pixel of the image to be processed (figure 10, right). The pixel values are then multiplied with their respective weights (in this case 0.25). Net input is then calculated using a summation function (S) before an activation function (discussed in section 2.4) is used to determine whether the value of the neuron falls in to one of two categories. In this example, ‘bright’ if the sum is greater than -0.1 or ‘dark’ otherwise (figure 10, right).

As can be seen from this example, perceptron’s are limited in their classification ability due to being linear binary classifiers. This means they are unable to adapt to situations where nonlinear classification is required.

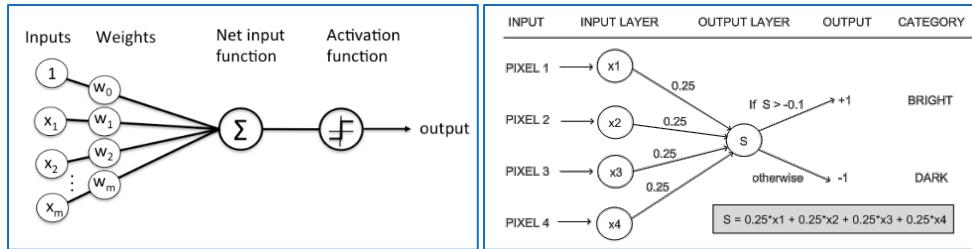


Figure 10 – Left: Overview of a perceptron network. Right: A basic example of a perceptron network. (Sharma, 2017) (Raschka, 2015)

2.4 Convolution

Unsurprisingly, convolution is one of the main building blocks of Convolutional Neural Networks. The purpose of the convolutional layer is to detect different features such as textures, edges and shapes within the input image.

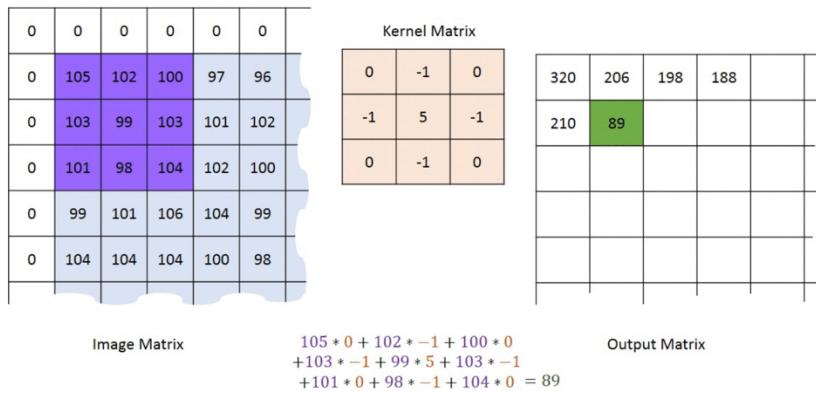


Figure 11 - Example of convolution. (Machine Learning Guru)

Using figure 11 as an example. A 3x3 kernel (also referred to as a window or filter) moves across the image, one pixel ‘stride’ at a time, starting in the top left corner. At each step the value in each purple cell is multiplied with the corresponding value in the peach kernel cell. The sum of the products is then inserted into the output matrix.

The values within the kernel can change so that different features of the original image are detected. An example of an edge detection kernel can be seen in figure 12.

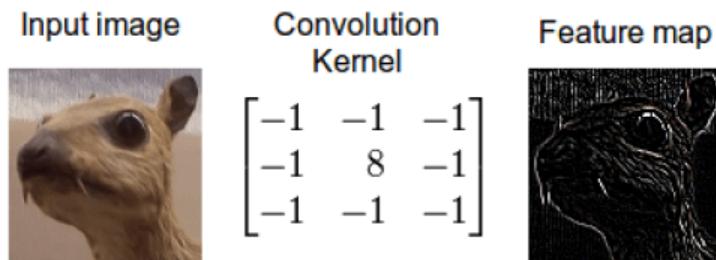


Figure 12 - Example of an image before and after an edge detection convolution has been applied.

2.5 Activation Layers

Neural networks would just be linear regression without activation layers, making them vital to CNNs. There are three activation techniques that have been used in the pursuit of finding the best activation function for convolutional neural networks; Sigmoid, TanH and Rectified Linear Unit (ReLU) are the most prominent. However, due to Sigmoid and TanH suffering from unwanted characteristics, only ReLU and variants of it are prevalent today.

To retain clarity within the convoluted image, the ReLU operation changes all the negative values to zero and keeps all values equal to or greater than zero the same (equation 1, figure 13).

$$f(y) = \begin{cases} 0, & \text{for } y < 0 \\ y, & \text{for } y \geq 0 \end{cases} \quad (1)$$

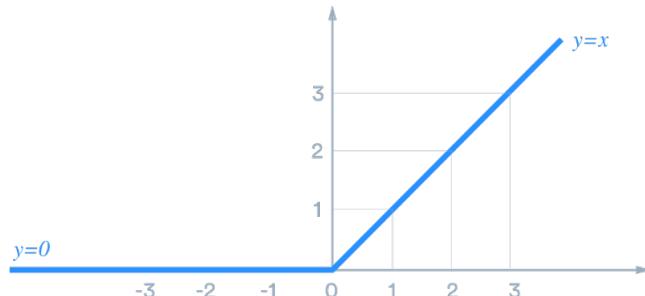


Figure 13 – Graph of a rectified linear unit function. (Liu, 2017)

One drawback of ReLU is that when a neuron gets assigned a zero it is hard for it to be of any use to the network again and over time whole sections of a network could become redundant. One solution to this is to use a variant of ReLU called Leaky ReLU which performs in a very similar way to ReLU whilst retaining some negative information. It achieves this by multiplying negative numbers by a small amount such as 0.01, which achieves a small ‘slop’ instead of discounting the neuron all together.

2.6 Pooling

There are three main pooling techniques: max, mean and sum pooling. The pooling layers within a CNN down-sample the layer before them by extracting the largest number from within the kernel, an example can be seen in figure 14.

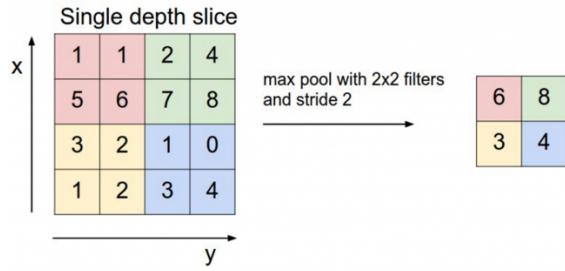


Figure 14 - Example of max pooling. (**Machine Learning Guru**)

A kernel (in this case, 2x2) moves over the image, with a stride of 2, at each stage the maximum value within the kernel's matrix is taken and entered into a new image. Thus, the original image is reduced in size with only its prominent features reserved.

2.7 SoftMax

As discussed in section 1.5, by the end of the model a prediction is required as to whether or not the ROI is deemed to be cancerous. This prediction needs to sum to 1 otherwise the output would be rendered useless. To achieve this, firstly a ‘fully connected layer’ is used. This is a layer that takes every pixel value from the layer preceding it and creates a vector using just those values. All of the values in that matrix are then processed by the SoftMax function which allocates a prediction based on the knowledge that the prediction is split between two outputs, cancerous and benign.

During training, the results of the output are then compared to the label that the image has (for example 1=cancerous, 0=benign). This allows the model to know how well it has done and allow for corrections to the model to take place using backpropagation, which discussed in the next section.

2.8 Backpropagation

For a neural network to ‘learn’ it needs to know two things: what prediction value have I given this image and what prediction value should I have given this image. Using figure 15 as an example, this network is ‘learning’ how to classify numbers. It has an input image of 28x28 (784) pixels. Once the pixels have been processed, they are given a prediction value. In this case the number ‘2’ is being processed and the model thinks there is a relatively low likelihood (0.2) that the image is a ‘2’. Since the model is being trained it knows what the value should be and can use this knowledge to go back through the network and correct all the weights and biases for each connected neuron in favor of the ‘2’ prediction being higher and the other number predictions being lower. The system uses thousands of images to adjust the weights and biases of every connection until the accuracy of the system is deemed sufficiently high.

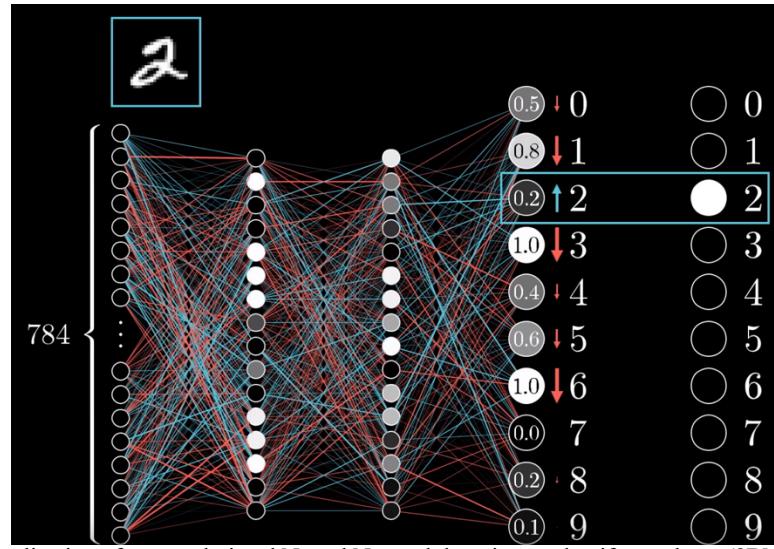


Figure 15 – Visualization of a convolutional Neural Network learning to classify numbers. (3Blue1Brown, 2017)

2.9 Deep Neural Networks

Deep learning is a method for connecting many different combinations of the convolution, activation and pooling layers then repeatedly processing training data until the algorithm ‘learns’ what weights and biases accurately represent the many features of each respective class. New images can then be processed by the algorithm and assigned a classification based on the learnt patterns.

Deep learning has become much more viable over the last few years, in part due to the availability of powerful Graphics Processing Units (GPUs) enabling the processing of vast amounts of data in a timeframe that is acceptable for researchers. Another factor that has contributed to the viability of deep learning is GPU libraries that enable processing parallelization such as CUDA and OpenCL along with open source software packages for efficient GPU implementation – Caffe, TensorFlow, Theano and Torch.

Although Deep Convolutional Neural Networks (DCNNs) have proven themselves as excellent natural image classifiers (Rawat & Zenghui, 2017), the need for large amounts of training data has limited their progress in the field of medical imaging. Image augmentation and transfer learning have been the main antidotes to lack of training data.

2.10 Image Augmentation

DCNNs cannot learn spatial invariance. Therefore, they cannot tell the difference between images that are the same aside from the location and orientation of the ROI. Figure 16 shows an example

of how one image can be augmented into 20 images that will appear to be different images to the DCNN.

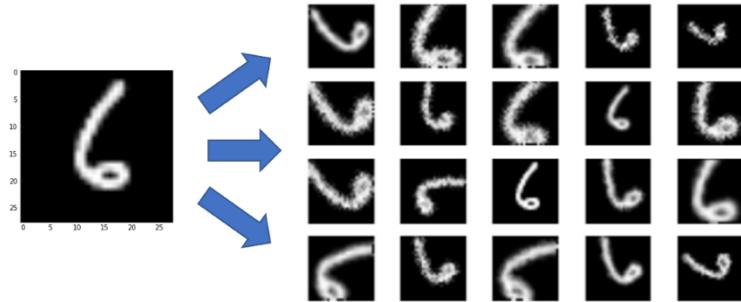


Figure 16 – Visualization of an image being augmented to represent many images. (Raj, 2018)

This means image augmentation techniques such as rotation, scaling and translation can be used to convert a dataset of hundreds of images into a dataset of thousands, this helps the DCNN to generalise. The most popular image augmentation techniques within deep learning are:

- Flip
- Rotation
- Scale
- Crop
- Translation
- Gaussian Noise

These techniques can be used on their own or in any combination.

2.11 Transfer Learning

Transfer learning is the process of using already learnt generic features, such as lines and curves, from DCNN's that have learnt from hundreds of thousands of images unrelated to a given field.

It is known that generic features are found towards the start of the hidden layer section of a network. As one moves through the network the features at each layer become increasingly image specific (figure 17). Therefore, when large amounts of data are unavailable to train a given network, it is a popular technique to extract the start of a network that has highly trained generic features and use it as the start of an image specific network. This results in a highly accurate field specific network that only requires hundreds of training images rather than thousands.

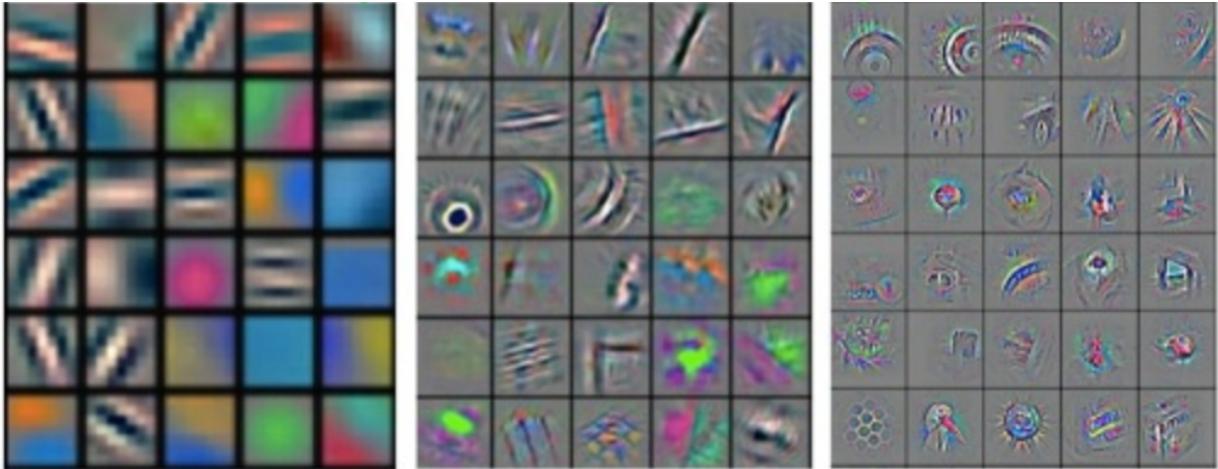


Figure 17 – Example visualization of low, middle and high-level features. The deeper into the network one goes the more detailed and image specific the features are. (Zeiler & Fergus, 2013)

2.12 Evaluation Metrics

The most common method of CNN evaluation is to measure the ‘sensitivity’ (equation 2) which is the ability of the model to accurately predict ‘true positives’ and ‘specificity’ (equation 4) which is the ability of the model to accurately predict ‘true negatives’.

$$Sensitivity = True\ Positive\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$False\ Positive\ Rate = \frac{False\ Positive}{False\ Positive + True\ Negative} \quad (3)$$

$$Specificity = 1 - False\ Positive\ Rate \quad (4)$$

‘Sensitivity’ is then plotted against ‘specificity’ which generates a Receiver Operating Characteristic (ROC) curve then the Area Under the Curve (AUC) is used to measure the effectiveness of the model. The larger the AUC the better the model is performing.

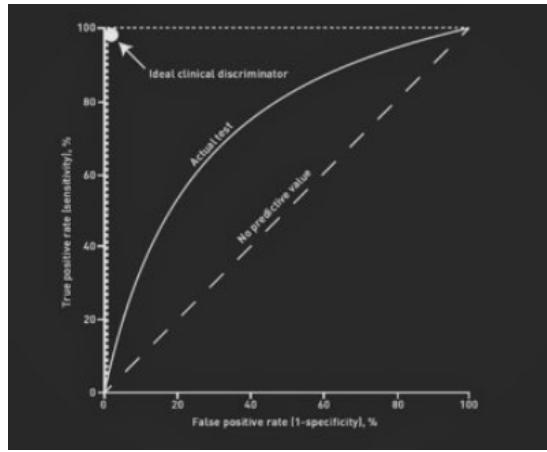


Figure 18 – Example Receiver Operating Characteristic. (Chan, Carmen. 2017)

Another popular evaluation metric is Positive Predictive Value (PPV). This is a measurement (in this research) of the number of cancerous subjects that are positively predicted as having cancer measured by the number of ‘true positives’ divided by the total ‘true positive’ and ‘true negative’ predictions (equation 5).

$$PPV = \frac{\text{True Positive}}{\text{True Positive} + \text{True Negative}} \quad (5)$$

3. Deep Learning in Mammography

Deep learning in mammography covers a few different nested topics, namely, tumour detection, segmentation and classification. Papers and articles published on these topics tend to use different datasets and testing modalities making it difficult to compare them head-to-head. Therefore, due to the nature of this project, focus will be on the techniques that have been used to classify mammography tumours over the last few years and where the field stands today.

3.1 Tumour Classification

As discussed in previous sections, the ability to assign a probability of malignancy to a given abnormal mass could have a large impact on a doctor’s confidence to not refer a patient for a biopsy and thus lead to a reduction in unnecessary intrusive biopsies for false positives (Elmore, 2009). Like many other fields, it is only recently that ‘hand-crafted’ features have started to be replaced by deep learning models that determine their own features. Examples of image features could be the number of white pixels above a certain threshold or whether a given shape exists within an image.

Because tumours are inherently complex it is an impossible task for a human to explicitly define a set of features that consider every eventuality. Nonetheless, just as a human brain can be trained to detect differences in more simple images through being shown many examples, machines can be trained in the same way. Using such an approach, a sufficiently sophisticated system can process and ‘learn’ from thousands of images and deduce their own complex feature set which may produce performance with better accuracy than a human.

A study in 2015 found that switching from hand-crafted techniques to a CNN with only two layers, for the purpose of classifying mammography tumours, improved their results from AUC = 0.79 to AUC = 0.86 (J. Arevalo, 2016). Since then deep learning combined with transfer learning has proven itself as the ‘go-to’ combination for classification tasks with minimal amounts of data.

The main element limiting deep learning’s effectiveness in medical imaging is its need for large amounts of labeled training data. Due to the rapid rise in deep learning, hospitals and health centres have been unable to provide large amounts of labeled data because they had not envisioned a need for it. Therefore, techniques have been developed to create state-of-the-art classifiers with as little as a couple of hundred labeled images. Known as ‘transfer learning’ its effectiveness has been researched in several medical imaging domains and proven to be a very useful tool (Shie et al, 2015).

4 Preprocessing and Image Registration

4.1 Data Source

Mammography images and related data from the OPTIMAM image database were uploaded to a Google Bucket (online storage platform) for transfer by a Principal Computer Scientist at the Royal Surrey County Hospital. Due to the size of the data (752GB) it was downloaded to Centre for Vision Speech and Signal Processing servers at the University of Surrey for processing. The data comprises of 1456 subject files containing a total of 91,204 objects which were received in the following file structure:

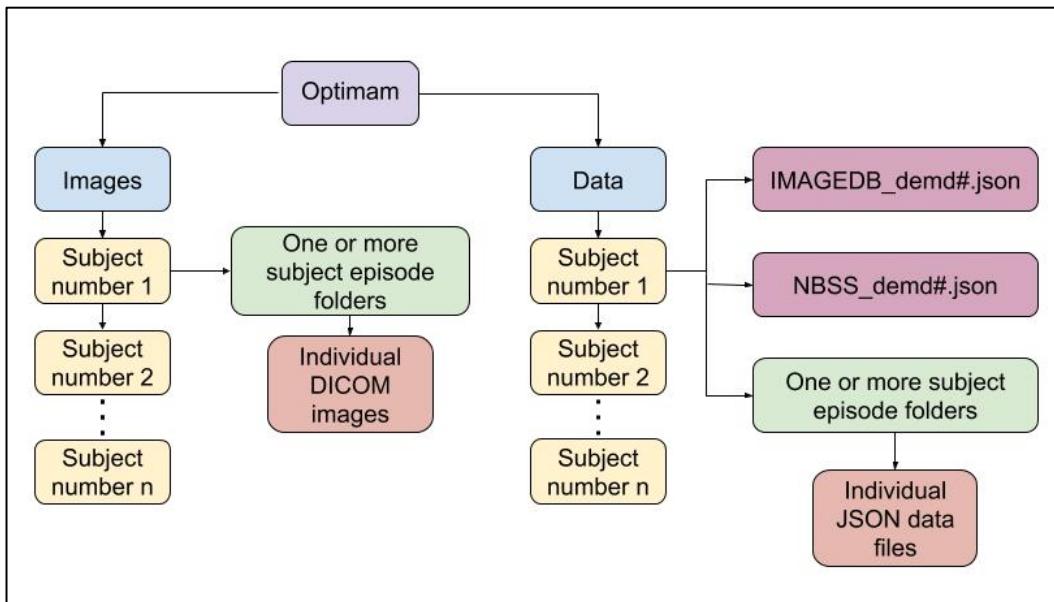


Figure 19 – The file structure the data was received in.

The top-level data and image folders contain matching subject folder numbers so that images can be easily linked to their corresponding data files. The Data subject episode folders correspond with matching image episode folders. Every DICOM image has a corresponding JSON data file. Each subject's data folder contains an Image Database (IMAGEDB) file and a National Breast Screening System (NBSS) file. The IMAGEDB file contains coordinates of tumors for specific images and the NBSS file contains information describing the diagnosis of the subject. All images are 2560x3328pixels in size.

4.2 Benign/Malignant Sort

Due to the aim of the project being to classify tumors as benign or cancerous, the data was first separated into benign and cancerous folders to clearly define the subjects from the start.

Using data from the NBSS files, the left and right breast opinion at every episode (stage) was checked for one of two values: benign or malignant. If there had ever been an episode in which the subject was diagnosed as malignant the subject was moved into the malignant folder. If there had never been a malignant episode the subject was moved into the benign folder. The sorting process resulted in; 535 benign subjects and 921 cancerous subjects.

Initially the classification at the end of the file was used but was later found to not be indicative of the opinion at each episode of the subject's treatment path i.e. a tumor could be diagnosed as cancerous; the tumor could then have been removed, and the subject would then be classed as

benign at the end of the file. This led to subjects being wrongly classified as benign in the sorting process.

The misclassification was overcome by checking the opinion at each episode rather than the overall classification located at the end of the file.

4.3 Spot Classification Sort

There are four types of image; full breast, magnification, biopsy and spot compression. Full view images are mammograms of the whole breast; Magnification images contain large areas of the breast (figure 19) limiting its usefulness in tumour positioning; biopsy images did not present enough breast tissue area and were largely whole breast mammograms not suitable for image registration, and spot compression images which present clear areas of breast tissue. Therefore, the decision was made to only use spot compression images for the remainder of the research.

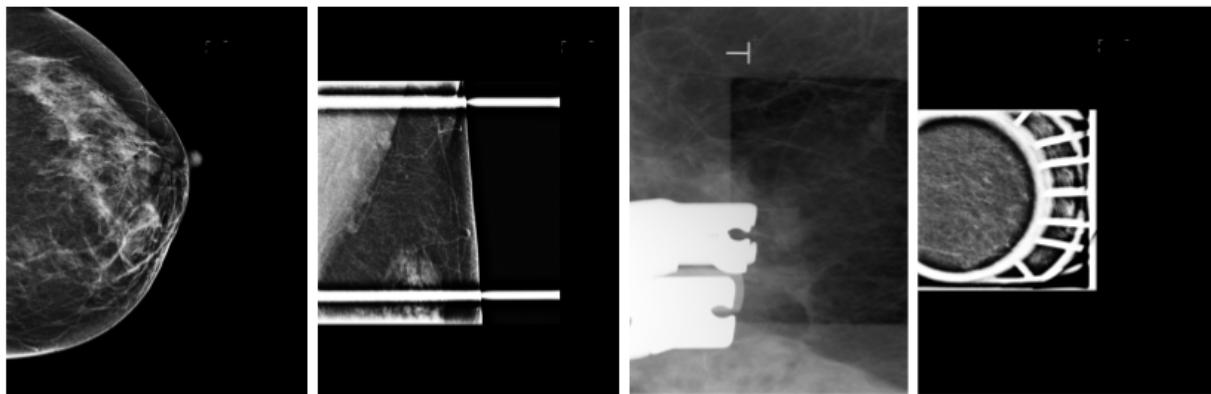


Figure 20 – Left to right: Full breast, Magnification, Biopsy and Spot Compression mammograms.

All subjects were organised into their corresponding magnification, biopsy and spot compression folders by extracting the ‘View Modifier Code Sequence’ key value pair from the DICOM header information of each none full view image. These were one of three values: Spot Compression, Magnification or Other. The ‘Other’ value was observed to actually be biopsy images and therefore each subject was moved into one of either the Spot Compression, Magnification or Biopsy folders.

4.4 Moving of JSON Files

For ease of reference, the IMAGEDB and NBSS JSON files were moved from each subjects data folder (figure 19) to each subjects Image folder (figure 21). Due to the data being sorted in the same way as the image folders, a lot of moving in and out of folders was required in the script to locate and copy each file successfully.

4.5 Image Pairing

For image registration to work the spot compression mammogram needed to be paired with the full breast mammogram of the same breast (left or right) and the same view (cranio-caudal or mediolateral-oblique). Therefore, each subject's files were organised into the following structure:

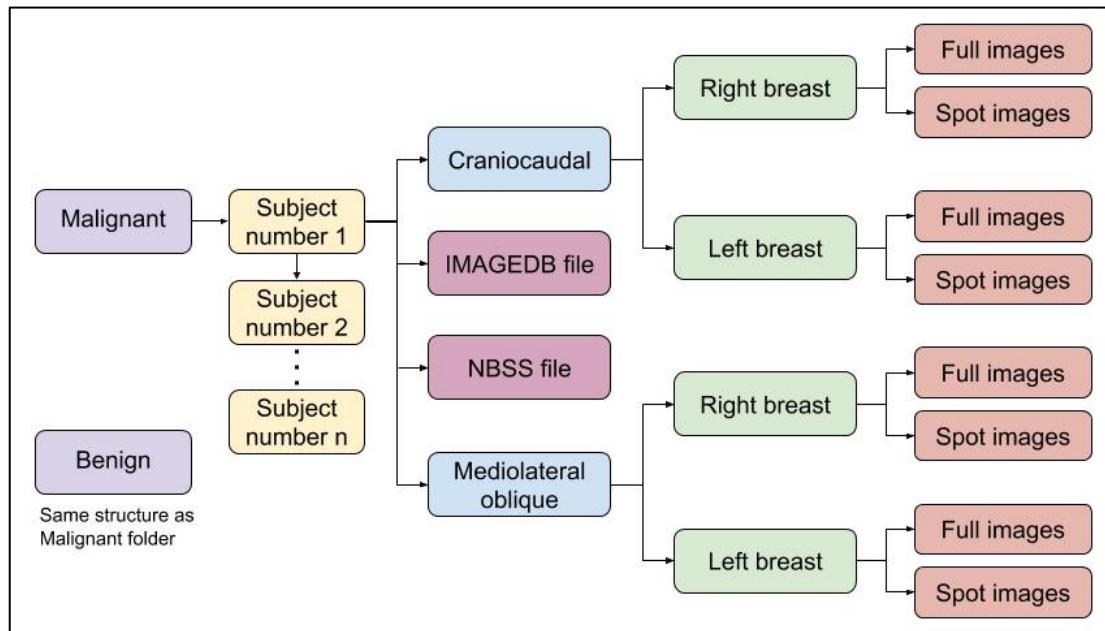


Figure 21 – Data structure after initial sorting process.

Using the JSON data file information, the DICOM header tag 'Series Description' (0008,103E) was located, the tag can have one of four values: 'L CC', 'R CC', 'L MLO', 'R MLO'. Once the orientation (Left or Right) and the view (CC or MLO) had been determined the respective folders were created as in the structure seen above (figure 21). If the 'Series Description' tag was not located, then two other tags were located instead, (0020,0062) nested under parent tag 'Patient Orientation' (0020,0020) which contained values 'L' or 'R' for left and right respectively and (0008,0104) under parent tag 'View Code Sequence' (0054,0220) which in the case of mammography contains either 'cranio-caudal' or 'mediolateral-oblique'.

Once the orientation and view had been determined the relevant folders were created and the files were then moved into them; This was repeated for every subject.

Having learnt afterwards that the Series Description tag is optional but Patient Orientation and View Code Sequence tags are required it would be more efficient to check for those at an earlier stage. Some image files had '.dcm.dcm' as their file extension as opposed to just .dcm which confused the algorithm.

4.6 Extracting Patches from Spot Compression Images

Initially it was thought that spot compression mammograms all have similar forms, that is a 2560x3328 image in which a highlighted circle is visible (figure 22, right). The highlighted circle contains clear breast tissue which can be extracted from the image as a new image patch. A script could then be written to locate the circle and extract the center automatically. However, spot compression mammograms can take various forms and be of various sizes (figure 22). Therefore, a script was written which semi-automated the process. The script finds the processed spot compression mammogram in each subject's CC or MLO subfolder, reads the DICOM file, displays the image to the user who can then manually select a crop area. The cropped area is then saved in a new folder called Processed Pair.

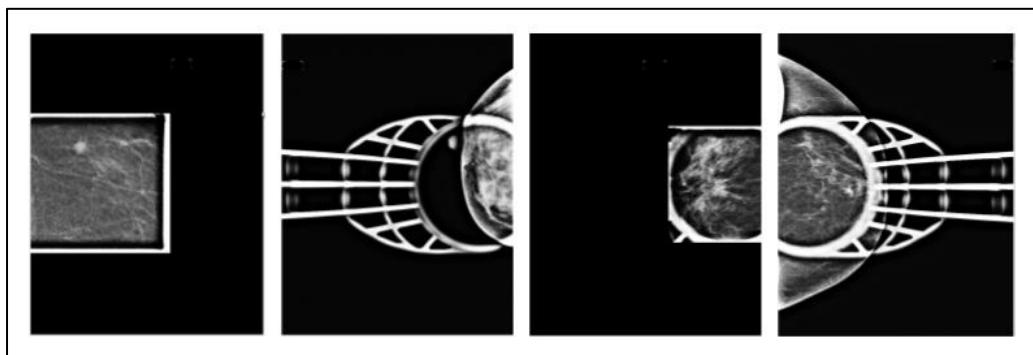


Figure 22 – Examples of some of the different forms spot compression images can take.

The corresponding processed full view mammogram image is then copied from the full image folder to the processed pair folder, resulting in the new file structure:

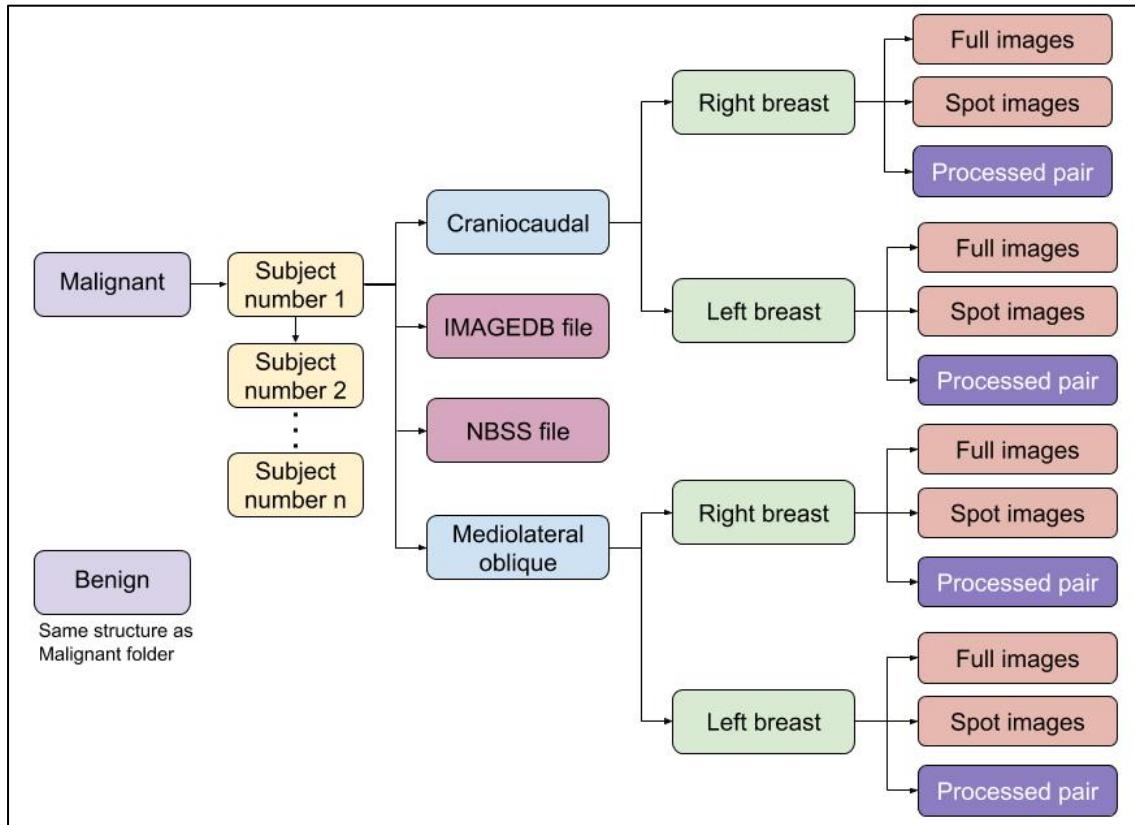


Figure 23 – Data structure after full and spot processed images had been paired.

4.7 Move Large Area Files

Some spot compression images contained large areas of the breast which would not have assisted in tumour location. During the semi-automated patch extraction stage, the ‘Processed Pair’ folder was only created if the spot compression was of the desired form. This meant that any folder without a Processed Pair subfolder was not required and was moved to a separate ‘Large Area’ folder.

4.8 Extract Coordinates from IMAGEDB Files

As discussed in section 4.1 the coordinates required for locating the ground truth location of the lesion/suspicious area are located in the Data folder and thus required locating, extracting and moving to a new file located within each Processed Pair folder. The coordinates for each image were heavily nested in non-uniform ways making it tricky but necessary to extract them into a new file rather than use them in their existing form.

4.9 Image Registration

Each malignant subject has at least one full breast mammogram which has coordinates X1,Y1 (top left) and X2,Y2 (bottom right) (figure 24, middle) which have been recorded in the Image Database file of each subject by a doctor, the coordinates encapsulate the region of interest in a rectangle. Resulting in a ground truth tumour location for each image. For this reason, all image registration tests were performed on malignant images.

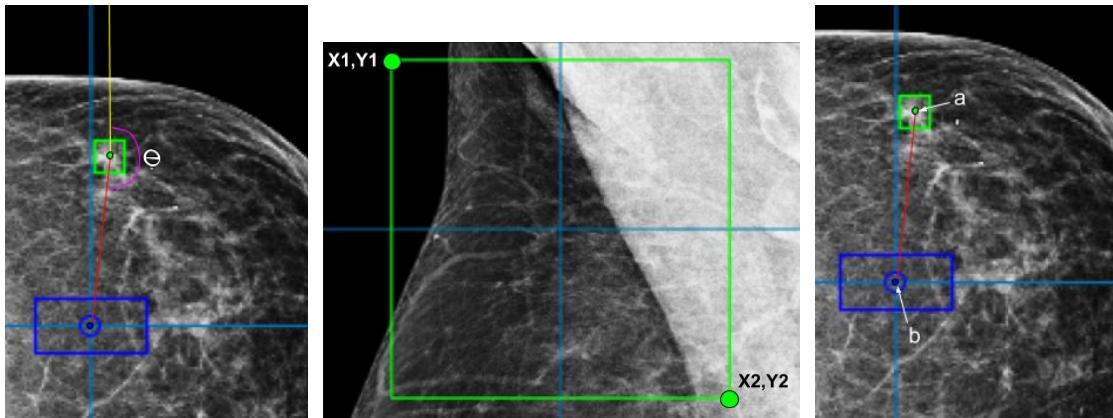


Figure 24 – Left: visualization of the angle being calculated for each image registration attempt. Middle: example of what X1,Y1 & X2,Y2 extracted from each IMAGEDB file are referring to. Right: Example of a distance measurement from a ground truth coordinate (a) and an image registration maximum correlation point (b).

Two image registration techniques were tested; Sum of Square Difference (SSD) and Normalized Cross Correlation (NCC). To evaluate the performance of the two techniques, the distance between the center of the coordinate rectangle and the maximum correlation point produced by the image registration were measured and recorded (figure 24, right) along with angle between the coordinate images vertical plane and the maximum correlation point for every image pair (figure 24, left). Although SSD is computationally efficient it is less robust to noise than NCC.

The image registration process worked well for 354 of the 627 image pairs, however three issues were observed while performing the semiautomated patch extraction step.

1. Not all spot compression images contained lesions. This means the image registration could yield a perfect result, but the patch is useless for training purposes.
2. Some image registration attempts failed due to no observable similarity between the full view mammogram and the spot compression mammogram.
3. Some of the correctly registered images had lesions that lay outside of the crop area resulting in useless image patches for machine learning purposes.

4.10 Semiautomated Patch Extraction

Due to the drawbacks of the image registration script results a solution was required so that the maximum amount of benign tumours could be extracted for the machine learning stage of the project.

Therefore, a script was created which built upon the image registration script by displaying the result with an interactive cross hair, enabling the user to manually select the tumour within the image while observing the crop and the maximum correlation point from the image registration (figure 25). The selection generates an x,y coordinate centered on the tumour. A 227x227 patch is then automatically extracted from the full view image and saved in a new folder ready for additional processing.

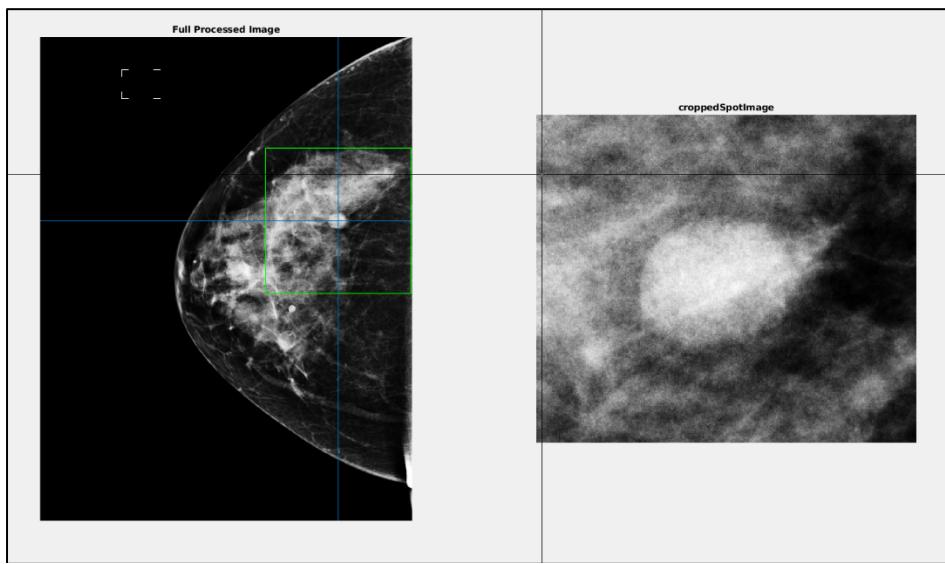


Figure 25 – Example of semi-automated patch selection. The black crosshair is interactive, ready to select the center of the tumour in the full view image.

When the user selects an area close to the edge of an image the crop could result in an image which is not 227x227 for example 190x227 which is not the correct size for AlexNet.

To solve this problem the script was modified to work out the distance (depending on what edge the selection has overlapped) from the edge to the center and move the extraction square just the right amount to maintain the 227x227 patch size.

4.11 Image Augmentation to Increase Image Count

After data processing, only 128 benign images were viable for machine learning and therefore image augmentation was required to boost the numbers of images suitable for training the machine learning model.

Each original image was rotated 90°, 180° and 270° degrees. Each original image was also flipped on its vertical axis and then rotated 90°, 180° and 270° degrees (figure 26). This technique produced an extra 7 images for every original image and resulted in a dataset of 1024.

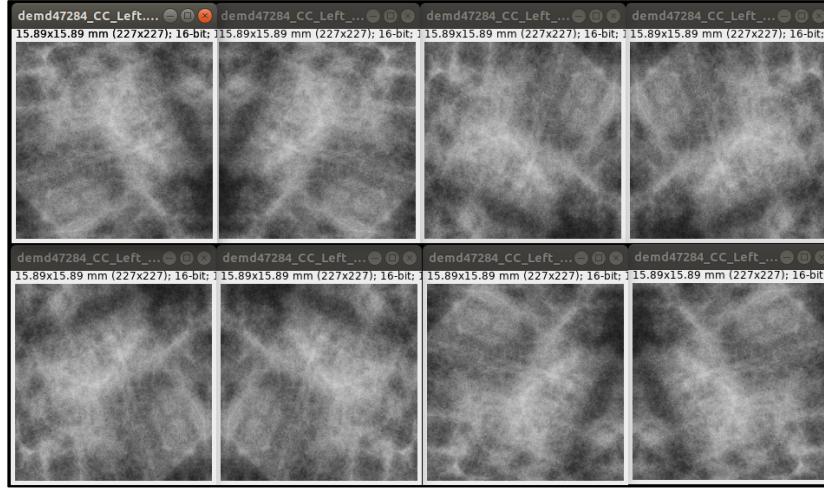


Figure 26 – Example of one image being augmented into seven new images. Top left is the original image.

4.12 Converting DICOM to TIF and Colour

The majority of machine learning models do not accept DICOM format images, therefore the images needed converting to an acceptable image format. After researching the different options PNG and TIF were decided upon as the best solutions due to their lossless conversion properties (figure 27). In addition to this new requirement, AlexNet only accepts ‘colour’ images, that is image with a depth (z value) of 3. However, the mammograms that are being used only have two dimensions x and y (227x227).

dcmimage							tifImage																		
227x227 uint16							227x227x3 uint16																		
1	2	3	4	5	6	7	1440	1507	1372	1492	1524	1426	1426	1333	1420	1187	1498	1200	1350	1370	1507	1533	1363	1470	1240
1	1440	1507	1372	1492	1524	1426	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1561	1693	1414	1445	1718	1397	1321	1606	1173
2	1498	1200	1350	1370	1507	1533	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
3	1458	1502	1596	1501	1443	1570	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
4	1561	1693	1414	1445	1718	1397	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
5	1862	1750	1587	1627	1716	1627	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
6	1870	1731	1619	1528	1596	1631	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
7	1724	1894	1646	1505	1579	1402	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
8	1750	1612	1706	1849	1558	1454	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
9	1662	1537	1716	1496	1445	1356	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
10	1917	1869	1618	1566	1594	1516	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
11	1770	1799	1747	1549	1806	1636	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
12	1468	1686	1916	1395	1591	1780	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
13	1618	1501	1514	1655	1606	1640	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
14	1588	1347	1463	1622	1533	1531	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
15	1628	1716	1336	1658	1802	1445	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
16	1549	1622	1457	1678	1537	1487	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
17	1880	1668	1646	1612	1751	1528	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
18	1880	1738	1636	1628	1655	1618	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
19	1702	1738	1697	1766	1769	1795	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519
20	1837	1802	1800	1607	1885	1600	1	1458	1502	1596	1501	1443	1570	1436	1531	1321	1562	1750	1587	1627	1716	1627	1190	1449	1519

Figure 27 – Example of a lossless image conversion. Left is the DICOM file format, right is the TIFF format for the same image. The blue square is highlighting a section to show the values are the same.

Therefore, each image was read using MatLab's 'dicomread' function, the '.dcm' was removed from the image name and '.tif' was appended to it. The full image was then remapped to a new image which contained a second and third layers which were duplicates of the first layer. Resulting in images of size 227x227x3 where the 3 represents the 3 duplicated layers (figure 28). Each new image is then written to the same folder using MatLab's 'imwrite' function and is thus saved as 3-dimensional TIFF image ready for machine learning.

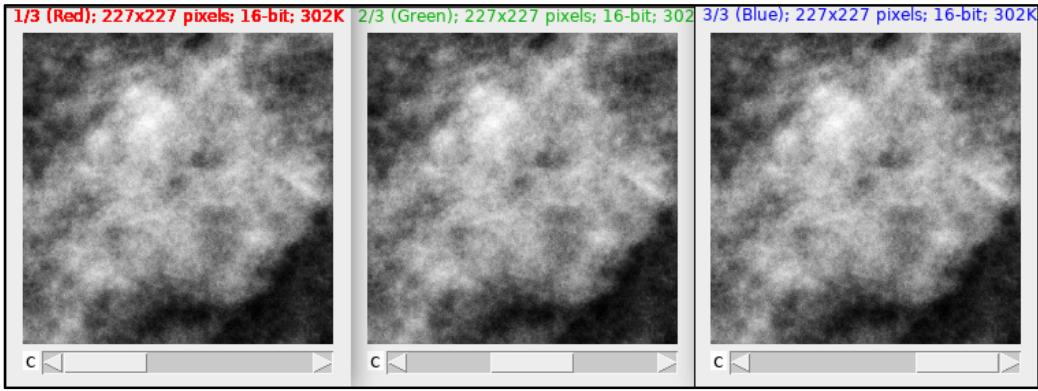


Figure 28 – Converting a grayscale image to Red, Green, Blue (RGB) ‘colour’ image by repeating the gray image 3 times

5 Deep Learning

As discussed in section: 2.11, when one's dataset does not contain enough images to train a CNN from scratch it is best practice to use the start of a highly trained model and retrain the last few layers to a specific categorization problem. The CNN chosen for this research is called AlexNet (Krizhevsky et al, 2012) which is a network that has been trained on more than a million images from the ImageNet database (Deng, J. et al, 2009), has 60 million parameters, 650,000 neurons and took six days to train on two GTX 580 3GB GPUs. All but the last three layers of the pretrained network were used and three ‘empty’ layers were appended ready for training, in this case a Fully Connected, SoftMax and Classification layer (table 2).

AlexNet CNN	
Layer	Description
Data	227x227x3 images
Convolution 1	96 11x11x3 convolutions with stride [4 4] and padding [0 0 0]
ReLU 1	ReLU
Normalization 1	Cross channel normalization with 5 channels per element
Pool 1	3x3 max pooling with stride [2 2] and padding [0 0 0]
Convolution 2	2 groups of 128 5x5x48 convolutions with stride [1 1] and padding [2 2 2]
ReLU 2	ReLU
Normalization 2	Cross channel normalization with 5 channels per element
Pool 2	3x3 max pooling with stride [2 2]
Convolution 3	384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
ReLU 3	ReLU
Convolution 4	2 groups of 192 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
ReLU 4	ReLU
Convolution 5	2 groups of 128 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
Pool 5	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
Fully Connected 6	4096 fully connected layer
ReLU 6	ReLU
Dropout 7	50% dropout
Fully Connected 8	1000 fully connected layer
Probability	Softmax
Output	Classification

Table 2 – AlexNet layers.

When performing machine learning with many images - due to their size - it is not possible to load the whole dataset ready for processing. Hence, an ‘image data store’ class was created which is a structure that contains file paths to all the images and their benign/malignant classification. The images are then randomly split into training and validation sets every iteration (epoch), 70% training (1052 images) 30% validation (452 images).

Initially the network was trained using the following options:

1. Mini batch size: 10
2. Max epochs: 6
3. Initial learn rate: 0.0001

The training and validation was repeated 5 times and the averages of the validation results were:

Sensitivity	0.7035
Specificity	0.6296
PPV	0.6687
Accuracy	0.6663
AUC	0.7443

Table 3 – Initial Deep Learning results.

The Initial Learn Rate was then changed to 0.00001, the training and validation process repeated, and the following average validation results were recorded:

Sensitivity	0.7752
Specificity	0.6563
PPV	0.6973
Accuracy	0.7123
AUC	0.8068

Table 4 – Deep Learning results after decrease in Initial Learn Rate.

Due to the noticeable increase in all measurement factors the Initial Learn Rate was kept the same. The Maximum Epoch was then changed from 6 to 10, the training and validation process repeated, and the following average validation results were recorded:

Sensitivity	0.7920
Specificity	0.6537
PPV	0.6963
Accuracy	0.7229
AUC	0.8213

Table 5 – Deep Learning results after increase in Maximum Epoch size

6 Results

6.1 Image Registration Results

As outlined in section 4.9 there were a number of qualitative reasons the fully automated image registration failed to work. However, there was one quantitative measurement that contributed to the decision to semi-automate the process. The distance from coordinate center to maximum correlation center was measured for every image and distances below 400 pixels were deemed suitable due to the original patch edges being 500 pixels from the center, thus encapsulating any

tumours that were not absolutely central. The results, which can be seen in figure 29, show that although 300 of the cropped spot compression images successfully registered, there were over 250 that were deemed unsuccessful due to being over the defined threshold of 400 pixels.

The rose graph, which can be seen on the right-hand side of figure 29, shows in which direction - from the ground truth coordinates - the cropped image was registered. The angle was measured to see if there was a large bias in images being registered in one direction from the ground truth. However, it would appear that the direction is reasonably well distributed.

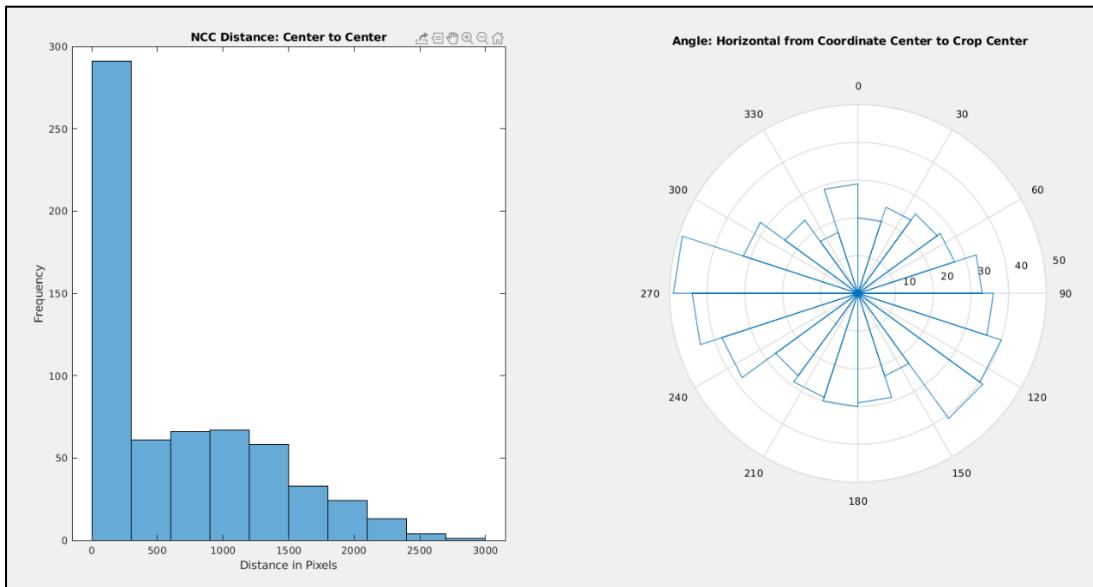
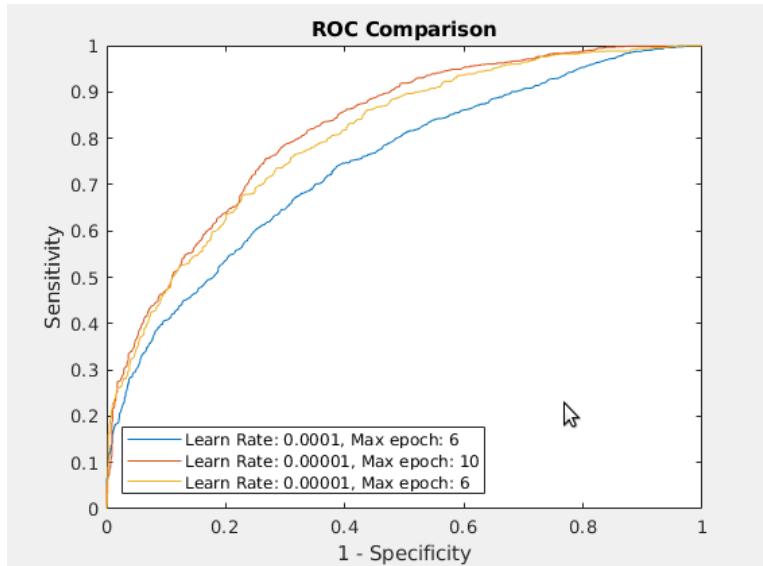


Figure 29 – Left, histogram depicting the result of each image pairs center to center distance. Right: Rose graph showing in which direction the crop image has registered in relation to the coordinate center.

6.2 Deep Convolutional Neural Network Results



Graph 1 – ROC comparison: Decreasing the Learn Rate and increasing the Max Epoch yields a better AUC.

As can be seen in graph 1 above, a decrease in Learn Rate from $0.0001 (10^{-4})$ to $0.00001 (10^{-5})$ and an increase of Max Epochs from 6 to 10 yielded a better AUC 82% compared to 74%. Tests were also performed at $0.000001 (10^{-6})$ learn rate but there was no increase in performance.

7 Discussion

Three hypotheses were tested during this project:

Is it possible to automatically extract patches from Spot Compression images?

Is it possible to perform image registration with Spot and Full mammograms and automatically extract patches from the full image?

Is it possible to classify tumours as benign or malignant using Deep Convolutional Neural Networks?

Extracting patches from spot compression images would be useful in an image registration pipeline. However, it would require complex computer vision based algorithms due to the variety of forms spot compression images can take (figure 21). Once the semiautomated script was written it took less than an hour to crop over 300 images. This had the added benefit of the crops being centered on regions of interest as opposed to the whole area leading to better image registration results later.

With regards to the second hypothesis, it is indeed possible to perform image registration with Spot and Full mammograms and automatically extract patches from the full image. However, a

correct registration does not equal a useful patch. This is due to the contents of spot compression images not necessarily containing a region of interest, for example a spot compression with no tumour present. Therefore, human input was also required at this stage to maximise the amount of useful tumour patches for the DCNN stage. Once the script was written the process of selecting tumours was fast, efficient and resulted in many more useful patches.

The preliminary results of the Deep Convolutional Neural Network are positive (Table 5) However, there is definitely scope for improving these results, possibly by trying other networks that have been trained on ImageNet and adjusting the parameters of the network.

The preprocessing stage of the project took approximately 90% of the project time. This was unforeseen and a big learning curve.

8 Conclusion and Proposed Future Work

The central research questions have been answered and show that a semi-automated approach to tumour patch extraction is preferred, this is due to the many variations that spot compression images can take (figure 22) and the variety of locations a tumour can reside within a spot compression image.

Although Image registration using DICOM images has been shown to work there is no telling whether or not the image registration location is of any use for machine learning without observing every case, rendering the full automation of patch extraction counterproductive for machine learning purposes.

Using machine learning to classify tumours has been shown to work to an extent with results of: AUC 82%, Sensitivity 0.79%, Specificity 65% and PPV 70% recorded. However, there is room for improvement with more fine-tuning of the model required.

Future work would include the aforementioned thorough deep learning tests. Another avenue of research could be to split the whole full mammogram images into patches and use them all to train a deep learning model. This would have the dual benefit of providing large amounts of data for both benign and malignant classes and the patches could include data that is pertinent to classification but unseen by humans.

References

- (n.d.).
- Dr Henry Knipe. (2018). *Axillary tail breast cancer*. Retrieved from radiopaedia:
<https://radiopaedia.org/cases/axillary-tail-breast-cancer>
- 3Blue1Brown. (2017, 10 5). *3Blue1Brown*. Retrieved from YouTube:
<https://www.youtube.com/watch?v=aircArUvnKk&t=1021s>
- breastimaging.vcu.edu. (2017, 05 08). *Diagnostic Mammography*. Retrieved from breastimaging.vcu.edu:
<https://breastimaging.vcu.edu/services/imaging/diagnosticmammo.html>
- Cancer Research UK. (2018, 10 19). *Breast Cancer Statistics*. Retrieved from Cancer Research UK:
<https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer>
- Comprehensive Diagnostic Imaging. (n.d.). *3D-HD MAMMOGRAPHY WITH C-VIEW*. Retrieved from Comprehensive Diagnostic Imaging: http://cdiok.com/?page_id=12
- Dronkers, D. J., Hendriks, J. H., Holland, R., & Rosenbusch, G. (2002). *The Practice of Mammography*. Stuttgart: Thieme.
- Elmore, J. e. (2009). *Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy*.
- Arevalo, John. González, Fabio A. Ramos-Pollánb, Oliveira, Raúl. Jose L. Angel Guevara Lopez, Miguel. (2016). *Representation learning for mammography mass lesion classification with convolutional neural networks*. Comput. Methods Programs Biomed., 127 (2016), pp. 248-257
- Forrest, P. (1986). *Breast cancer screening: report to the Health Ministers of England, Wales, Scotland & Northern Ireland*. London: H. M. S. O.
- Gøtzsche, P. C., & Jørgensen, K. J. (2013). *Screening for breast cancer with mammography*. John Wiley & Sons, Ltd.
- Graziano, L. (2012, March). *Breast Lymphoma: A pictorial review*. Retrieved from Research Gate:
https://www.researchgate.net/figure/Mammography-Left-breast-CC-and-MLO-mammograms-reveals-a-nodular-image-in-the-lower-inner_fig13_272576996
- Health, R. (2012, November 5th). *Women with dense breasts may need annual mammograms*. Retrieved from Reuters: <https://www.reuters.com/article/us-health-mammography-dense-breasts-idUSKCN10X25W>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems*, (pp. 1097-1105). Toronto.
- Lim, M. (2018). *AP Computer Science Principles*. Retrieved from medium:
<https://medium.com/mslim/computer-vision-288f53a28b6b>
- Liu, D. (2017, November 30). *A Practical Guide to ReLU*. Retrieved from medium:
<https://medium.com/tinymind/a-practical-guide-to-relu-b83ca804f1f7>
- Machine Learning Guru. (n.d.). *Undrestanding Convolutional Layers in Convolutional Neural Networks (CNNs)*. Retrieved from Machine Learning Guru:
http://machinelearningguru.com/computer_vision/basics/convolution/convolution_layer.html
- Machine Learning Guru. (n.d.). *Undrestanding Convolutional Layers in Convolutional Neural Networks (CNNs)*. Retrieved from Machine Learning Guru:
http://machinelearningguru.com/computer_vision/basics/convolution/convolution_layer.html
- Mammographic Equipment Schematics. (n.d.). Retrieved from Crossfit Hartford:
https://crossfithartford.com/dummies_mammographic_equipment_schematics.php
- mammoguide. (n.d.). *Mammography Techniques*. Retrieved from mammoguide:
<https://www.mammoguide.com/mammo-techniques.html>
- National Breast Cancer Foundation Inc. (2016, 06 12). *Breast Cancer Biopsy*. Retrieved from National Breast Cancer Foundation Inc: <https://www.nationalbreastcancer.org/breast-cancer-biopsy>
- Raj, B. (2018, April 11). *Data Augmentation | How to use Deep Learning when you have Limited Data—Part 2*. Retrieved from Medium: <https://medium.com/nanonetshow-to-use-deep-learning-when-you-have-limited-data-part-2-data-augmentation-c26971dc8ced>

- Raschka, S. (2015, March 24). *Single-Layer Neural Networks and Gradient Descent*. Retrieved from Sebastian Raschka: https://sebastianraschka.com/Articles/2015_singlelayer_neurons
- Rawat, W., & Zenghui, W. (2017). *Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review*. Florida, South Africa: Massachusetts Institute of Technology.
- Screening, T. I. (2012). *The Benefits and Harms of Breast Cancer Screening*. London: Cancer Research UK.
- Sharma, S. (2017, September 9). *The Fundamentals of Neural Networks*. Retrieved from Towards Data Science: <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>
- Siemens. (2018, 09 17). *How CAD can help save time in mammography reading*. Retrieved from healthcare.siemens: <https://www.healthcare.siemens.co.uk/mammography/news/how-cad-can-help-save-time-in-mammography-reading.html>
- Tse, T., & Esposito, M. (2017, May 8). *Deep Learning and Neural Networks*. Retrieved from The Conversation: <https://theconversation.com/deep-learning-and-neural-networks-77259>
- UW Health. (n.d.). *Breast Center*. Retrieved from UW Health: <https://www.uwhealth.org/breast-center/digital-breast-tomosynthesis/50738>
- Zeiler, M. D., & Fergus, R. (2013). *Visualizing and Understanding Convolutional Networks*. European Conference on Computer Vision.
- Chan, Carmen. (2017). *What is a ROC Curve and How to Interpret It*. Retrieved from: <https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/>
- Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li, K. and Fei-Fei, L. (2009). *ImageNet: Large-Scale Hierarchical Image Database*. CVPR09
- Shie, Chuen-Kai. Chuang, Chung-Hisang. Chou, Chun-Nan. Wu, Meng-Hsi & Y.Chang, Edward (2015). *Transfer representation learning for medical image analysis*. 10.1109/EMBC.2015.7318461
- Krizhevsky, Alex. Sutskever, Ilya and Hinton, Geoffrey E (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Advances in Neural Information Processing Systems 25. Curran Associates, Inc.