Automatic Classification, Detection and Segmentation of Breast Arterial Calcification on Digital Mammography Images Using Deep Learning

Dominic Maguire

Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

School of Health & Society,

University of Salford.

2024

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors, Professor Sunil Vadera and Dr Katy Szczepura, for their invaluable guidance, support and patience over the last few years. Also thanks to Dr John Thompson who preceded them and to Dr Lucy Walton and Professor Farid Meziane who initiated the research. I was very lucky to have such a group of advisors over the course of the study. I am very grateful too for the financial support from the School of Health & Society which allowed me access to the OPTIMAM dataset.

Thanks to my colleagues in Tipperary University Hospital, Dr Marian O'Reilly and Dr John Hynes, who kindly participated in the observer reader studies and provided expert advice. Thanks also to my line manager, Siobhán Thompson, for facilitating time for supervision meetings and study leave.

Finally, thanks to my family, Carmel, Hannah and John, for all your love, encouragement, support and patience over the last 6-plus years. *Mo cheol sibhse.*

List of Abbreviations and Acronyms

AF Atrial Fibrillation.

AI Artificial Intelligence.

API Application Programming Interface.

ASCVD Atherosclerotic Cardiovascular Disease.

ASPP Atrous Spatial Pyramid Pooling.

AUC Area Under Curve.

BAC Breast Arterial Calcification.

BACCADS Breast Arterial Calcification and Coronary Artery Disease Scale.

BI-RADS Breast Imaging-Reporting and Data System.

BMI Body Mass Index.

CA Coronary Artherosclerosis.

CAC Coronary Artery Calcium.

CAD Computer-aided Diagnosis.

CAD Coronary Artery Disease.

CADe Computer-aided Detection.

CADx Computer-aided Diagnosis.

CBIS-DDSM Curated Breast Imaging Subset of Digital Database for Screening Mammography.

CC Craniocaudal.

CLAHE Contrast-limited Adaptive Histogram Equalisation.

CMR Cardiac Magnetic Resonance.

CNN Convolutional Neural Network.

CRF Conditional Random Field.

CRISP-DM CRoss-Industry Standard Process for Data Mining.

CT Computed Tomography.

CTCA CT Coronary Angiography.

CVD Cardiovascular Disease.

DBT Digital Breast Tomosynthesis.

DICOM Digital Imaging and Communications in Medicine.

FCN Fully Convolutional Network.

FDA US Food and Drug Administration.

FFDM Full Field Digtal Mammography.

FN False Negative.

FP False Positive.

FROC Free-Response Operating Characteristic.

GAN Generative Adversarial Network.

GPU Graphics Processing Unit.

HU Hounsfield Unit.

IoU Intersection Over Union.

JSON JavaScript Object Notation.

LDA Linear Discriminant Analysis.

MLO Mediolateral Oblique.

MP Megapixel.

MRI Magnetic Resonance Imaging.

NBSS National Breast Screening System.

NHS National Health Service.

NMS Non-Maximum Suppression.

ONNX Open Neural Network Exchange.

PACS Picture Archiving and Communication System.

PAN Path Aggregation Network.

PNG Portable Network Graphic.

ReLU Rectifier Linear Unit.

ROC Receiver Operating Characteristic.

RPN Region Proposal Network.

SPP Spatial Pyramid Pooling.

SVM Support Vector Machine.

TN True Negative.

TP True Positive.

YOLO You Only Look Once.

Abstract

Cardiovascular disease (CVD) is the leading cause of premature death in the United Kingdom with one type, coronary artery disease, killing more than twice as many women as breast cancer. Conventional CVD risk factors have been shown to have less accuracy for females who are considered low-risk. Recently, researchers have noted that breast arterial calcification (BAC), which is regularly observed as an incidental finding on mammograms, could be used to risk-stratify women for CVD.

In 2023, almost 2 million women attended breast screening clinics in England. Automatic BAC detection on mammograms could provide vital additional cardiovascular information, without the need for further invasive tests or radiation exposure, and could direct patients to relevant clinical pathways or therapies.

As a first step in automating the BAC grading process, I developed deep learning models for BAC classification, object detection and segmentation using an anonymised dataset which was annotated for the presence and location of BAC under the guidance of two consultant radiologists. Data augmentation was used in both the classification and object detection networks, increasing the training data size.

My modified ResNet22 network showed promise in classifying the presence or absence of BAC at image level, attaining a test accuracy of 80%, indicating that this method could be used as a simple flag for this purpose. I also used this network for feature extraction in Faster R-CNN and YOLO BAC object detection models. Despite improving on a recent similar study, these latter networks performed poorly with very low average precision scores at several thresholds. As an improvement, this study developed a DeepLabv3+-based BAC segmentation network which doubled the IoU obtained by another study using a similar model and achieved a BFScore of over 70% specifically for BAC.

Based on the findings of this research, a two-step pipeline is recommended with our classifier triaging mammographic images for BAC and our segmentation model providing an indication of the extent of its presence. This could provide the basis for further research in

order to realise the potential of concurrent, automatic BAC grading for women undergoing mammographic imaging.

1 Introduction

1.1 Research Motivation

Cardiovascular disease (CVD) is the leading cause of premature death in the United Kingdom with one type, coronary artery disease (CAD), killing more than twice as many women as breast cancer (British Heart Foundation, 2022). While risk algorithms such as QRISK3 (Hippisley-Cox & Coupland, 2017) can predict the likelihood of cardiovascular events over the next ten years, conventional risk factors such as high blood pressure, smoking and cholestrol levels have been shown to be less accurate when applied to those considered low-risk and especially to women (McClintic, McClintic, Bisognano, & Block, 2010). Magni et al. (2023) believe that this has led to both women and primary care physicians underestimating the risk of developing CVD, noting that the latest prediction models, while supplying age- and sex-specific multipliers, do not include female-specific risk factors.

Breast arterial calcification (BAC), shown in Figure 1.1, is regularly observed on mammograms as a non-actionable incidental finding. It results from the diffuse calcification of the media of small to medium-sized mammary arteries (Abi Rafeh et al., 2012) and is found along the circumference of the media giving a “train track” appearance (Polonsky & Greenland, 2017). Its prevalence was found to be 12.7% in breast screening programmes, rising from 10% in 40 year olds to 50% in 80 year olds (Hendriks et al., 2015). Recent research has led to BAC being considered as a potential female-specific risk factor for cardiovascular disease. Nudy, Asmaro, Jiang, and Schnatz (2022), in a ten-year follow-up prospective study of 1039 patients, found that BAC-positive participants were 3.14 times more likely to develop coronary artery disease and 5.10 times more likely to have a stroke. Similarly, Iribarren et al.

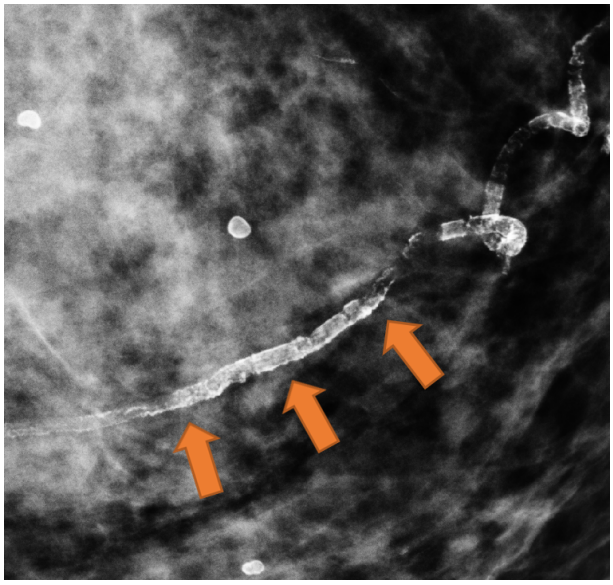


Figure 1.1: Breast arterial calcification.

(2022), in a cohort study of 5059 post-menopausal women aged 60-79 years, found that BAC is associated with a 1.51 times increased risk of incident atherosclerotic cardiovascular disease (ASCVD) and a 1.23 times increased risk of incident global CVD in multivariate models that adjusted for traditional CVD factors. The latter authors believe that these results show that the assessment and reporting of BAC status has the potential to change clinical practice and impact primary CVD prevention for women. Furthermore, utilising mammographic images to screen for BAC affords the possibility of no additional invasive tests or radiation exposure to the patient while providing useful supplementary diagnostic information (Magni et al., 2023).

The grading of BAC, essential for the triage and management of potential cardiovascular patients, is not a straightforward task, however. Arterial calcifications can exhibit topological complexity (K. Wang, Khan, & Highnam, 2019), are sometimes subtle and can be easily overlooked, especially as patterns can be mimicked by image noise or image processing algorithms (Mordang, 2018). Manual BAC segmentation is known to be tedious, expensive, time-consuming and subjective (J. Z. Cheng, Chen, & Shen, 2012; Ghamdi, Abdel-Mottaleb, & Collado-Mesa, 2020). Despite Mantas and Markopoulos (2016) suggesting it would be a reasonable policy to inform women of the presence of BAC on their mammogram, it is not routinely reported. Some researchers have used deep learning to address these challenges

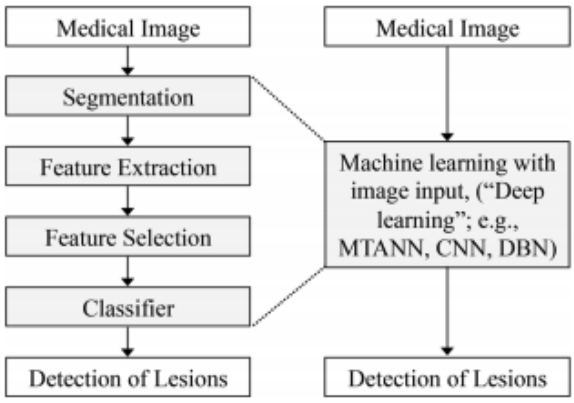


Figure 1.2: Deep learning skips segmentation, feature extraction and selection steps (Suzuki, 2017).

in order to automate the process.

Deep learning is an area of machine learning that uses multi-layer artificial neural networks to automatically learn representations that are expressed in terms of other simpler representations with each layer in the network refining the representation into more abstract levels (Gonzalez & Woods, 2018; Goodfellow, Bengio, & Courville, 2016). It has subsequently replaced the need for manual segmentation, feature extraction and selection traditionally associated with machine learning as shown in Figure 1.2. In a review of deep learning in medical imaging, Soffer et al. (2019) note that deep neural networks have been used for a wide range of tasks including organ classification and disease detection, image segmentation and optimisation. They also state that convolutional neural networks (CNNs), a type of network especially suited to the grid-like structure of image data, are now state-of-the-art in image analysis.

A small number of recent papers using deep learning have shown promising results in BAC classification, detection and segmentation (Ghamdi et al., 2020; Guo et al., 2021; R. Khan & Masala, 2023; Mobini et al., 2023; J. Wang et al., 2017; K. Wang et al., 2019). Limitations of these papers included self-described low image quality (Ghamdi et al., 2020), small datasets (Ghamdi et al., 2020; R. Khan & Masala, 2023; Mobini et al., 2023; K. Wang et al., 2019) with relatively larger datasets limited to a single site and equipment manufacturer (Guo et al., 2021). No paper employed full-image BAC classification and only one paper used object detection techniques (K. Wang et al., 2019). Two papers (Guo et al., 2021;

K. Wang et al., 2019) used their own custom metrics to evaluate BAC segmentation making comparisons difficult.

1.2 Research Methodology

Burrell and Morgan (1979) identified two approaches to social science research methodology, namely idiographic and nomothetic. The idiographic approach places an emphasis on getting close to one's subject and allowing them to relate its nature and characteristics during the investigative process mainly from case studies and action research. The nomothetic approach, on the other hand, emulates methods employed in the natural sciences with formal mathematical analysis, hypothesis testing and both experimental and non-experimental methods such as laboratory experiments and surveys respectively. Iivari, Hirschheim, and Klein (1998) add a third approach for information systems development, constructive methods, which are concerned with the design and creation of artefacts that may be conceptual such as models and frameworks or technical such as software. Mingers (2001) notes that different research methods focus on different aspects of reality and therefore a richer understanding can be obtained by combining several approaches in a pragmatic and pluralist methodology.

Oates, Griffiths, and McLean (2022) note that most recent computer systems research is based on a positivist scientific method similar to the nomothetic approach mentioned above. With this paradigm, the world is ordered and regular, not random, and it can be investigated objectively, mainly through experiments. This allows for the testing of hypotheses and the use of quantitative data and mathematical techniques in a process of observation and measurement. Combining this approach with constructive methods allowed us to iteratively develop models for BAC classification, detection and segmentation.

The above theoretical methodological approaches were implemented using CRISP-DM (CRoss-Industry Standard Process for Data Mining)(Chapman et al., 2000), which is the de facto standard for applying a process model in data mining projects (Schröer, Kruse, & Gómez, 2021). This technology-agnostic, iterative process, shown in Figure 1.3, consists

of six phases: business understanding, data understanding, data preparation, modelling, evaluation and deployment. CRISP-DM's informal methodology also provides flexibility by not enforcing a rigid framework, evaluation metrics or correctness criteria (Niakšu, 2015). The sequence is not fixed and can move back and forth between stages. (Namora & Jan Everhard Riwurohi, 2022).

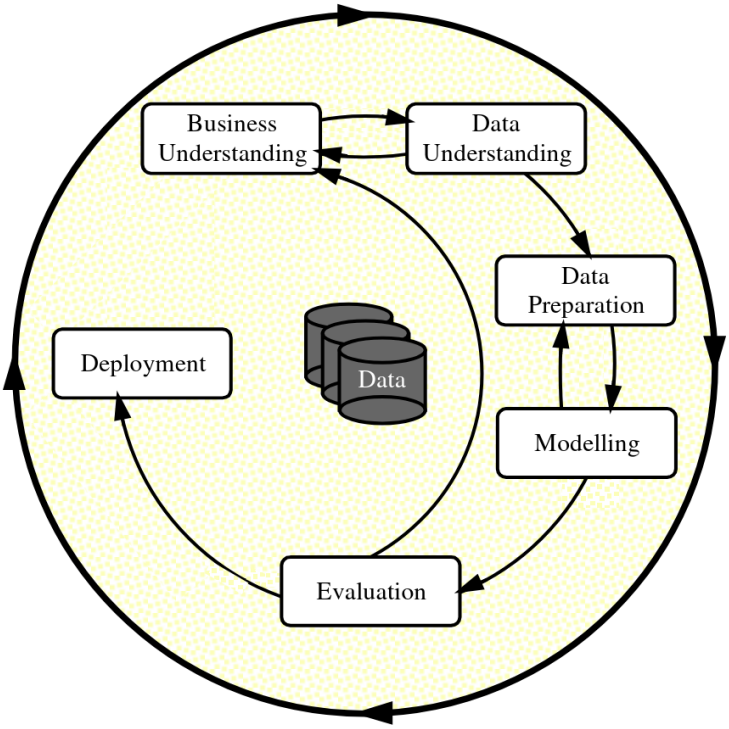


Figure 1.3: Phases of the CRISP-DM Process Model for Data Mining. From Wirth and Hipp (2000).

The first stage, **business understanding**, converts the core project objectives into a deep learning problem and a preliminary project plan designed to achieve the objectives. This was achieved by performing a systemic literature review which investigated BAC, computer-aided detection and deep learning in order to identify gaps and potential areas for improvement in automatic BAC detection. The development of three separate models for classification, detection and segmentation aimed to assist radiologists in determining the presence or absence of BAC, where it is located and how extensive it is.

The second stage, **data understanding**, concerns collecting and understanding the data, identifying any problems and/or initial insights. Wirth and Hipp (2000) state that there is a close link between the data understanding and business understanding phases, i.e. de-

vising the deep learning problem and project plan requires at least some understanding of the data. Niakšu (2015) proposed an extension to CRISP-DM, CRISP-MED-DM, to reflect the non-uniformity of healthcare data in general. This was not needed with our dataset as it was composed of breast screening mammography cases, each in DICOM format.

Stage 3, **data preparation**, encompasses all activities to construct the final dataset. This included data annotation and validation and also image pre-processing such as DICOM to PNG conversion, image cropping and padding, data augmentation and the application of CLAHE (Contrast Limited Adaptive Histogram Equalisation).

In the fourth stage, **modelling**, various modelling techniques are chosen and applied and their parameters optimised. In this project, this entailed selecting suitable architectures for the deep learning task at hand and refining various hyperparamters such as learning rate, dropout rate, batch size, loss function and normalisation techniques.

Stage 5, **evaluation**, thoroughly evaluates the model and reviews the steps executed to construct it, making sure it achieves the business objectives from phase 1. In our case, models were evaluated using pre-stated metrics and compared to the results of similar projects in the literature. It was then ascertained whether the research question, hypotheses and objectives had been achieved.

Schröer et al. (2021) note that the last phase, **deployment**, is absent from most research projects using CRISP-DM. If present, the model should be organised and presented in a way that customers can use it. It was intended in this study that the trained deep learning models would be made available on Github.com for other researchers to access.

The relationship of CRISP-DM phases to the thesis structure is shown in Figure 1.5 in Section 1.5.

1.3 Research Aim & Objectives

Although adoption and deployment issues are an important consideration for the use of AI in a clinical setting, the first step is to develop an accurate model. The aim of this study is to answer the research question, namely:

- What are the most effective deep learning architectures for the automatic detection, classification and segmentation breast arterial calcification in digital mammography images?

In order to achieve this aim, the following objectives will need to be met:

- To conduct a review of the literature regarding BAC and its detection, classification and segmentation using automatic, computerised methods.
- To obtain, pre-process, augment and annotate a dataset of mammography images in preparation for BAC model training.
- To validate BAC ground truth annotations using an observer reader study involving two consultant radiologists.
- To develop and train deep learning models using MATLAB suitable for BAC detection, classification and segmentation.
- To compare and evaluate the model results with other studies in the literature.

Hypotheses to be tested include:

- Our deep learning algorithms can be used successfully for image-level classification of BAC.
- Our deep learning algorithms can be used successfully for region-level BAC object detection.
- Our deep learning algorithms can be used successfully to more accurately segment BAC at pixel-level than the current state-of-the-art.

1.4 Thesis Outline

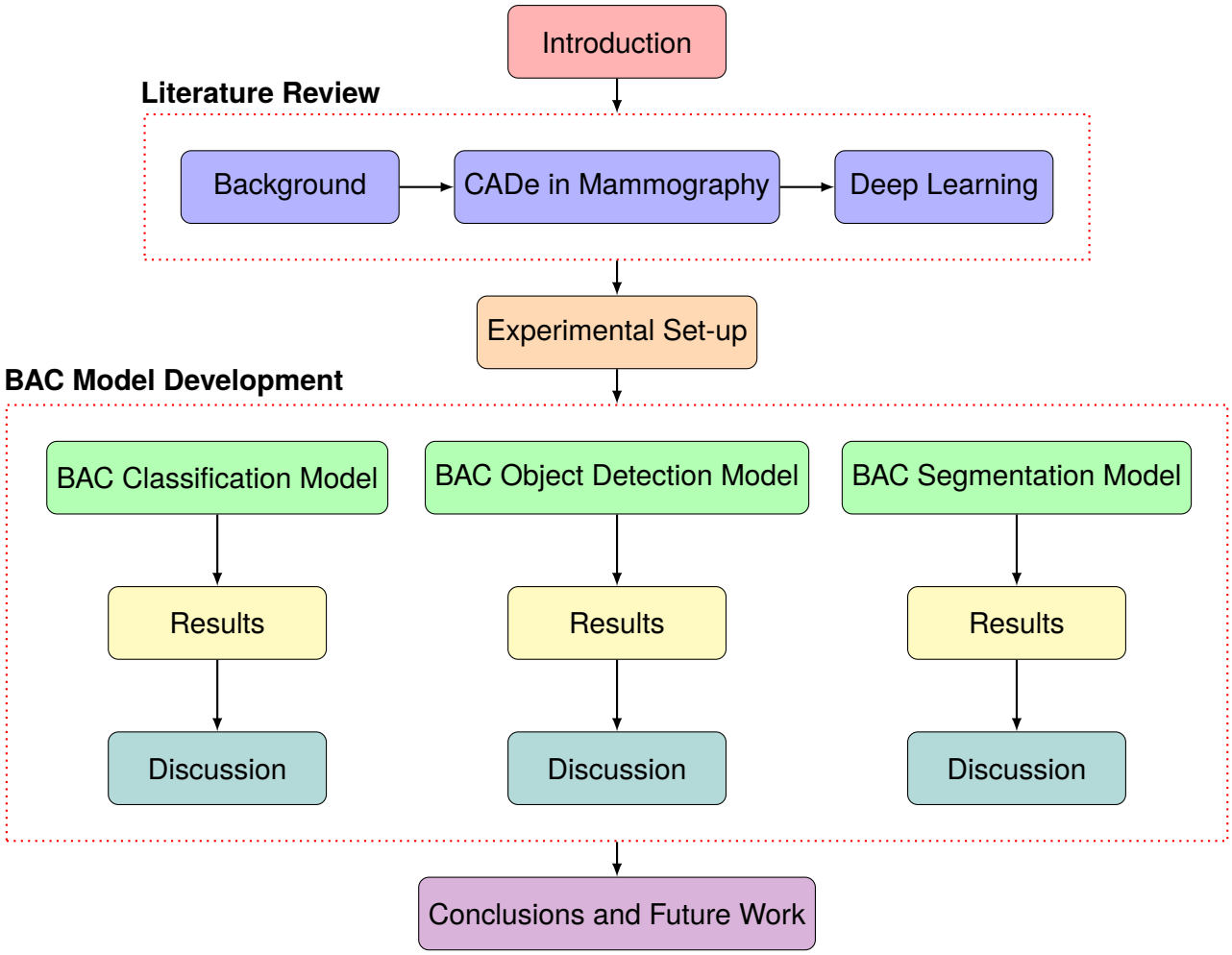


Figure 1.4: Thesis structure.

The structure of the thesis is shown above in Figure 1.4 and each chapter is summarized as follows:

Chapter 1. Introduction This chapter outlines the motivations for the research as well as the theoretical basis for the methodology used in addition to identifying research hypotheses and objectives.

Chapter 2. Background Chapters 2 to 4 make up the main literature review. Chapter 2 provides the background to breast arterial calcification and its links to cardiovascular risk factors and coronary atherosclerosis. It finishes with an overview of current practices in BAC

reporting and grading.

Chapter 3. Computer-aided Detection in Mammography This chapter introduces CADe and its use in mammography along an overview of full field digital mammography. It concludes with an examination of BAC detection using CADe which traditionally employed machine learning techniques.

Chapter 4. Deep Learning This chapter documents the rise of deep learning and the success of CNNs in particular. CNN components are outlined along with networks suited to classification, object detection and segmentation tasks. Applications of deep learning in medical imaging and breast imaging are investigated, finishing with an overview of the core literature relating to its use in BAC classification, object detection and segmentation.

Chapter 5. Experimental Set-up This chapter outlines the various experimental aspects of the study beginning with the dataset and followed by the annotation, image-wise and pixel-wise, and pre-processing, including data augmentation, of the images. An annotation validaton reader study involving two radiologists is also described. The chapter finishes with an outline of the development environment hardware and software.

Chapter 6. BAC Classification Model Chapter 6 describes the ResNet22 BAC classification model including transfer learning and the network training undertaken with reference to optimisation and metrics. Results are presented with a subsequent discussion and summary.

Chapter 7. BAC Object Detection Model This chapter begins with an overview of multi-stage and single-stage detectors and progresses to describing the training one of each type, namely Faster R-CNN and YOLOv4 respectively. Results are presented with a subsequent discussion and summary.

Chapter 8. BAC Segmentation Model Chapter 8 outlines a preliminary evaluation stage where several models were investigated to acsertain their suitability for BAC segmentation. A DeepLabv3+-ResNet18 model was chosen and its training is further described. Results are

presented in addition to evaluating a promising post-processing technique. This is followed by a discussion and summary.

Chapter 9. Conclusions and Future Work In the final chapter, conclusions are made including the main contributions of the research. Limitations of the study and possible future research directions are also outlined.

1.5 CRISP-DM Phase Relationship to Thesis Structure

The relationship between CRISP-DM phases and the thesis structure is shown in Figure 1.5. Business understanding correlates to the introduction and literature review chapters while the two data phases are reflected in the experimental set-up chapter. Modelling and evaluation phases correspond to the BAC model development chapters and the final conclusions and future research recommendations. There was no deployment phase required in this project although trained models were made available on Github.com.

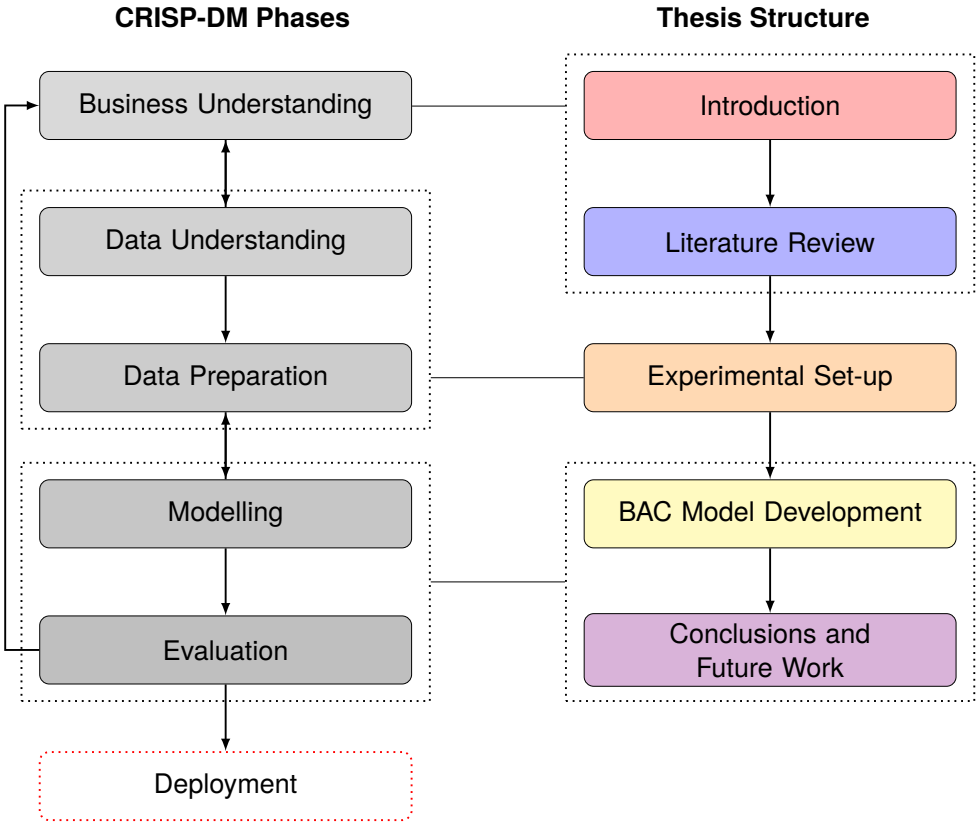
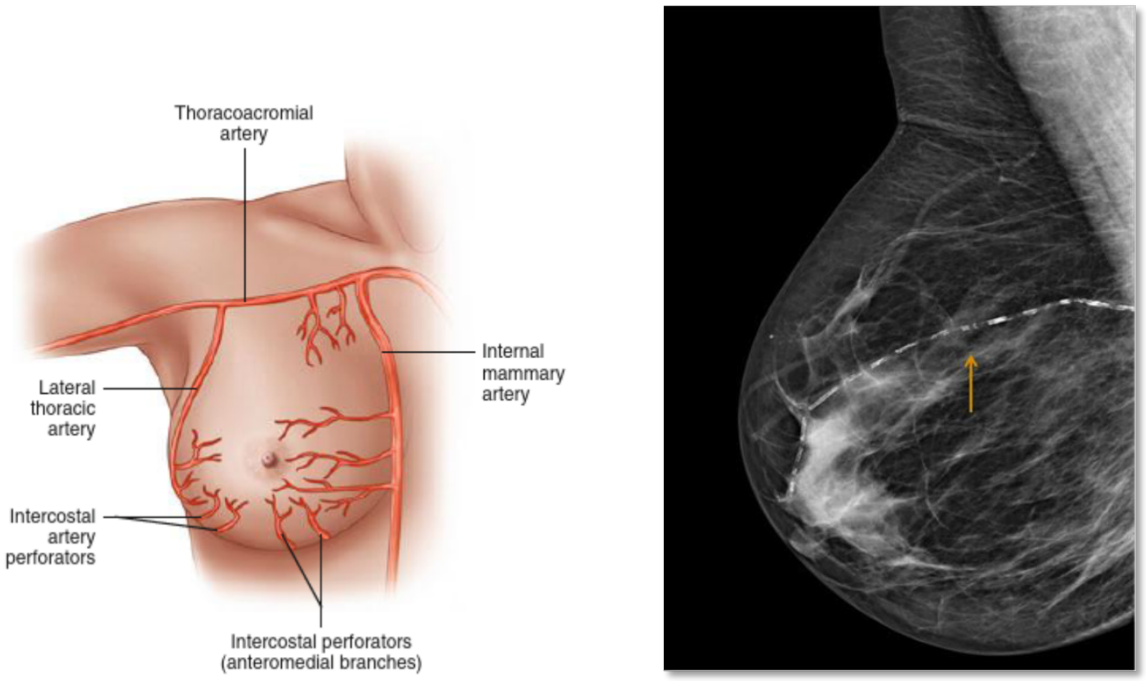


Figure 1.5: Relationship of CRISP-DM phases to the thesis structure.

2 | Background

2.1 Breast Arterial Calcification



(a) Breast vascular anatomy. From Hogg et al. (2015).

(b) BAC on mammogram

Figure 2.1: Breast vascular anatomy and BAC.

Figure 2.1a above shows the vascular anatomy of the breast. Around sixty percent of the total breast arterial supply arises from the perforating branches of the internal mammary artery. The rest is derived from the thoracoacromial artery, the lateral thoracic artery and the intercostal arteries (Hogg et al., 2015). Venous drainage typically mimics the arterial supply (McGuire, 2016).

Breast arterial calcification (BAC), shown in Figure 2.1b above, results from the diffuse

calcification of the media of small to medium-sized mammary arteries and is a common incidental finding on mammograms (Abi Rafeh et al., 2012). BAC is found along the circumference of the media giving a "train track" appearance (Polonsky & Greenland, 2017) and makes the vessel stiffer (Shah et al., 2014) although it remains non-occlusive, i.e. it does not block blood flow (Trimboli et al., 2021). It is associated with increasing age and its prevalence was found to be 12.7% in breast screening programmes rising from 10% in 40 year olds to 50% in 80 year olds (Hendriks et al., 2015). Reddy, Son, Smith, Paultre, and Mosca (2005) note that BAC prevalence varies based on race/ethnicity with Hispanic women having the highest prevalence of BAC at 34%, whereas Asian women have the lowest prevalence of only 7%. Prevalence was found to be 25% in the white population and 24% in African-American women. Lower prevalence was also found among those who are Ashkenazi[1], nulliparous[2] or pre-menopausal or those who have dense breasts or breast implants or who are currently using hormone replacement therapy (Montgomery et al., 2022).

2.2 BAC and Cardiovascular Risk Factors

Researchers have investigated the link between BAC and cardiovascular risk factors and whether BAC itself could be considered as a potential female-specific risk factor for cardiovascular disease.

The idea of cardiovascular risk factors was first mooted during the Framingham Heart Study which amalgamated several factors into a single risk score to estimate the absolute CVD risk over a ten year period (Mahmood, Levy, Vasan, & Wang, 2014). Risk factors have been categorised as medical or lifestyle (British Heart Foundation, 2018), controlled or conventional (Hajar, 2017) and, perhaps most accurately, as modifiable or non-modifiable (World Heart Foundation, 2017). In a recent study examining the impact of cardiovascular risk factors on cardiac structure and function, Petersen et al. (2017) found the most important risk factors to be: age, sex, ethnicity (non-modifiable) and systolic blood pressure, diastolic blood pressure, smoking, levels of physical activity/exercise, body mass index (BMI), high

[1] One of two major ancestral groups of Jewish people whose ancestors lived in France and Central and Eastern Europe, including Germany, Poland, and Russia.

[2] Women who haven't given birth.

cholesterol, diabetes and alcohol intake (modifiable). A recent state of the art review (Bays et al., 2022) adds unhealthy nutrition, kidney dysfunction and genetics/familial hypercholesterolemia.

Iribarren et al. (2004) identified links between BAC and age, having had three or more children and diabetes, while finding an inverse association with smoking. They noted that BAC was significantly associated with a 1.32-fold increased risk of coronary heart disease, a 1.41-fold increased risk of ischaemic stroke and a 1.52-fold increased risk of heart failure. They concluded that screening mammograms may be a useful additional tool for the early detection of CVD in women.

Hendriks et al. (2015), in a systematic review and meta-analysis covering data from 52 articles and over 90,000 women, also found that BAC appears to be associated with an increased risk of cardiovascular disease events. Increased prevalence of BAC was found to be associated with several risk factors including age, diabetes, having children and breast-feeding. Smoking and hormone replacement therapy were found to be associated with a decreased prevalence of BAC. This could be due to the effects of smoking on weight and estrogen metabolism but the authors state that there is no satisfactory explanantion from the literature. No associations were found for hyperlipidemia or hypertension. These findings suggest that the aetiology of BAC may be different to the intimal atherosclerotic process and they conclude that BAC may provide a novel route to a better understanding and treatment of cardiovascular disease.

A more recent 23 year retrospective cohort study (Galiano et al., 2022) also examinied the relationship between BAC and cardiovascular events, finding that BAC is an additional risk factor for those women 59 years of age and under, especially diabetics. They recorded a much higher number of deaths among those with BAC (42.1%) than those without (3.1%). These results were echoed by another 10 year cohort study (Nudy et al., 2022) which found that patients with BAC were 5.10 times more likely to have a stroke and 3.14 times more likely to develop coronary artery disease.

2.3 BAC and Coronary Atherosclerosis

Coronary atherosclerosis (CA) is a disease with multiple clinical manifestations ranging from asymptomatic to stable angina, acute coronary syndrome, heart failure and sudden cardiac death (Boudoulas, Triposciadis, Geleris, & Boudoulas, 2016). Plaque develops in the wall of a coronary artery, remodelling the artery so that the luminal area of the vessel is enlarged. Several papers recommend coronary arteriography in all patients with suspected CA given the limitations of other diagnostic methods such as stress testing (could be done too late) and multi-slice computed tomography (CT) (may not define the degree of stenosis) (Arbab-Zadeh, 2016; Bober & Jahangir, 2015; Gould, 2009). Henein, Vancheri, Bajraktari, and Vancheri (2020) indicate that the choice of imaging technique depends on the cardiovascular risk of the patient with non-invasive modalities, such as CT, CT coronary angiography (CTCA) and cardiac magnetic resonance (CMR), being more suitable in primary prevention for low-to-intermediate populations in order to improve risk stratification and to identify individuals who may benefit from individual treatment. The coronary artery calcium (CAC) score determined by CT, therefore, plays an important role in cardiovascular risk stratification, showing a significant association with the medium- or long-term occurrence of major cardiovascular events (Neves, Andrade, & Monção, 2017). Calcification is identified on a CT image as areas of hyper-attenuation of at least $1mm^2$ with >130 Hounsfield units (HU), which is a measure of radiodensity, or 3 adjacent pixels (Agatston et al., 1990).

The relationship between BAC and coronary atherosclerosis as diagnosed on coronary angiography has been investigated by several studies. Both Abi Rafeh et al. (2012) and Jiang, Clark, Singh, Juhn, and Schnatz (2015) noted that BAC on mammography may not be a benign finding as its presence increases the risk of having obstructive coronary artery disease (CAD) by 60% and it is significantly associated with stroke. Maas et al. (2007) similarly identified the association of BAC to subsequent CAD development. Forty-four (76%) women with BAC on mammograms at baseline had coronary artery calcifications after a mean period of 9 years, whereas 218 (49%) women without BAC had coronary calcifications at follow-up. A more recent study by Kelly et al. (2018) found that the presence of

BAC correlated to a CAD-RADS (coronary artery disease classification) score of 3, i.e. requiring functional cardiac imaging. Similar results with asymptomatic women led Yoon et al. (2018) to conclude that BAC evaluation provides an independent and incremental value over conventional risk algorithms.

Chadashvili, Litmanovich, Hall, and Slanetz (2016) undertook a study of 145 patients who underwent coronary CT within a year of screening or diagnostic mammography and found that BAC correlated with a CAC score of >11 which indicates mild or greater risk of developing CAD. They concluded that BAC can be used as a potential marker for increased risk of developing CAD. This finding was replicated by Minssen et al. (2022) in a retrospective study of patients who had mammogram and CT thorax exams between 2009 and 2018. Interestingly, the highest diagnostic accuracy of BAC to detect CAC (93.2%) was noticed in women under 60 years.

Margolies et al. (2016) also sought to determine whether BAC could predict CAC. BAC was evaluated using a score (from 0 to 12) derived from the number of vessels involved in each breast, the longest length of vessel involvement and the density of calcium in the most severely affected segment. They found a strong quantitative association between BAC with CAC and considered BAC to be superior to the Framingham Risk Score and Pooled Cohort Equations in area under curve (AUC) comparisons of BAC > 0 and risk scores for the presence of CAC. Nasir and McEvoy (2016) reiterate the latter study's positive predictive value of nearly 70% for identifying women with the presence of CAC and state that BAC detection should be actively pursued in all mammograms performed and its reporting and management tracked as part of core quality measures.

In some studies the association between BAC and CAC has been weak or absent. A cross-sectional study by Moradi, Adibi, and Abedi (2014) found no significant correlation of BAC severity to CAC severity. Matsumura et al. (2013) carried out a case-control study comparing 98 women with BAC with a control cohort of 104 without BAC and found that BAC was not predictive of a CAC score > 0 although in an age-adjusted model BAC presence did correlate with a high risk calcium score. Another more recent study examining BAC and atherosclerotic CVD in an Australian cancer population (S. C. Lee et al., 2023) found that

while patients with BAC were 1.37 times more likely to experience a cardiovascular event, it was not statistically significant.

Erbil et al. (2018) previously recommended age-controlled studies with longer follow-ups in order to address this discrepancy and to maximise the potential of mammography as a "vascular screening" tool. Since then, a recent study (Iribarren et al., 2022) found a 1.51 times increased risk of incident atherosclerotic CVD (ASCVD) in post-menopausal women with the authors stating that BAC should be considered a risk-enhancing factor for ASCVD and that it has potential to change clinical practice in primary CVD prevention. The same team also found an association between BAC and atrial fibrillation (AF) in women over 70 (Iribarren et al., 2023), noting that BAC could inform who to screen for AF in that age cohort.

2.4 BAC Reporting and Grading

Iribarren and Molloi (2013) noted that early BAC evaluation studies were based on the presence or absence of BAC rather than a gradation and they suggested future studies should look at the role of gradation in the prediction of CVD outcomes. Polonsky and Greenland (2017) advocated a grading system that does not compromise workflow and recommended mild, moderate and severe grade values. A BAC score from zero to twelve was introduced by Margolies et al. (2016) based on the number of vessels in each breast, the longest length of vessel involvement and the density of calcium in the most severely affected segment. In addition to the significant association of the total BAC score, each of the three components of the BAC score was quantitatively related to CAC.

Several studies (Heaney et al., 2022; Mostafavi et al., 2015; Ružičić et al., 2018a, 2018b) use or recommend a four-point Likert scale as illustrated in Figure 2.2 with Grade 0 representing no vascular calcification and Grade 3 indicating severe, coarse, or tram track calcifications affecting three or more vessels. The Canadian BAC awareness study (Heaney et al., 2022) also adds a suggested script for grade 3: "*There are severe coarse vascular or tram track calcifications affecting three or more vessels in the breasts (Grade 3). A strong association between breast arterial calcifications and cardiovascular disease has been identified in*

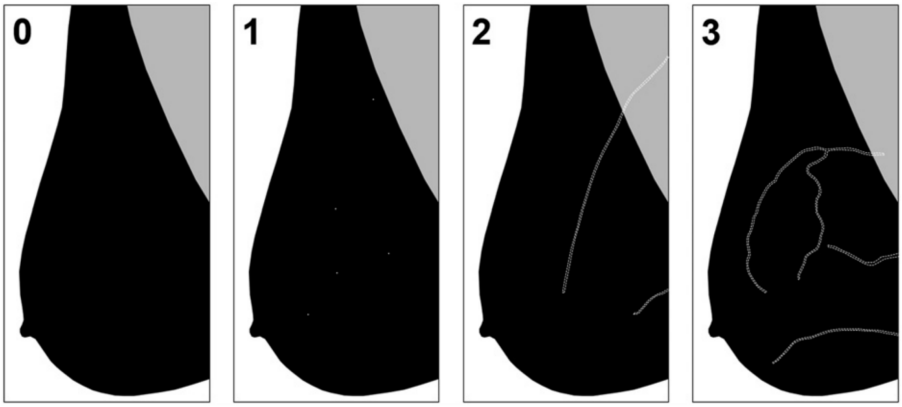


Figure 2.2: Grading of breast arterial calcifications: Grade 0: No vascular calcification. Grade 1: Few punctate vascular calcifications, without coarse, tram-track, or ring calcifications. Grade 2: Coarse vascular calcification or tram-track calcification in fewer than three vessels. Grade 3: Severe, coarse, or tram track calcifications affecting three or more vessels. From (Heaney et al., 2022).

multiple studies. Consider correlation with cardiovascular risk factors as clinically indicated."

Ružičić et al. (2018a) added "four recognized risk factors" - age, blood sugar, total cholesterol and BMI to their BAC Likert scale to create the BACCADS (Breast Arterial Calcification and Coronary Artery Disease Scale) scoring system. Ordinal values of risk factor levels (total CAD score) and the BAC Likert scores were combined as shown below to create the total BACCADS score:

- BAC score (0-3).
- The total CAD score (0-3) = (age category + glycated haemoglobin category + total triglycerides category + body mass index category) / 4.
- The total BACCADS score (0-6) = BAC score + total CAD score.

The BAC score, total CAD score and BACCADS score were evaluated on their ability to predict a patient's known SYNTAX score, which is a measurement of CAD severity (Neumann et al., 2018). As Figure 2.3 shows all three scales were very good at detecting patients with a SYNTAX score > 22 (intermediate-to-high severity CAD) with the BACCADS score better than either risk factors or the BAC scale alone.

In a multi-ethnic quantitative study of BAC gradation and CVD, Iribarren et al. (2017) used a BAC mass value determined by image densitometry applied to raw digital mam-

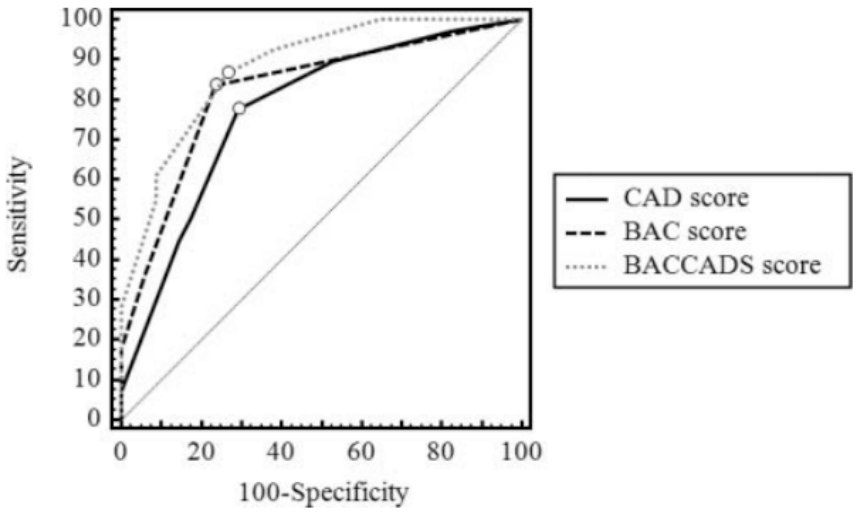


Figure 2.3: The ROC (Receiver Operating Characteristic) curves of BAC scale, CAD score and the total BACCADS score in the dissociation SYNTAX score patient groups (examined status - SYNTAX score >22). From (Ružičić et al., 2018a).

mograms. This technique was developed previously using standard digital mammography hardware and considered the effects of scatter correction, breast thickness, breast density, anatomic background and magnification error (Molloi et al., 2009; Molloi, Xu, Ducote, & Iribarren, 2008). Their study found the overall prevalence of BAC to be 26% with a mean (SD) BAC mass of 12 (23) mg and a range of 0-342 mg. The authors stated that, in future, they aimed to study BAC mass associations with multiple risk factors and markers and investigate the role of BAC in the prediction of CVD and whether adding BAC mass to prediction models based on traditional risk factors improves reclassification of risk for CVD as a whole. This work has still to come to fruition.

Three studies investigated the awareness among radiologists in Canada, Europe and the United States of the association of BAC and higher cardiovascular risk (Brown, Wahab, Zhang, Smetherman, & Mahoney, 2022; Heaney et al., 2022; Trimboli et al., 2021). They found a high level of knowledge of the link and a wide range of BAC reporting levels from 16% in Canada to 61.9% and 87% in Europe and the United States respectively. Despite this, the three studies found that most reported BAC in a binary fashion for presence/absence and it was rarely reported quantitatively. Only 1% of radiologists in the US study provided a BAC score and only one radiologist (of 234 respondents) in the EU study performed “quantitative measurement”. Trimboli et al. (2021) believe this may be due to the lack of validated and

reproducible quantification methods that allow for cardiovascular risk stratification and further state that this may be a barrier to standardised BAC reporting.

The literature has shown that BAC has the potential to be a female-specific risk factor for cardiovascular disease with BAC-positive patients being over 5 times more likely to have a stroke and over 3 times nore likely to develop coronary artery disease. Despite this, there is a wide range of reporting levels and grading methods with the vast majority of radiologists not providing a BAC score or any quantitative measurement. The situation could be improved by applying computer technology to automate the latter processes. The next chapter examines computer-aided detection in mammography and how it has been used for BAC detection using traditional machine learning techniques.

3 | Computer-aided Detection (CADe) in Mammography

3.1 Introduction

Computer-aided detection (CADe) is a type of CAD (Computer-aided Diagnosis) which has been defined as a diagnosis made by a radiologist who uses the output of a computer analysis of images when creating their report (Nishikawa, 2010). CADe software identifies and marks suspicious areas on an image in order to aid radiologists in minimising interpretation errors. CADx (Computer-aided Diagnosis) is another type of CAD that helps radiologists decide whether a patient needs a biopsy or not. CAD refers to the whole field and comprises both CADe and CADx.

Winsberg, Elkin, Macy, Bordaz, and Weymouth (1967) first proposed an automated system for reading mammograms using a facsimile scanner and a CDC 160A computer. The reason for doing so was the “problems inherent in the routine viewing of large numbers of examinations of presumably asymptomatic patients”. These problems still exist as the number of mammograms each year in the United States, for example, has risen to 37 million while radiologist numbers have decreased (Harvey et al., 2019).

Boyer, Balleyguier, Granat, and Pharaboz (2008) noted that breast density and difficulty of interpretation, even for radiologists, led to fatigue, lack of attention and failure of detection. Pisano et al. (2005) found that 45% of women who had cancer were given a normal diagnosis on screening mammography. Significant inter- and intra-observer variation also exists with Beam, Layde, and Sullivan (1996) reporting that the sensitivity of radiologists reading

screening mammograms could vary by as much as 45%.

3.2 The Introduction of CADe in Mammography

FDA (US Food and Drug Administration) approval was granted to two commercial mammography CADe systems, R2 image checker® and iCAD Second Look®, in 1998 and 2002 respectively. These systems converted analog screen film images to digital using a feeder as shown in Figure 3.1 and applied machine learning techniques before presenting areas of interest to the radiologist.

Figure 3.1: iCAD Second Look® input tray.

Earlier studies had indicated the potential of CADe. Chan et al. (1990) found that CADe could improve radiologists' performance in detecting clustered micro-calcifications. An observer study by Getty, Pickett, D'Orsi, and Swets (1988) showed that general radiologists, when aided by CADx in classifying breast lesions as benign or malignant, could perform at a level comparable with unaided expert breast radiologists. After FDA approval, Warren Burhenne et al. (2000) reported that image checker® had the potential to reduce radiologists' miss rate by 77%.

Cole et al. (2014) examined the impact of two CADe systems (R2 image checker® and

iCAD Second Look®) on radiologist performance. Three hundred cases were retrospectively reviewed by 29 radiologists using either system. The authors found that although both systems increased the AUC (area under curve) and the sensitivity of the readers, the average differences were not statistically different. Interestingly, they state that radiologists rarely changed their initial diagnostic decision after using CADe, regardless of which system was used.

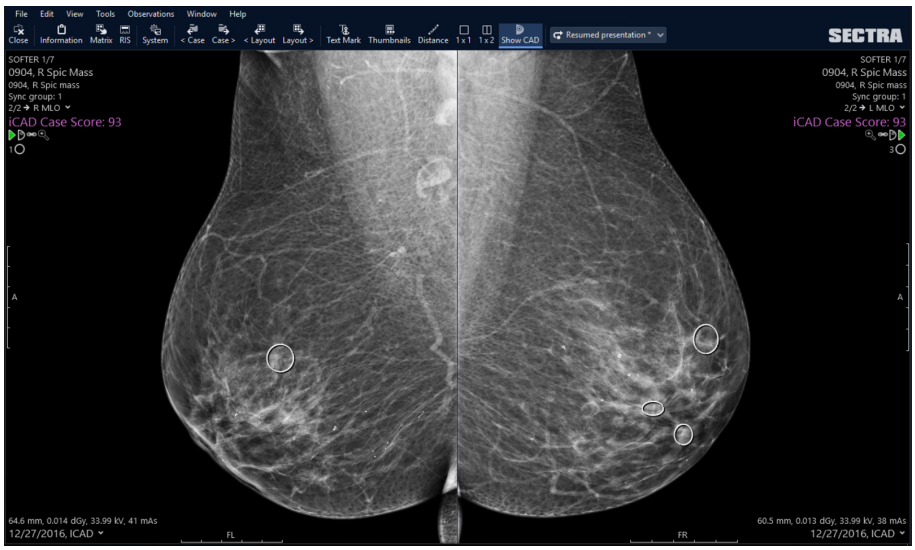


Figure 3.2: iCAD Second Look® System

CADe was initially seen as a tool to augment radiologists by lowering false negatives and reducing the frequency of false positives. It did this by overlaying markings on top of the mammography image, as shown in Figure 3.2, indicating areas where CADe has processed and detected as potentially representing a malignant feature (Harvey et al., 2019). Early algorithms utilised feature extraction via machine recognition of hand-engineered visual motifs. For example, image checker® M1000 detected spiculated lesions by identifying radiating lines emerging from a 6mm centre within a 32mm circle (Taylor, Champness, Given-Wilson, Johnston, & Potts, 2005). Feature extraction enabled rule-based classification using decision trees, support vector machines (SVMs) or multi-layer perceptrons.

3.3 Full Field Digital Mammography

Kotre and dos Reis (2015) acknowledge that mammography is one of the most technically de-

manding examinations in radiology ranging from 20-100μm high density micro-calcifications to low contrast masses - all against a background of mixed densities. Despite the fact that pathology can sometimes be difficult to demonstrate, they note that mammographic technology aims to identify structural or morphological differences in tumours such as tissue masses, angiogenesis, asymmetry and architectural distortion.

The advent of full field digital mammography (FFDM) replaced film screen technology that had been in use for 100 years. FFDM uses a digital detector to capture x-rays passing through the breast in order to produce a latent image that is subsequently processed by a digital computer. The digital image processing techniques enhance the visibility of detail and contrast of the image in order to maximise the detectability of breast lesions (Seeram, 2019). The main components of the system and processing steps are shown in Figure 3.3.

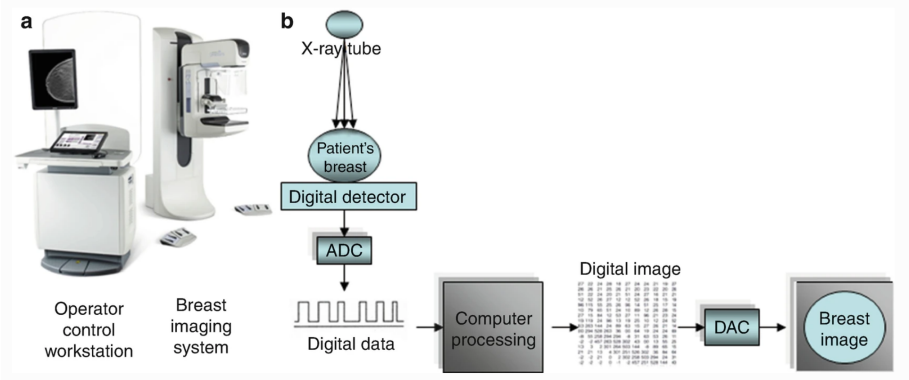


Figure 3.3: (a) FFDM equipment (b) The basic steps in producing the image. From Seeram (2019).

The main advantages of FFDM include a wider dynamic range, greater bit depth, the use of post-processing techniques to enhance image quality and the ability to communicate with a Picture Archiving and Communication System (PACS) (Seeram, 2019). FFDM also allows for a more efficient acquisition of the mammogram through (a) the absorption of most of the x-rays transmitted through the breast, (b) the reduction of quantum noise and (c) a reduction in radiation dose (Yaffe, 2009). The introduction of FFDM meant that mammography imaging data became available in a format that lent itself to computational analysis (Krupinski, 2010). Katzen and Dodelzon (2018) also believe that the permitting of simultaneous billing for FFDM and CAD in the US encouraged the growth of CADe there.

3.4 BAC Detection using CADe

Studies evaluating CADe for breast cancer detection have mostly used commercial systems whereas studies investigating BAC detection have been more experimental and small-scale, employing machine learning techniques. Machine learning is a subset of artificial intelligence (AI) that has been used in a variety of domains to analyse complex datasets and to find patterns and relationships therein without being explicitly programmed (Lakhani et al., 2017). The performance of a machine learning algorithm depends on the representation of the data given to it with each piece of information included in the representation known as a *feature* (Bengio, Courville, & Vincent, 2012). Traditionally, in machine learning, these features were hand-engineered or provided by domain experts.

J. Z. Cheng, Chen, Cole, Pisano, and Shen (2012) leveraged calcification and "vesselness" cues into an integrated framework for detection of vessels with calcium deposits on mammograms. They used a tracking with uncertainty scheme to generate multiple sampling paths and subsequently compiled these paths into super-paths which are then linked using an iterative process. Their study also correlated the association of BAC severity to the degree of cardiovascular risk. Several quantitative measurements for the evaluation of BAC severity were computed including the number of vessels with calcium deposits, vessel length and diameter and calcification density. The authors concluded that their proposed method could potentially be used as a convenient BAC measurement tool in replacement of tedious manual delineation tasks.

J. Ge et al. (2008) employed a k-segments algorithm to find a set of line segments that could suggest the presence of calcified vessels. A four-feature linear discriminant analysis (LDA) reduced segments not linked to BAC. Adjacent segments were linked and dilated with morphological dilation to identify areas of BAC.

In order to remove BACs as false positives in cancer lesion detection, Mordang, Gubern-Mérida, den Heeten, and Karssemeijer (2016) used a GentleBoost classifier trained on micro-calcification features describing their shape, topology and texture. This increased the performance of the CADe system in finding micro-calcifications and showed that the same sen-

sitivity at one false positive per 50 cases in the CADe system without BACs removal can be achieved at one false positive per 80 cases in the CADe system with BACs removal.

More recently, Mazidi, Roobottom, and Masala (2019) used a line-strength algorithm followed by a region-growing algorithm to automatically detect and grade BAC on a small dataset of 26 patients (104 images). Despite this, they achieved 82% accuracy in classifying low-grade and high-grade BAC. Classifying four individual grades of severity was less successful at 40.9%, however, although they concluded that this type of system would be useful to radiographers and radiologists in diagnosing and grading BAC.

Since 2015, machine learning techniques in CADe have largely been replaced by deep learning with extracted features being superceded by automatic feature extraction. This has been aided by more robust networks and the ability to handle more complex data (Loizidou, Elia, & Pitris, 2023). The next chapter gives an overview of deep learning and its application in breast imaging in general, finishing with an analysis of the literature relating to its use for BAC detection, classification and segmentation.

4 | Deep Learning

4.1 Introduction

Deep learning is an area of machine learning that uses multi-layer artificial neural networks to *automatically* learn representations that are expressed in terms of other simpler representations with each layer in the network refining the representation into more abstract levels (Gonzalez & Woods, 2018; Goodfellow et al., 2016). Deep learning has subsequently replaced the need for manual segmentation, feature extraction and selection. It uses deep networks with many intermediate layers of artificial neurons between the input and the output which, emulating the visual cortex, learn progressively more complex feature detectors (see Figure 4.1). Feature detectors optimised for classification have allowed deep learning models to outperform systems using hand-crafted features (G. Hinton, 2018).

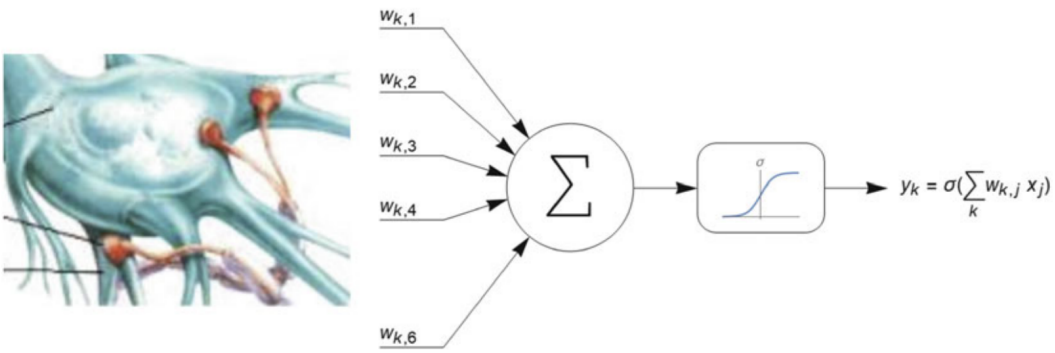


Figure 4.1: Left: synapses connect between neurons. A larger synapse is a larger weight. Repetitions let synapses grow, the basis of learning. Right: a simple artificial neural network summing weighted inputs and passing the output through a non-linear threshold. From ter Haar Romeny (2019).

Artificial neural networks, the technology supporting deep learning, have been around since the 1940s when they were known as *cybernetics*. One of the early proponents, Nor-

bert Wiener, viewed purposive behaviour as arising from a regulatory mechanism trying to minimise “error” (Goodfellow et al., 2016). In the mid-1980s, at least four different groups re-invented Bryson and Ho’s 1969 back-propagation algorithm (Bryson & Ho, 1969) leading to so-called connectionist models of intelligent systems which learnt by adjusting the strength of the connections (weights) between their nodes (neurons) according to some reinforcement learning algorithm (Russell & Norvig, 2013). Wooldridge (2021) believes that the development and potential of this approach at that time was thwarted by the poor performance and limits of simple single-layer perceptrons which were wrongly associated with the more useful multi-layer perceptrons.

LeCun et al. (1989) introduced the LeNet CNN (convolutional neural network) architecture with innovative convolutional and pooling layers to tackle hand-written digit recognition. By 1998 CNNs were state of the art for this task (LeCun, Bottou, Bengio, & Haffner, 1998). Unfortunately they were unable to scale to larger-scale problems due to the computational cost at the time. G. E. Hinton, Osindero, and Teh (2006) improved network performance by adding layers, and subsequently neurons and connections, making the network ‘deeper’, thus neural networks became deep learning.

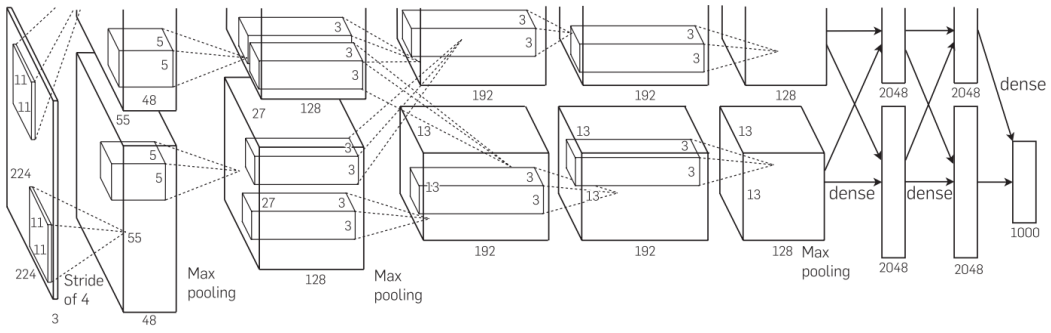


Figure 4.2: AlexNet architecture detail. From Krizhevsky, Sutskever, and Hinton (2012).

Deep learning came to prominence again in 2012 after the AlexNet algorithm won the ImageNet challenge with an error rate that was 41% better than the algorithm which came second (Krizhevsky, Sulskever, & Hinton, 2012). The ImageNet challenge evaluated algorithms for object detection and image classification at large scale. AlexNet consisted of five convolutional layers, using 60 million parameters and 650,000 neurons, and a final 1000-way softmax classifier. The general architecture is shown in Figure 4.2 above. An efficient ap-

proach to GPU (Graphics Processing Unit) implementation, where GPUs only communicate with each other at certain layers, made the algorithm very fast. The use of the then recently discovered (by the same team) regularisation method, dropout, in the fully connected layers prevented over-fitting (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). Deep learning has been applied successfully in areas such as speech recognition, spam filters and drug discovery (Goodfellow et al., 2016). CNNs are now state of the art in image analysis (Soffer et al., 2019).

Despite the above success deep learning is not without critics. Marcus and Davis (2019) decry deep learning's "narrow" AI and describe it as unreliable, greedy (needs a lot of data), brittle (perfect in one situation, wrong in another) and cryptic (experts struggle to understand why certain decisions are made). They believe speech recognition and object detection are not intelligence and that reasoning, language and analogy are not handled by current technology. These views are echoed by Wooldridge (2021) who notes that we have no idea how to interpret the knowledge and representations that neural networks embody.

4.2 Convolutional Neural Networks

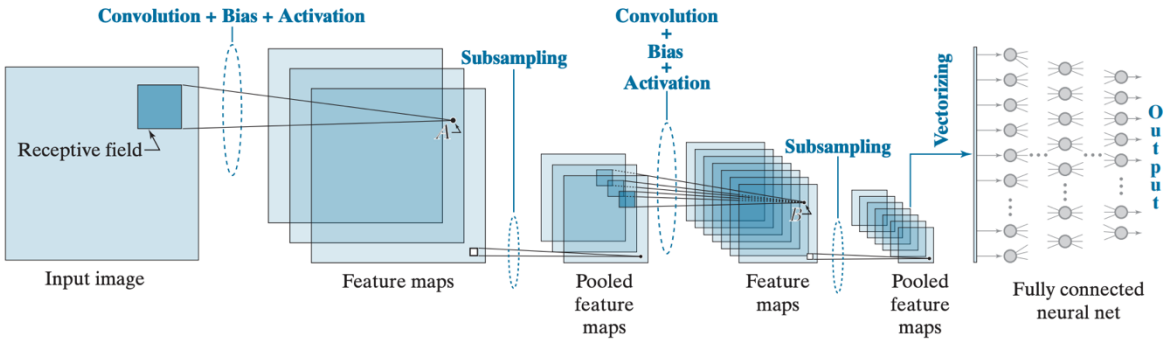


Figure 4.3: A CNN containing all the basic elements of a LeNet (LeCun et al., 1989) architecture. From Gonzalez and Woods (2018).

Convolutional neural networks (CNNs), shown above in Figure 4.3, are a special kind of multi-layer deep learning network for processing data that has a known grid-like topology such as image data (Goodfellow et al., 2016). CNNs are the most widely used model for supervised learning (Litjens et al., 2017) and are particularly suitable for common tasks in

computer vision such as image classification, object detection and segmentation (Soffer et al., 2019). These tasks have been described as “the core of deep learning methods for medical imaging” (P. M. Cheng et al., 2021).

Low-level (high-resolution) features in the images are found by the initial convolutional and pooling layers. As the pooling layers reduce the resolution, these features are combined into higher-level features that represent more complex objects. Therefore, the initial layer(s) may find things like points, lines, and edges which are combined by later layers to identify the main image subjects, e.g. faces or objects (Erickson, 2019). Their width, breadth and height properties allow CNNs to share weights.

4.2.1 CNN Components

CNNs are simply neural networks that use convolutions in place of general matrix multiplication in at least one of their layers (Goodfellow et al., 2016). They compute a sum of products between pixel values and a set of kernel weights at every spatial location in an image and the result at each location (x, y) in the input is a scalar value (Gonzalez & Woods, 2018).

The following sections outline the commonly used layers in CNNs as described by Erickson (2019), Hamidinekoo, Denton, Rampun, Honnor, and Zwiggelaar (2018) and Selvikvåg Lundervold and Lundervold (2018).

4.2.1.1 Input layer

The input layer supplies the convolutional layers with data. Data augmentation, mean-subtraction or feature-scaling may happen here.

4.2.1.2 Convolutional layer

There may be several layers of convolutionals at the input in image-based tasks. They contain three stages of operational units which learn the features needed for successful training:

Convolutional filters: Figure 4.4 shows how the first convolutional layer processes an image. The 3x3 matrix in the middle is the convolutional filter, also known as the *kernel*.

This passes over the image pixel by pixel, multiplying each pixel value in the receptive field by the corresponding value in the kernel and sums them to give the value of the centre pixel in the resulting feature map. The *stride* is the number of spatial increments by which a receptive field is moved and strides greater than one can be used for data reduction. Figure 4.5 illustrates a convolution kernel with optimised weights for edge detection and the resultant image.

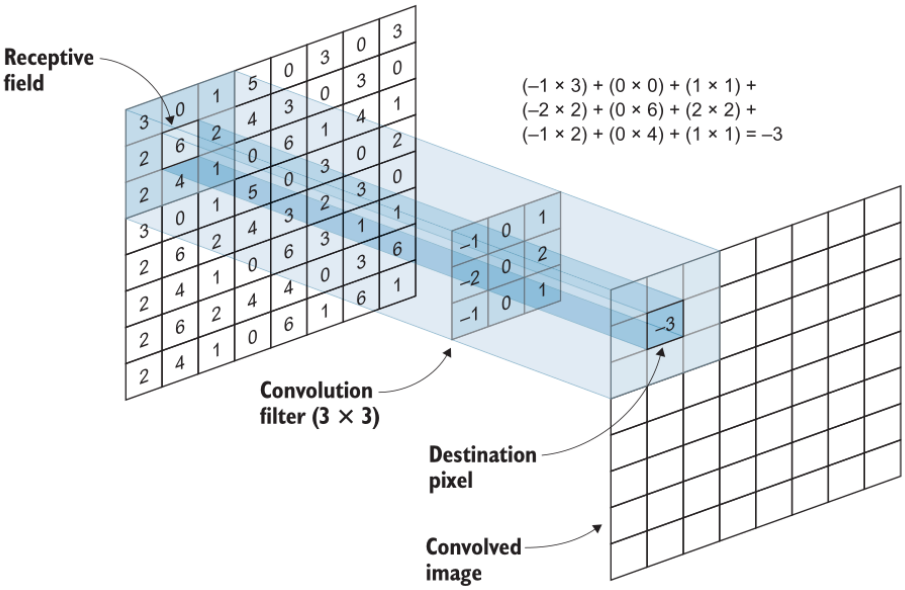


Figure 4.4: How the first convolutional layer processes an image. From Elgendy (2020).

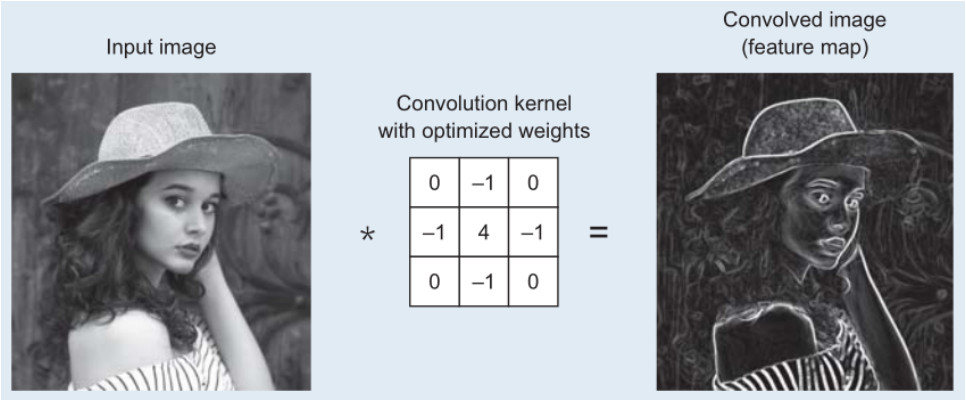


Figure 4.5: Applying an edge detection kernel on an image. From Elgendy (2020).

Pooling: Performs down-sampling for the spatial dimension of the input which results in a reduced-resolution output feature map which is robust to small variations in the locations of

features in the previous layer. The most common pooling function is the “max pool” which simply passes on the maximum value of its current window on to the next layer.

Activation: Feature maps from a convolutional layer are fed through non-linear activation functions making it possible for the entire neural network to approximate almost any non-linear function. Non-linear activation functions ensure there are no large swings in the value of outputs such as those encountered using a “hard” thresholding function that outputs +1 or -1, for example (Gonzalez & Woods, 2018). The latter can cause stability issues in subsequent nodes. Three common activation functions are shown in Figure 4.6.

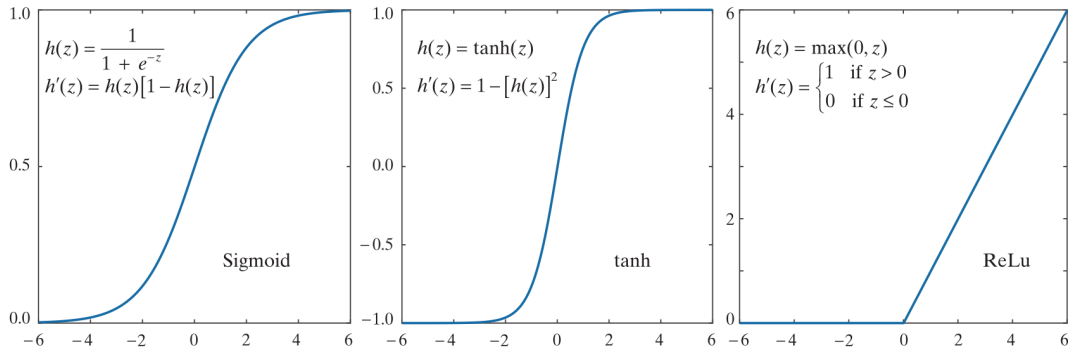


Figure 4.6: Sigmoid, hyberbolic tangent and rectifier linear unit (ReLU) activation functions. From Gonzalez and Woods (2018).

The output of a neuron based on the sigmoid function is

$$a = h(z) = \frac{1}{1+e^{-z}}$$

where z is the result of the computation performed by the neuron. The hyperbolic tangent $(tanh(z))$ has the same shape as the sigmoid function but is symmetric around the axes which improves convergence of the backpropagation algorithm (Gonzalez, Woods, & Eddins, 2020). The Rectified Linear Unit (ReLU) activation function has proved to be very efficient for image processing applications. It outputs zero for any negative input and outputs the input if it is positive. Modified versions include leaky ReLUs which allow a small positive gradient when the unit is not active and exponential ReLUs in which some non-zero output is used for a negative input.

4.2.1.3 Normalisation layer

Usually located after activation layers, producing normalised activation maps by subtracting the mean and dividing by the standard deviation for each training batch. Periodically forces the network to change its activations to zero and unit standard deviation. This speeds up training and negates the need for careful parameter initialisation.

4.2.1.4 Dropout regularisation layer

Dropout is an averaging technique based on stochastic sampling of neural networks. Neurons are randomly removed during training resulting in slightly different networks for each batch of training data. Weights of the trained network are then optimised to these multiple variations of the network.

4.2.1.5 Fully connected layers

Compute the final output. They treat their input as a simple vector and output a single vector.

4.2.1.6 Residual layer

Uses a "bypass" layer which is basically the identity function. This is compared to the output of a layer or a group of layers. Reduces the number of potential parameters to adjust when learning and also reduces the likelihood of overfitting to the training data.

4.2.2 CNN Training

The objective of training a CNN is to minimise the difference between the predicted output and the actual output of the network. A forward pass through the network is performed to classify all the patterns of the training set and to determine the classification error (Goodfellow et al., 2016). Backpropagation directs this error back through the network to update the weighted parameter values (Hamidinekoo et al., 2018). There are four steps in the process (Gonzalez et al., 2020):

1. Input $a(0)$, the set of image pixel values in the input volume, to layer 1.

2. For the feed forward pass, for each neuron corresponding to location (x, y) in each feature map in layer ℓ compute:

$$z_{x,y}(\ell) = w(\ell) \star a_{x,y}(\ell - 1) + b(\ell)$$

$$a_{x,y}(\ell) = h(z_{x,y}(\ell)); \ell = 1, 2, \ldots, L_c$$

where $a_{x,y}(\ell)$ is the neuron output and $w(\ell)$ is the weight and $b(\ell)$ is the bias. $a_{x,y}(\ell)$ is obtained by passing $z_{x,y}(\ell)$ through an activation function h. "$\star$" denotes a convolution. L_c is the number of convolutional layers.

3. For backpropagation, for each neuron in each feature map in layer ℓ compute:

$$\delta_{x,y} = h'(z_{x,y}(\ell))[\delta_{x,y}(\ell + 1) \star \text{rot180}(w(\ell + 1))]; \ell = L_c - 1, L_c - 2, \ldots, 1$$

4. Update the weights and biases for each feature map using:

$$w_{l,k}(\ell) = w_{l,k}(\ell) - \alpha\delta_{l,k}(\ell) \star \text{rot180}(a(\ell - 1))$$

and

$$b(\ell) = b(\ell) - \alpha \sum_x \sum_y \delta_{x,y}(\ell); \ell = 1, 2, \ldots, L_c$$

where α is the learning rate and k and l are the dimensions of the kernel.

This process is repeated until the error reduces to an acceptable level.

4.2.3 CNN Architectures

4.2.3.1 Classification

Examples of popular CNNs for image classification include VGGNet (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015) and ResNet (He, Zhang, Ren, & Sun, 2016):

- VGGNet: The architecture of this network is known for its simplicity of alternating convolution and dropout layers. It was the first to use multiple 3 x 3 filters in each convolutional layer in comparison to AlexNet's 11 x 11 kernels. Although simple in architecture,

it is computationally expensive in terms of memory as increasing kernels is associated with higher computational time and a larger model (Chougrad, Zouaki, & Alheyane, 2018).

- GoogLeNet: The objective of GoogLeNet was to reduce computational complexity compared to other CNNs. It did this by using "inception modules", shown in Figure 4.7 below, containing multiple convolutional filter sizes in each block. Stacking the modules on top of each other and using occasional max-pooling layers with a stride of 2 allowed the resolution of the grid to be halved. It did this by using the 1x1 convolutions before the more expensive 3x3 and 5x5 convolutions. GoogleNet had more layers (22) than any CNN before but had much less network parameters (7M) than AlexNet (60M) or VGG-19 (138M) (Alom et al., 2018).

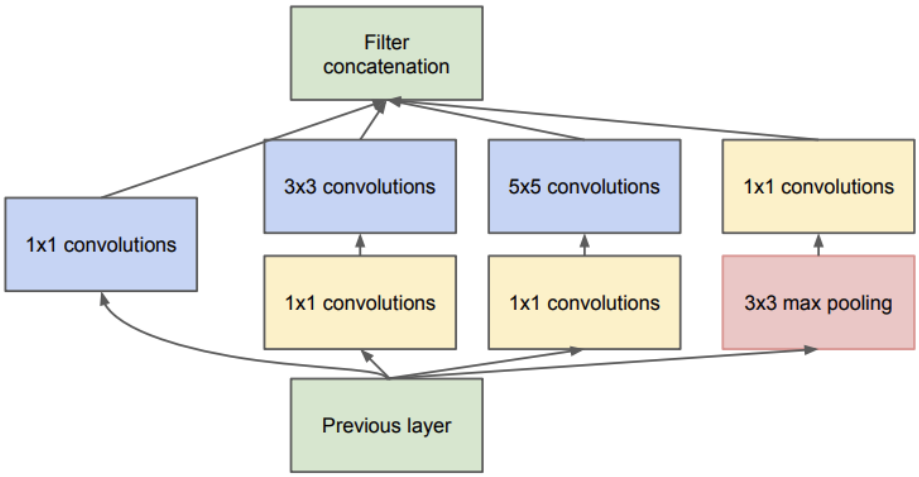


Figure 4.7: Inception module. From Szegedy et al. (2015).

- ResNet: Residual networks take a standard deep CNN and add shortcut ("skip") connections that bypass a few convolutional layers at a time. The shortcut connections create residual blocks, shown overleaf in Figure 4.8, where the output of the convolutional layers is added to the block's input tensor. Having direct connections that shortcut the convolutional layer allows gradients to more easily flow backward through the network during training (Szeliski, 2022) thus solving the issue of vanishing or exploding backpropagated gradients when using deeper networks (Glorot & Bengio, 2010).

Computational overhead is kept low by these connections which also provide a rich combination of features (Chougrad et al., 2018).

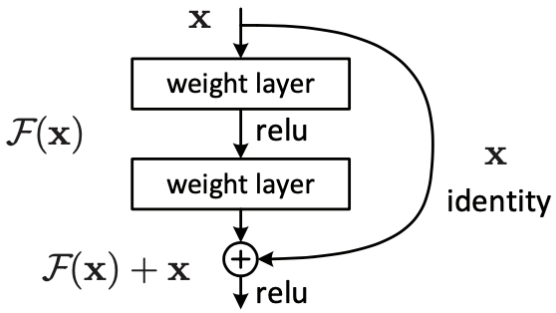


Figure 4.8: Residual learning building block. From He et al. (2016).

4.2.3.2 Object Detection

Faster R-CNN (Ren, He, Girshick, & Sun, 2017) is a network for object detection that built on R-CNN (Girshick, Donahue, Darrell, & Malik, 2014) and Fast R-CNN (Girshick, 2015) by utilising a CNN to extract features and obtain regions of interest using a region proposal network (RPN) that shared full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network (FCN) that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection.

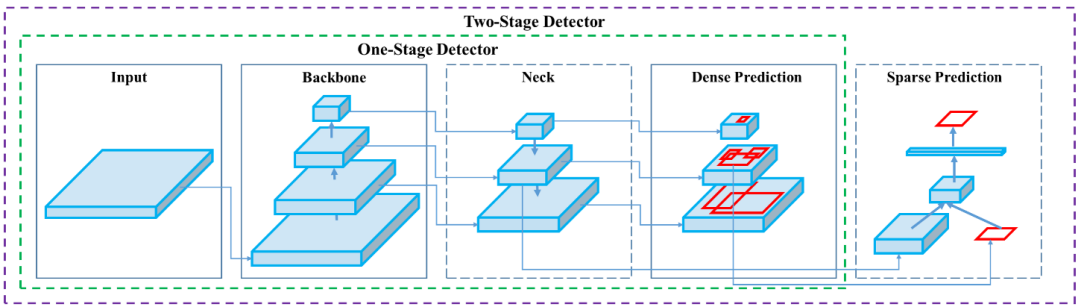


Figure 4.9: YOLOv4 “You Only Look Once” Object Detector. From Bochkovskiy et al. (2020)

YOLOv4 (Bochkovskiy et al., 2020), shown in Figure 4.9, also uses a feature extraction backbone and combines this with a neck and head, the latter which can be one-stage or

two stage. The neck comprises a spatial pyramid pooling (SPP) module and a path aggregation network (PAN). These identify the most relevant features and aggregate feature maps and pass them on to the head which predicts bounding boxes, objectness scores and classification scores.

4.2.3.3 Segmentation

The encoder-decoder architecture has been widely used for pixel-level medical image segmentation (Su, Zhang, Liu, & Cheng, 2021). U-Net (Ronneberger, Fischer, & Brox, 2015) is a popular exemplar of this structure. With this model, named after its shape, there are two steps to feature extraction, down-sample and up-sample. The image is reduced to the key component being searched for, this is found at the bottom of the "U". Once identified, "bypass layers" using pixel data from higher resolution versions refine the key component until the original resolution is achieved. Segnet (Badrinarayanan, Kendall, & Cipolla, 2017) improves U-Net's memory overload by pooling indices transferred from the compression path to the expansion path.

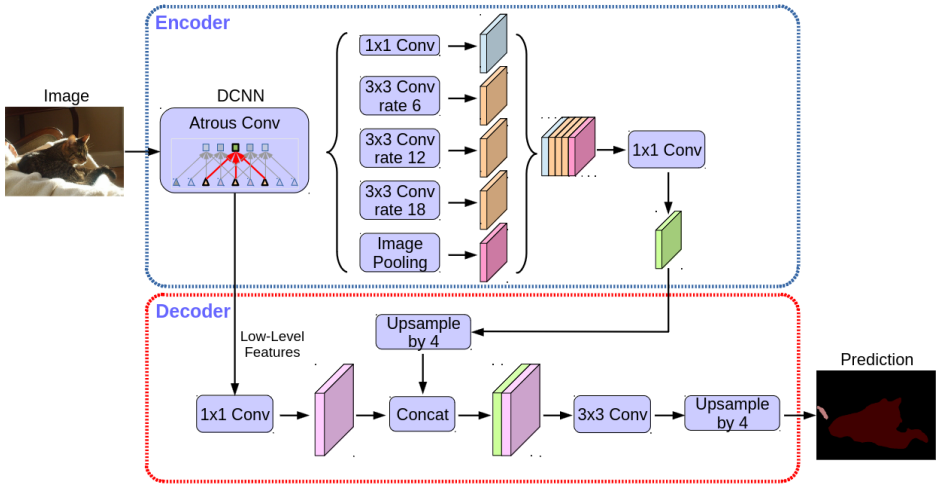


Figure 4.10: DeepLabv3+ encoder-decoder architecture. From L. C. Chen, Zhu, et al. (2018).

DeepLabv3+ (L. C. Chen, Zhu, et al., 2018), shown in Figure 4.10, combines the benefits of an encoder-decoder architecture and a spatial pyramid pooling module to provide multiple effective fields-of-view and sharper object boundaries. Dilated convolutions aggregate context around a feature allowing it to be better segmented. DeepLabv3+ performed best in a recent study comparing networks (including U-Net and SegNet) for segmenting MRI

(Magnetic Resonance Imaging) prostate studies, achieving a Dice score of 92.8% (Z. Khan, Yahya, Alsaih, Ali, & Meriaudeau, 2020).

4.2.3.4 Generative Adversarial Networks

Generative adversarial networks (GANs) comprise two competing networks which are usually CNNs. A generative network creates samples that a discriminative network classifies as originating in the generative network or in the training data (Selvikvåg Lundervold & Lundervold, 2018). Originally GANs were used to test deep learning systems but have proved useful in medicine in a number of ways including the creation of additional training and testing images and providing interpretability of both deep learning systems and disease detection (Erickson, 2019).

4.3 Deep Learning in Medical Imaging

In a review of deep learning in medical imaging, Soffer et al. (2019) note that deep neural networks have been used for a wide range of tasks in including organ classification and disease detection, image segmentation and optimisation. CNNs are the core of deep learning methods for imaging with classification, object detection and segmentation being the three most popular applications (P. M. Cheng et al., 2021). Classification refers to categorisation of a specific group or type of lesion(s) from one class to others while object detection highlights a specific subregion in an image which is likely to contain a localised lesion (Montagnon et al., 2020). Segmentation delineates or volume extracts a lesion or organ based on image analysis. Mazurowski, Buda, Saha, and Bashir (2018) believe that recent successes are due to availability of data, increased processing power, rapid development of algorithms and the availability of GPUs.

Despite the above successes, Mukhlif, Al-Khateeb, and Mohammed (2022) note that there are challenges too as large datasets are needed and labelling images takes a significant amount of time and effort. Datasets also may have quality issues such as a lack of diversity and large class imbalances.

Research publications in the area are increasing year on year. The following sections will briefly describe their impact and application in breast imaging and, more specifically, BAC detection.

4.4 Deep Learning in Breast Imaging

A number of network architectures have been used in recent breast imaging literature including CNNs, GANs and an extended U-Net.

Liu et al. (2018) used an Inception V2 ResNet pre-trained on ImageNet to automate breast density reporting. This network builds on the Inception architecture by replacing filter concatenation with ResNet skip connections. They investigated training specific views separately but found the network was more accurate whenever all views were trained together. Mohamed et al. (2018) also used a CNN, this time based on AlexNet, to differentiate "scattered density" and "heterogeneously dense" BI-RADS (Breast Imaging Reporting and Data System) breast density categories. They obtained an area under curve (AUC) of 0.9421 after training the network from scratch and an AUC of 0.9265 using a pre-trained model and only 500 images.

Yi et al. (2018) used a ResNet pretrained on ImageNet to investigate predicting the image view, laterality (right or left), breast density and benign versus malignant masses. Accuracy for view and laterality was 99% while that for both breast density and lesion analysis were found to be 68%. Pan, Chu, Wang, Merck, and Lourenco (2018) used a MobileNet CNN, intended for embedded mobile applications, in order to predict breast nodule malignancy from ultrasound images. They fine-tuned the network using data augmentation, dropout probability and the learning rate. A mean AUC of 0.869 was achieved and at a threshold of 0.10, the authors state the model would reduce the number of negative core biopsies by 40% while maintaining 95% sensitivity.

Singh et al. (2018) used a conditional generative adversarial network (cGAN) to optimise breast mass segmentation and shape classification. The generative network learned the intrinsic features of tumours while the adversarial network forced segmentations to be similar

to ground truth. The Dice co-efficient was found to be > 94%. Shams, Platania, Zhang, Kim, and Lee (2018) combined CNNs with GANs in order to overcome the limitations of annotations from real medical settings. They found that enhanced feature learning with a GAN coupled with hybrid training with regions of interest resulted in more accurate classification. They recorded an AUC of 92.5% for the small INBreast (Moreira et al., 2012) dataset.

S. Mehta et al. (2018) used a Y-Net (extended U-Net) for segmentation and classification of breast biopsy images. The Y-Net extends U-Net by adding parallel branch discriminative map generation and by supporting convolutional block modularity. This allows the user to adjust the network without changing the network topology. The authors report state-of-the-art segmentation with almost seven times reduction in parameters than its closest rival. An increase of 7% in classification accuracy over the current state-of-the-art was also recorded.

Three recent large studies have showed excellent results in relation to breast screening mammography. Using over a million images from 220,000 mammogram studies Wu et al. (2019) achieved an AUC of 0.895 in predicting the presence of cancer when tested on the screening population. They believe their success was due to:

- A two-stage architecture and training procedure incorporating a patch-level network based on pixel-level labels.
- A custom ResNet optimised for large images.
- Pre-training on screening BI-RADS classification.
- Combining multiple input views.

They also found the combination of the model and a human reader was more accurate than either of the two separately.

Lotter et al. (2019) presented an annotation-efficient approach that out-performed five of five full-time breast imaging specialists by improving absolute sensitivity by an average of 14%. Their ResNet50-based model demonstrated state-of-the-art classification which also extended to digital breast tomosynthesis. They note the challenges inherent in deep learning and breast imaging – severe class imbalance (95% of images are normal), datasets are costly

and impractical to collect, high resolution images that are ten to twenty times the resolution of natural image datasets.

A cross-Atlantic study by McKinney et al. (2020) using over 30,000 cases also showed a deep learning system capable of surpassing human experts in breast cancer prediction. Significant reductions were achieved in both false positives and false negatives. In conjunction with a human reader, the AI maintained a non-inferior performance and reduced the workload of the second reader by 88%.

Y. Chen, Taib, Darker, and James (2023) evaluated human readers and a commercially available AI algorithm (Lunit Insight MMG (*Lunit Inc.*, n.d.)) reading two test sets of 60 challenging breast imaging cases and found the diagnostic performance equal. They note that although AI is not currently ready for deployment outside of clinical trials, it seems increasingly likely that AI will eventually play a part in the interpretation of screening mammograms. When this occurs, it may also be prudent to screen for arterial calcification. The next section examines the literature relating to using deep learning to do just that.

4.5 Deep Learning and BAC

Table 4.1 on page 42 shows all the papers that have investigated the feasibility of automated and accurate BAC detection on mammograms using deep learning methods. There are a diverse range of networks, datasets, metrics and approaches.

J. Wang et al. (2017) were the first to do so. The problem was scoped as a pixelwise, patch-based, two-class classification problem based on the presence or absence of BAC. Patches around each pixel were directly entered into a custom 37-layer CNN trained to classify whether the central pixel belonged to a BAC class or not. Performance was measured using a two-round human reader study, free-response receiver operating characteristic (FROC) and calcium quantification analysis. At a true positive rate of 60%, the area of false positives was found to be 0.4762cm^2, comparable to one of the human readers. Calcium quantification analysis was used to compare the predicted calcium to ground truth. The coefficient of determination R^2 of the fitted model was found to be 0.9624, indicating a close correlation.

K. Wang et al. (2019) tested deep learning techniques for BAC detection and segmentation but concluded that a more traditional Hessian-based multi-scale filter coupled with a self-adaptive thresholding algorithm showed the highest accuracy on validation data. The deep learning algorithms, YOLOv2 for detection and UNet and DeepLabv3+ for segmentation, performed extremely poorly, achieving an IoU of 0, 0 and 0.05 respectively. Their small dataset contained both FFDM and DBT (Digital Breast Tomosynthesis) unprocessed images which may have contributed to this. With DBT,https://www.github.com a sequence of shorter exposures is made as the x-ray tube gantry moves through an arc rather than a single exposure like that used to create an FFDM image (Sechopoulos & Sá dos Reis, 2022). A pseudo-3D image can be created using DBT as well as a single synthetic planar image. The latter is the type used in this study. The authors note that detection of BAC can be challenging due to its topological complexity and the fact that its intensity values are not uniform and are sometimes similar to healthy breast tissue.

More promising results were achieved by two later studies using variations of the U-Net model. Ghamdi et al. (2020) extended the U-Net model with DenseNet blocks that, unlike ResNets, concatenate outputs from the previous layers instead of using summation before passing them on to the next layer. The patch-based model was evaluated using 5-fold cross-validation and achieved a test accuracy of 91.47%, sensitivity of 91.22% and specificity of 92.01%, out-performing human experts. These results were achived using what the authors describe as a “small-sized and low-quality dataset”. Their dataset is the only one in all the papers that came from a publically-available source, the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) (R. S. Lee et al., 2017). Images in this dataset were originally analog screen film images that were scanned in and converted to digital.

Guo et al. (2021) adapted the U-Net model to make it deeper and thinner by extending the down-sampling layers to five and by using only a quarter of the convolutional channels compared to the standard U-Net. This reduced the number of parameters to train from 13,395,329 to 7,782,281. They used 512x512 patches that were concatenated together to recover the whole segmentation mask for the original large-size FFDM image. Excellent

results were achieved for accuracy (>99.8%) but less so for precision (<69%). They also devised their own metrics to quantify the predicted calcium mass and obtained a quantitative correlation of over 95%. The authors note that such measurements of breast arterial calcification can offer a personalized, non-invasive approach to risk-stratify women for cardiovascular disease at no additional cost or radiation.

Two recent studies also recorded satisfactory results although both seem to have issues with their methodology. R. Khan and Masala (2023) investigated using deep learning to classify four grades of BAC severity. They tested muliple networks and achieved a test accuracy of 94% using MobileNet. They seem to have used their full dataset of 104 images for the validation and test data and created augmented images for the training set from the same data. Furthermore, they resized their images to 300x400 which is 2% and 1.2% of their two original image sizes, 2082x2800 and 2800x3518.

Using a VGG-16-based network, Mobini et al. (2023) obtained a test accuracy of 94% with an AUC-ROC of 0.98. They also found a strong correlation between the Grad-CAM++ predicted lengths of BAC and manual measurements. Their training, validation and test data were imbalanced class-wise, however, with the BAC+ve to BAC-ve ratio being 24% to 76% in the training set and 10% to 90% in both the validation and test sets. They state the ratios reflect the prevalence of BAC in clinical cohorts.

The largest dataset used in the literature was the 6573 unprocessed DBT images used by K. Wang et al. (2023) in their study investigating BAC segmentation using UNet and DeepLabv3+ networks. The data was divided in to a 80:10:10 train:validate:test split and images were initially annotated by non-experts resulting in a Dice similarity score of 0.3771. After finding many false positives such as areas of microcalcifications, they developed an automatic label correction algorithm for the training set and also manually re-annotated the test set using a domain expert. The Dice score subsequently increased to 0.4849. They also developed a novel length-based Dice score to negate any large negative effects of small differences in predicted BAC areas and ground truth as well as recognising that BAC is clinically measured by length rather than by area. The length-based Dice score was found to be 0.6261. They conclude that deep learning models have shown promise in BAC segmen-

tation and this promise can be optimised by using larger image sizes, larger models, better quality annotations and appropriate image contrast adjustment.

In October 2023, Curemetrix® were granted approval by the FDA for a BAC detection and localisation application (Hall, 2023). Unfortunately, no research paper has, as yet, been published on it making it impossible to know whether machine or deep learning techniques were used in its development. It will be the first of many.

4.6 Gap Analysis

The literature review identified a number of gaps, namely:

- No publically available dataset of BAC-annotated mammograms.
- No BAC studies using full-size mammography images as deep learning network inputs.
- No BAC studies that used a pre-trained breast cancer detection model for transfer learning and weight initialisation.
- A class imbalance in studies with some papers only using BAC-positive images for classification and others not having equal numbers of each class (BAC and non-BAC).
- Only one BAC paper used object detection techniques.

Chapter 5, Experimental Set-up, and the subsequent model development chapters outline how these gaps were addressed.

Paper	Task	Network	Dataset	Images	Modality	Input Type	Results
J. Wang et al. (2017)	Classification, Segmentation	Custom CNN	Restricted	840	FFDM	Patch, 95x95	True Positive Rate: 60%
K. Wang et al. (2019)	Object Detection, Segmentation	YOLOv2, U-Net, DeepLabv3+	Restricted	135	FFDM, DBT,	Image, Size Not Known	IoU (Object Detection): 0 IoU (Segmentation): 0.05
Ghamdi et al. (2020)	Segmentation	DU-Net	CBIS-DDSM	826	FFDM	Patch, 640x640	Jaccard Index: 0.8517 Accuracy: 09147 F1-Score: 0.9219
Guo et al. (2021)	Segmentation	SCU-Net	Restricted	661	Scanned	Patch, 512x512	Jaccard Index: 0.583 Accuracy: 0.998 Precision: 0.685 Recall: 0.826 Quant. correlation: >95%
R. Khan and Masala (2023)	Classification	VGG-19, ResNet50, DenseNet121, InceptionV3, MobileNet	Restricted	104	FFDM	Image, 300x400	Accuracy in detecting BAC severity: 94%
Mobini et al. (2023)	Classification, Segmentation	VGG-16	Restricted	3444	FFDM	Image, 1536x768	Accuracy: 0.97 F1-Score: 0.80 AUC-ROC: 0.95 Grad-CAM++ BAC length correlation: $\rho = 0.88$, $p < 0.001$
K. Wang et al. (2023)	Segmentation	U-Net, DeepLabv3+	Restricted	6573	DBT	Image, 1600x1600, 1024x1024	Dice Score: 0.4849 Length-based Dice Score: 0.6261

Table 4.1: Papers using deep learning to classify, detect or segment breast arterial calcification.

4.7 Summary

In Chapter 2, we looked at the background to BAC and its association with cardiovascular risk factors and atherosclerosis. We then described computer-aided detection (CADe) in mammography and the machine learning techniques to detect BAC. This was followed, in this chapter, by an overview of deep learning and CNNs culminating in a review of the literature relating to deep learning and BAC detection, classification and segmentation. The latter informed our methodology in the next chapter in relation to the use of full-size, high quality mammography images in network training, a proper class balance and a validated annotation process.

5 | Experimental Set-up

This chapter outlines the experimental basis of the study beginning with the dataset and its annotation for BAC at an image, region and pixel level. Image pre-processing is then described including cropping, padding, image conversion, data augmentation and the application of a technique to reduce image noise and increase contrast. Finally, the software and hardware development environments used are detailed. The experimental set-up provides the necessary preparation for the model development chapters that follow.

5.1 Dataset

DICOM (Digital Imaging and COmmunication in Medicine) is a collection of standards that provide all the necessary tools for the diagnostically accurate representation and processing of medical image data (Pianykh, 2008). In addition to data transfer, storage and display protocols, it also encompasses an image file format. Each DICOM file has a header containing a diverse range of data including patient demographic information, acquisition parameters, referrer, practitioner and operator identifiers and image dimensions. The rest of the DICOM file contains the image data (Graham, Perriss, & Scarsbrook, 2005) as shown in Figure 5.1.

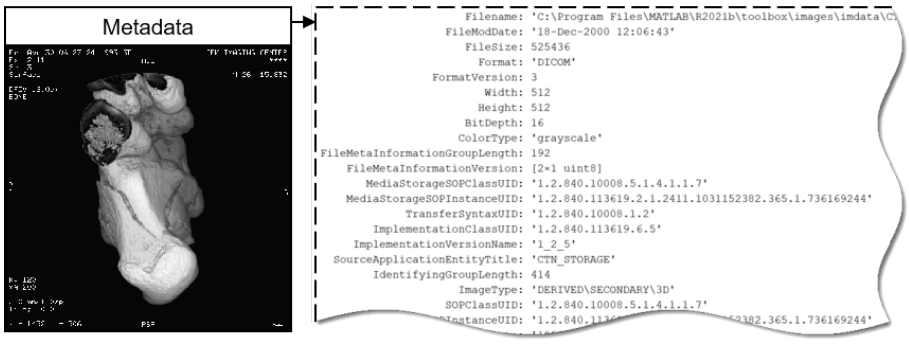


Figure 5.1: DICOM image file showing header and image data. From Mathworks (2023a).

There are currently no freely available BAC-annotated DICOM mammography datasets. Only one paper (Ghamdi et al., 2020) has used a public dataset, CBIS-DDSM (R. S. Lee et al., 2017), as a source of images for training deep learning models for BAC classification, detection or segmentation. Their images, described by the authors as "low quality", were originally produced using an analog screen film combination and were subsequently digitized. The authors also had to retrospectively annotate and segment the images for BAC.

The dataset used in this study was derived from the OPTIMAM Mammography Image Database (Halling-Brown et al., 2021), a collection of high quality FFDM (Full Field Digital Mammography) DICOM images collected from three NHS (National Health Service) clinical centres, and was used under a purchased licence from Cancer Research UK. The latter accept applications from commercial or non-profit organisations, healthcare institutions and academic centres for access to the database. Access is granted based on the following criteria:

- Scientific Merit - For example, if the scientific aims are achievable and realistic.
- Trustworthiness - What procedures are in place with regards to data storage security?
- Reputation - Does the Research Group have expertise in the subject area?
- Ethics - Is ethical approval present/being sought if applicable?

Ethical approval was granted in January 2019 by the University of Salford Research Ethics Panel and was amended in August 2021, post-COVID, in order to allow the recruitment of participants from clinical colleagues in the author's own hospital group in Ireland. The amendment also allowed an annotation observer study to be carried out at a suitable, local viewing facility there too. Both ethical approval and amendment douments are shown in Appendix A. The dataset licence granted access for three years from June 2019 (date of signature) and the signed cover is shown in Appendix B.

The dataset contained both processed ("for presentation") DICOM images and their unprocessed ("for processing") counterparts, typically of women who had a breast-screening examination. The data was pseudonymised at the point of collection using DICOM Standard 142 supplement compliance tools which outline clinical trial de-identification profiles.

The Information Commissioners Office states that the collection of de-identified data without patient consent is permissible provided there is “no likelihood of anonymisation causing unwarranted damage or distress” (Information Commisioners Office, 2012). The Royal College of Radiologists guidance reiterates that explicit consent is only required if the patient is, or may be, identifiable (Royal College of Radiologists, 2017).

Metadata such as patient age, image size and manufacturer were interrogated using the `pydicom` (Mason et al., 2020) Python library although an accompanying dataset API (application programming interface) is now available. The dataset comprised of FFDM images from 600 patients with each case containing at least four images, i.e. standard CC (craniocaudal) and MLO (mediolateral oblique) views of each breast as shown in Figure 5.2. Some cases included repeat views or additional views to cover all the anatomy. The total number of images was 2437. These were accompanied by clinical data from the NHS Breast Screening Programme’s National Breast Screening System (NBSS) with DICOM header tags extended to include expert annotations relating to cancer lesion status, position and classification. None of the clinical data or annotations included information relating to BAC so were not used in this study.

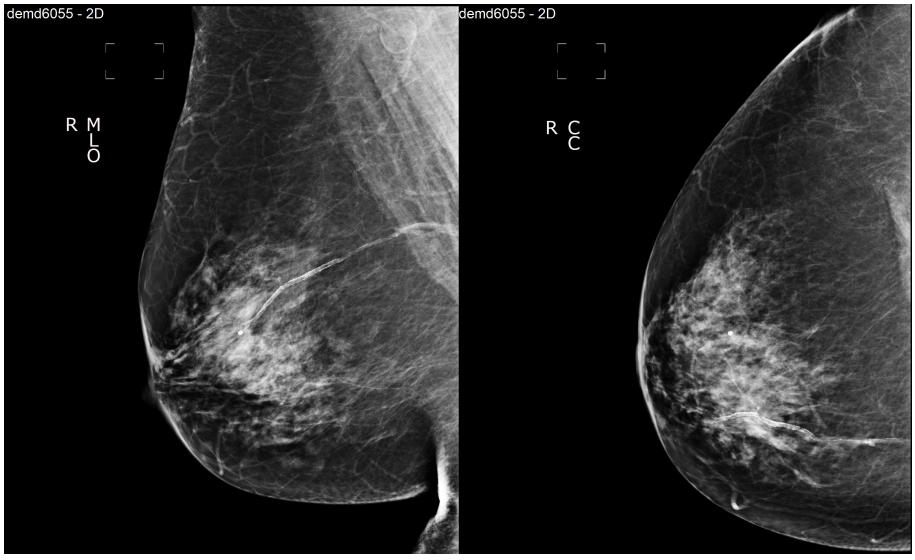


Figure 5.2: Left to right: MLO and CC standard mammography views of the right breast.

Because BAC is associated with increasing age, the age range of cases was limited to between 65 and 70 years in order to have a suitable number of images containing BAC. Images were produced using hardware from two companies, Hologic, Inc and GE Medical systems

(two machines each), and cases were divided equally among the four Volpara-designated (Highnam, Brady, Yaffe, Karssemeijer, & Harvey, 2010) breast composition categories from low to high density: A (almost entirely fatty), B (scattered areas of fibroglandular density), C (heterogeneously dense) and D (extremely dense). Volpara uses a combination of mammography physics and machine learning to generate an objective volumetric measure of breast composition. A summary of the number of cases per age range including breast density category is shown in Figure 5.3.

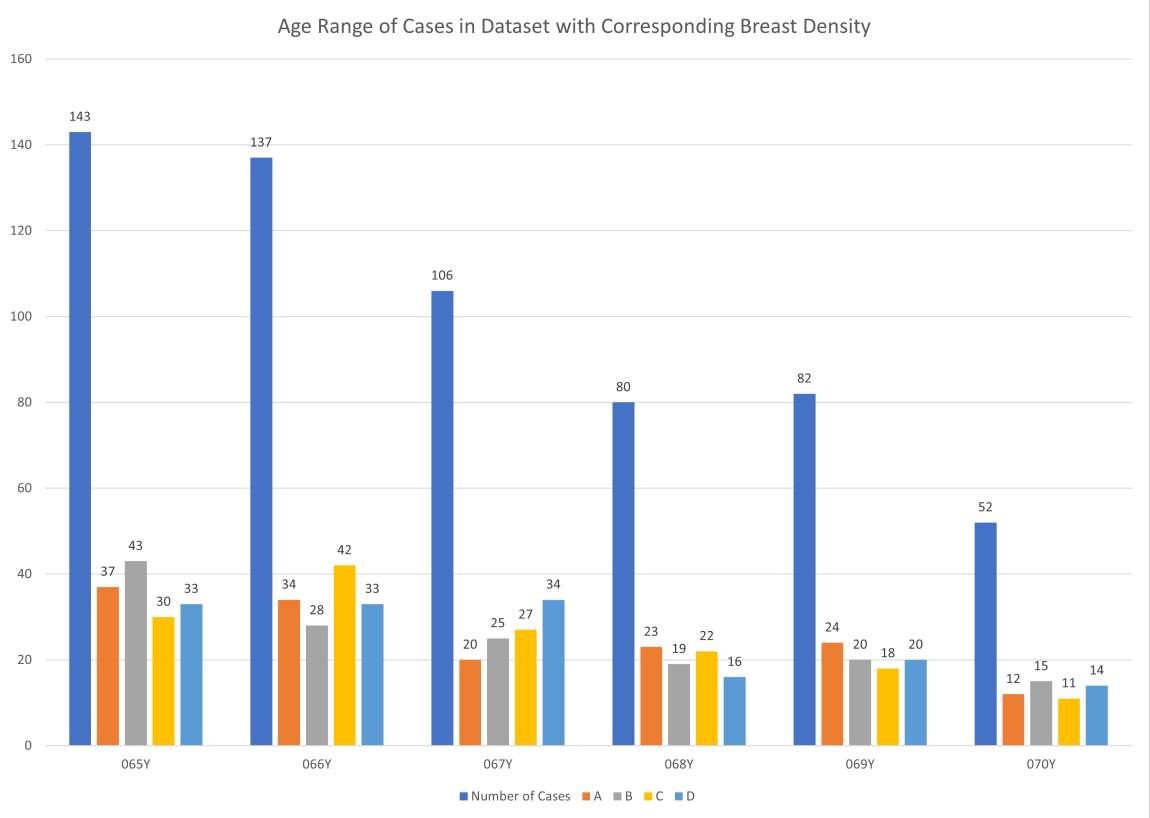


Figure 5.3: Number of cases per age range with associated breast density categories.

5.2 Image-wise Annotation

Willemink et al. (2020) note that the term *ground truth* usually refers to information acquired from direct observation such as biopsy or laboratory results although annotations such as image labels performed by radiologists can be considered ground truth if imaging is the reference standard, e.g. as in pneumothorax (collapsed lung). Such annotation of reference standard training data is costly and time consuming for relevant clinically experienced pro-

fessionals, however (P. Mehta et al., 2019). Some studies have addressed these challenges by using non-expert crowdsourced annotators with one pilot study finding that the latter were able to detect and refine innacurate liver contours with a quality similar to experts (engineers with domain knowledge, medical students and radiologists) (Heim et al., 2018). In this study, the author, who has almost 25 years medical imaging domain knowledge, albeit not in mammography, carried out the BAC annotation. This was then validated by two consultant radiologists in an observer reader study.

Images were manually examined and categorised using a demonstration version of MedX Viewer (Medical eXtensible Viewer, (Looney, Young, & Halling-Brown, 2016)), a software tool primarily used for breast cancer lesion detection remote reader studies. The software allows users to draw bounding boxes on FFDM DICOM images and create questions that easily facilitate image labelling. Answers and box coordinates could then be exported as a JSON (JavaScript Object Notation) file for easy import into MATLAB. A data export from one of the annotation validation reader studies is shown in Appendix D.

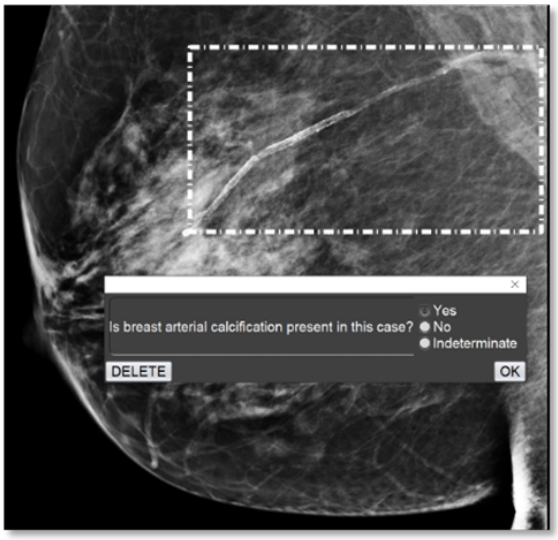


Figure 5.4: Using MedXViewer to annotate and label images.

A small number (n=5) of images were discarded due to the presence of cardiac pacemakers or silicon implants. Images were labelled as “BAC” or “NON_BAC” with bounding boxes drawn around BAC-positive areas (as shown in Figure 5.4). 127 (21.2%) cases were found to have at least one BAC-positive image with 408 (16.7%) BAC-positive images in the dataset as a whole.

5.2.1 Annotation Validation Reader Study

Annotation ground-truth was validated in a reader study by two consultant radiologists with 8 and 3 years breast imaging experience respectively. Participation in the validation study was requested via a formal letter. An information sheet outlining the rationale of the study was also provided as were contact details for the author and his supervisors. Participants were invited to sign a consent form. Confidentiality was assured and participants had the right to withdraw at any stage. An information sheet, risk assessment form and two signed consent forms for the reader study are shown in Appendix C.

The reader study was conducted on two separate days in a radiology reporting room in the author's own hospital group. The study was carried out under optimal viewing conditions with dual 5MP (megapixel) Barco® monitors and appropriate ambient room lighting (less than 10 lux) as outlined in the Royal College of Radiologists' guidelines on diagnostic display devices (Royal College of Radiologists, 2019). Participants were also given basic training on the MedXViewer software. Covid-19 regulations were adhered to at all times.

Each participant was given 10 randomly selected cases of 4 images (n=40 images) each. 5 cases were BAC-postive and 5 were BAC-negative. A total of 80 images were used in the study. As in Figure 5.4 above, participants were asked to mark the presence of BAC with "Yes", "No" or "Indeterminate". They were also asked to provide a rectangular bounding box around the area(s) in which they thought BAC was present. Each participant took approximately one hour to complete the tasks.

To assess the reliability of the author's BAC annotations, a widely used measure of inter-rater reliability , namely Cohen's Kappa Co-efficient (Cohen, 1960), was used. This was found to be 0.85 and 0.9 between the annotator and each radiologist respectively, denoting near perfect agreement. A summary of these results is shown in Figure 5.5.

Case	Rater 0	Rater 1	Result	R MLO	L MLO	R CC	L CC
demd133084	No	No	Agree				
demd139620	No	No	Agree				
demd9697	Yes	Yes	Agree				
demd101008	No	Yes*	Disagree				
demd57707	Yes	Yes	Agree				
demd127928	Yes	Yes	Agree				
demd58044	No	No	Agree				
demd124129	No	No	Agree				
demd134625	Yes	Yes	Agree				
demd103365	Yes	Yes	Agree				

Case	Rater 0	Rater 2	Result	R MLO	L MLO	R CC	L CC
demd126320	Yes	Yes	Agree				
demd139399	Yes	Yes	Agree				
demd139860	No	No	Agree				
demd125989	Yes	Yes	Agree				
demd140609	No	Yes*	Disagree				
demd113666	Yes	Yes	Agree				
demd71486	No	No	Agree				
demd37105	No	No	Agree				
demd120047	Yes	Yes*	Agree				
demd140972	No	No	Agree				

$\kappa = 0.9$

$\kappa = 0.85$

Cohen's Kappa Statistic

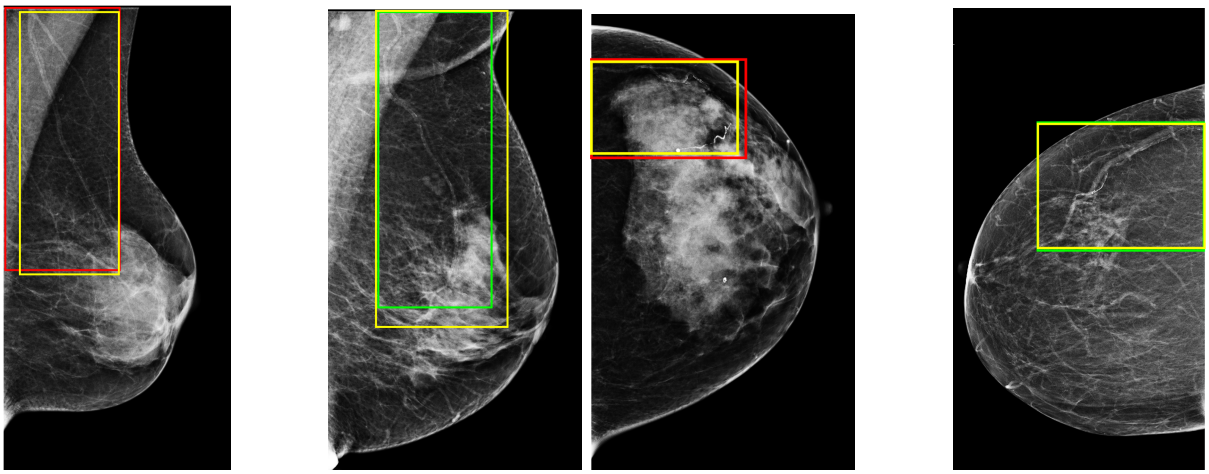
- 0 = agreement equivalent to chance.
- 0.1 – 0.20 = slight agreement.
- 0.21 – 0.40 = fair agreement.
- 0.41 – 0.60 = moderate agreement.
- 0.61 – 0.80 = substantial agreement.
- 0.81 – 0.99 = near perfect agreement
- 1 = perfect agreement.

Figure 5.5: Inter-rater reliability for BAC annotation results.

Each bounding box created by the two radiologists was compared to those outlined by the annotator. The metric *intersection over union (IoU)* was used to measure similarity between bounding boxes. IoU computes the ratio as the area of intersection between one box and another, divided by the area of the union of the two:

$$a_o = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$$

where B_p is the predicted area and B_{gt} is the ground truth. Everingham et al. (2015) note that IoU must exceed 50% to be considered a correct detection. The average IoU was found to be 0.51 and a detailed summary of the results is shown in Appendix E. A visual comparison of the author's bounding boxes compared to those of the radiologists is shown in Figure 5.6.



(a) $IoU = 0.84$ (b) $IoU = 0.80$ (c) $IoU = 0.88$ (d) $IoU = 0.96$

Figure 5.6: Comparison of several BAC bounding boxes of the author (yellow) and the two radiologist (red and green) readers.

5.3 Pixel-wise Annotation

A number of traditional segmentation methods implemented in MATLAB by Gonzalez et al. (2020) were investigated to try and lessen the manual pixel-labelling task. These included active contours, Frangi vesselness filter, k-means clustering, Sobel edge enhancement and graph-cut segmentation. Each method was evaluated visually for BAC enhancement with active contours being the most helpful. Despite this, the latter method increased the time needed for manual annotation compared to using assisted free-hand in MATLAB alone. Some of the results are shown in Figure 5.7.

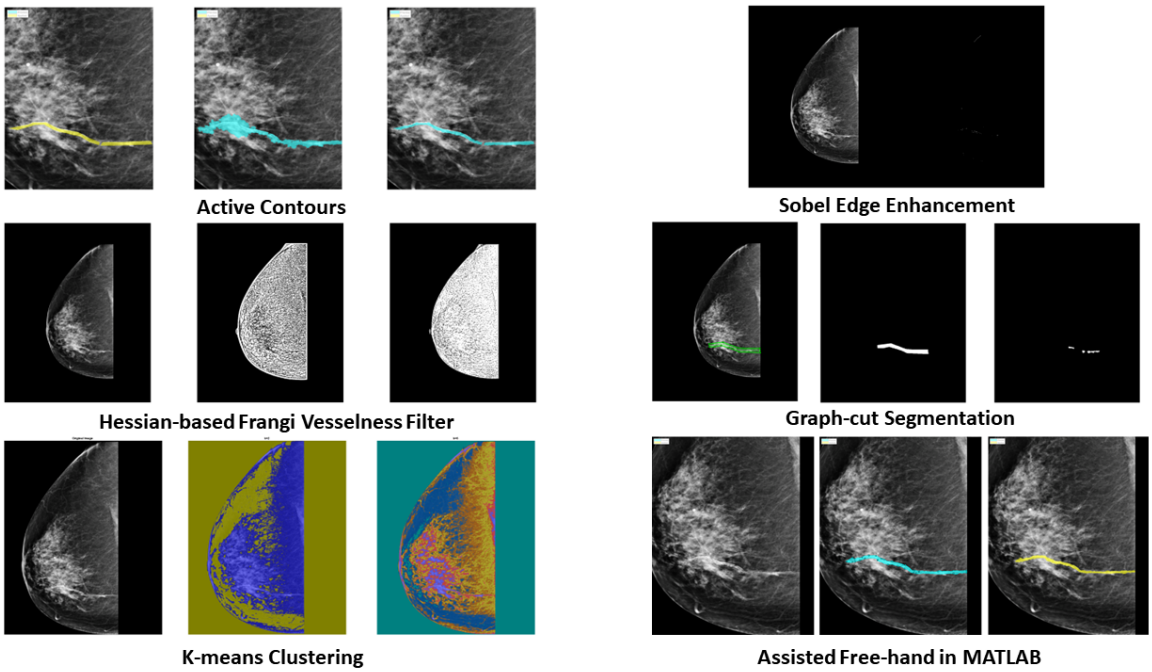


Figure 5.7: Image processing techniques investigated to ease the manual segmentation task.

Using MATLAB's Image Labeler app (MATLAB, 2022) under the guidance of a consultant radiologist with 8 years breast imaging experience, 220 BAC-positive images were manually segmented by the author using a Wacom Intuous S Tablet and Pen. BAC and background pixel labels were added, aided by the flood fill and assisted freehand tools. The remaining breast tissue was subsequently assigned programmatically, giving three classes in total: (i) BAC, (ii) background (area with no tissue) and (iii) breast. Manual segmentation using Image Labeler is shown in Figure 5.8.

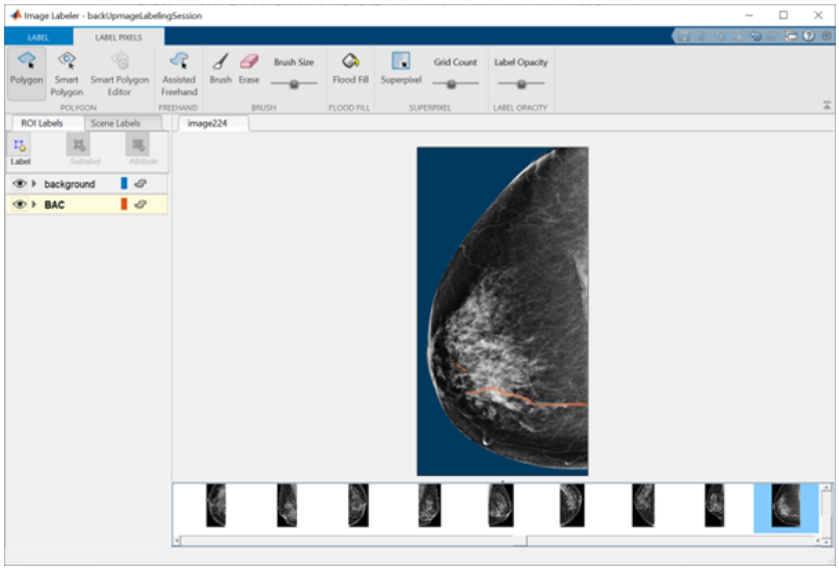
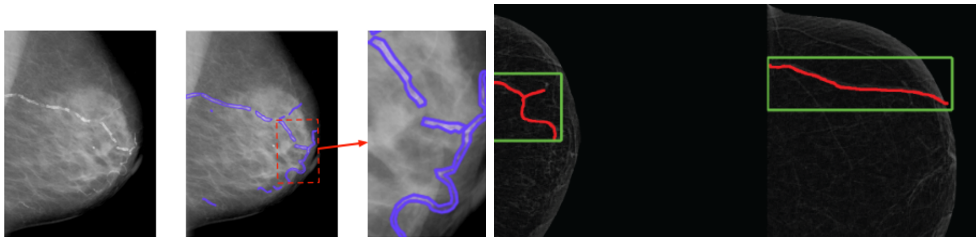


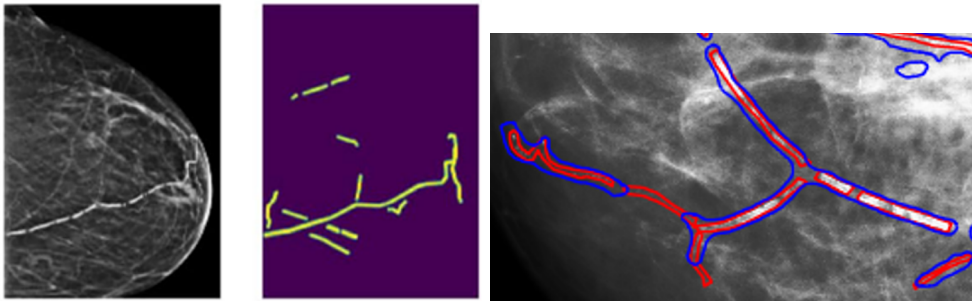
Figure 5.8: Using MATLAB's Image Labeller app (MATLAB, 2022) to segment BAC.

Figure 5.9 shows hand-crafted ground truth segmented BAC from four papers. Most outline the vessel completely although K. Wang et al. (2019) note that it is practically difficult to achieve a pixel-level perfection. Figure 5.10 shows examples of the author's ground truth incorporating background, breast and BAC for comparison.



(a) Ghamdi et al. (2020)

(b) K. Wang et al. (2019)



(c) Guo et al. (2021)

(d) J. Wang et al. (2017). In red.

Figure 5.9: Ground truth segmented BAC from four papers.

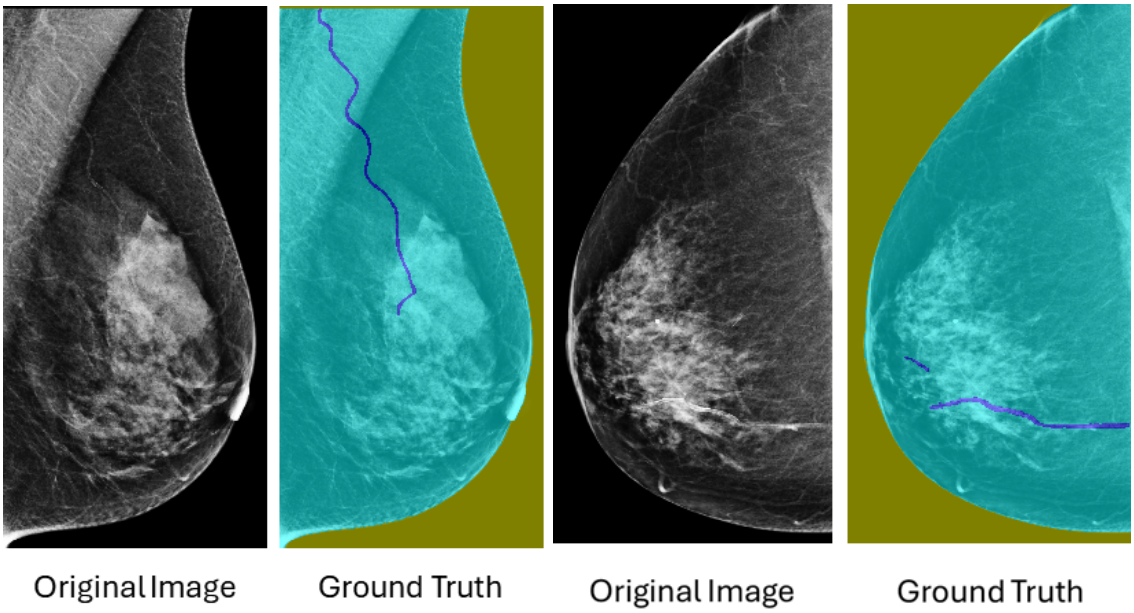


Figure 5.10: Examples of hand-crafted BAC ground truth segmented manually by author.

5.4 Image Pre-processing

5.4.1 12-bit DICOM to 16-bit PNG Conversion

In order to reduce computational overhead, the 12-bit DICOM images were converted to 16-bit PNG (Portable Network Graphics) files using an open source tool, `pydicom` (Mason et al., 2020). This allowed the images to retain their spatial resolution (number of pixels) while reducing their memory footprint (Wu et al., 2019). This process removed the header metadata and, on average, file sizes were reduced by a third.

5.4.2 Image Cropping and Padding

Images in the dataset came in four different matrix sizes and were generated on three mammographic units:

(i) 2560 x 3328 (63.5%, Lorad Selenia and Selenia Dimensions (Hologic, Inc)),

(ii) 3328 x 4096 (34.54%, Lorad Selenia and Selenia Dimensions (Hologic, Inc)),

(iii) 1914 x 2294 (1.04%, Senographe Essential (GE)),

(iv) 2394 x 3062 (0.95%, Senographe Essential (GE)).

Images were cropped to the breast using a Python function open-sourced by a recent study (Wu et al., 2019) that used deep learning to classify breast lesions. This resulted in images ranging from 567 x 1801 to 2953 x 4096 in size. While satisfactory for our segmentation network which extracted patches from images of differing sizes, our BAC classification and object detection network required all input images to be the same size. For this requirement, padding was added using the Python Pillow library (Clark, 2020) resulting in images of 3372 x 4140 in size that retained the spatial resolution of the area of interest. Bounding boxes were also amended accordingly for each image in both the above scenarios.

5.4.3 Data Augmentation

It has been noted that in image classification the input-output function should be insensitive to variations in position, orientation or illumination allowing the possibility of data augmentation to increase the dataset size (LeCun, Bengio, & Hinton, 2015). Goodfellow et al. (2016) also state that an image classification network needs around 5000 training images in order to be sufficiently robust.

Chougrad et al. (2018) note that data augmentation can increase the training dataset by a factor of ten, reducing the possibility of overfitting which occurs when the gap between the training error and test error is too large. They employed a range of techniques in their own study including transformations such as width and height shifts with a fraction of 0.25 from the total width or height of the image, a random rotation range of zero to 40 degrees, a shear range of 0.5 and a zoom range between 0.5 and 1.5. They also flipped the images horizontally and applied a “fill mode” strategy for filling in newly created pixels, which can appear after a rotation or a width/height shift. Jain and Levy (2017) filled these latter areas with the mean pixel value of the training set.

As shown in Figure 5.11, several transforms were applied to the OPTIMAM dataset images including horizontal and vertical flipping and a combination of both. All original and resulting images were then rotated by 5 degrees and cropped to the pre-rotation image size. This increased the train:validation:test dataset from 652:82:82 (BAC and non-BAC images)

to 5216:82:82 images. The total number of images pre- and post-augmentation per class are shown in Table 5.1. Bounding boxes were also amended accordingly for each BAC-positive augmented image.

Data augmentation was primarily used to increase the data size in order to reduce overfitting rather than addressing any class imbalance. For the classification model, the aim was to have equal amounts of each class in the training, validation and test sets. Pixel-level class imbalance was addressed in the segmentation model using inverse frequency as described in Section 8.2.

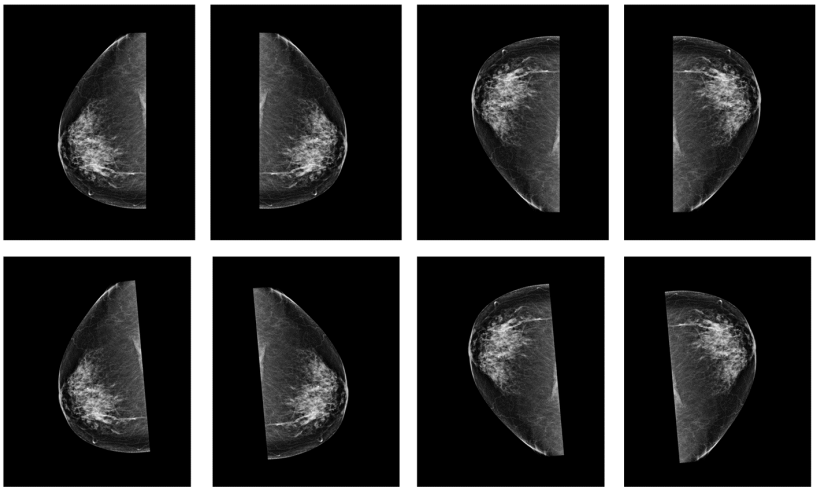


Figure 5.11: Dataset images with vertical and horizontal flipping and 5 degree rotation.

Table 5.1: Total number of images pre- and post-augmentation per class.

Class	**BAC**	**NON_BAC**
Pre-augmentation	408	408
Post-augmentation	3264	3264

5.4.4 CLAHE (Contrast-limited Adaptive Histogram Equalisation)

Some papers like Zeiser et al. (2020) have used CLAHE (contrast-limited adaptive histogram equalisation) in the pre-processing stage of mammography deep learning studies in order to remove image noise and increase contrast. CLAHE works by dividing an image into contextual regions of equal size, called *tiles*, and applying histogram matching for each tile individually. Neighbouring tiles are then combined to eliminate artificially induced boundaries. A

clip limit specifies a contrast enhancement cut-off that avoids amplifying image noise (Gonzalez et al., 2020). Cunha Carneiro, Lemos Debs, Oliveira Andrade, and Patrocinio (2019) noted better contrast between fibroglandular tissue and adjacent structures on dense breast mammograms using a clip limit of 0.01, a patch size of 15 x 15 and a uniform distribution. These parameters were implemented on our dataset to create a second set of images using the MATLAB `adapthisteq` function which implements CLAHE. Figure 5.12 shows the effect of various CLAHE clip limits and tile combinations on the original image on the left.

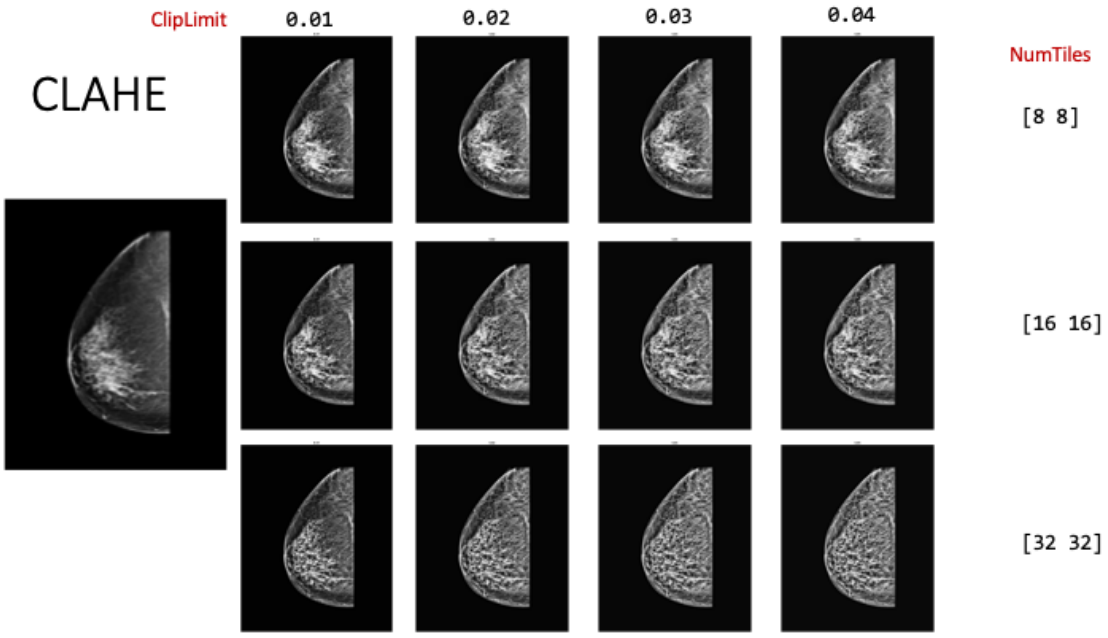


Figure 5.12: CLAHE: Effect of various clip limits and tile numbers.

5.5 Development Environment

MATLAB (MATLAB, 2022) was used as the main development environment as it provides a range of deep learning models, visualization capabilities, the ability to import models from other tools and built-in annotation tools. Other development environments could, of course, also have been used.

The hardware used consisted of a desktop PC running Windows 10 Pro with 32GB RAM, an Intel Core i7-8700 3.2 GHz processor and an 11GB Nvidia GeForce RTX 2080 Ti GPU.

6 | BAC Classification Model

Classification involves asking a computer program to specify which of k categories some input belongs to (Goodfellow et al., 2016). It does this by producing a function $f : \mathbb{R}^n \rightarrow \{1, \ldots, k\}$ and when $y = f(\mathbf{x})$, an input described by vector **x** is assigned to a category identified by numeric code y. Szeliski (2022) states that visual category recognition or image classification is an extremely challenging problem but notes that deeper networks and better training algorithms post-AlexNet have increased recognition accuracy dramatically. In fact, it is the main application for which deep networks were originally developed (LeCun et al., 1989).

6.1 Transfer Learning

Bengio et al. (2012) showed that the transfer of knowledge in networks could be achieved by first training a neural network on a domain for which there is a large amount of data, and then retraining that network on a related but different domain via fine-tuning its weights by either training the whole initialized network or by "freezing" some of the pre-trained weights. Applied to radiology, even though the algorithm is being trained largely on non-medical images, these networks perform just as well as those trained with purely medical images (McBee et al., 2018).

MATLAB has thirteen pre-trained networks of varying depths trained on ImageNet (Deng et al., 2009). Three pre-trained networks, namely AlexNet (Krishevsk, et al, 2012), SqueezeNet (Iandola et al., 2016) and GoogLeNet (Szegedy et al., 2015), were chosen for the BAC classification task. AlexNet was the breakthrough network for the ImageNet competition and is therefore often used as a benchmark network. GoogleNet reduced computational complexity

post-AlexNet by using inception modules which comprised of multiple parallel convolutional layers with different filter sizes. Squeezenet improved performance further by investigating what combination of 1x1 and 3x3 filters affected model size and accuracy.

Initially all three networks were trained without the augmented images resulting in poor results (approx. 50% test accuracy) which did not improve significantly when the additional images were added. Input sizes of 227x227 (AlexNet, SqueezeNet) and 224x224 (GoogLeNet) meant that the dataset images were resized to less than 1% of their size and transfer learning using these networks may have been more appropriate for image patches rather than larger images like those in the OPTIMAM dataset.

One recent large study investigating the use of deep learning in breast cancer screening for lesion detection developed a custom ResNet-based network specifically targeted at high resolution mammography images (Wu et al., 2019). Residual networks take a standard deep CNN and add shortcut ("skip") connections that bypass a few convolutional layers at a time. The shortcut connections create residual blocks where the output of the convolutional layers is added to the block's input tensor. They also solve the issue of vanishing or exploding back-propagated gradients when using deeper networks (Glorot & Bengio, 2010). Computational overhead is kept low by the connections which also provide a rich combination of features (Chougrad et al., 2018).

The network used by Wu et al. (2019) is based on a ResNet-18 (He et al., 2016) topology with an extra ResNet block (5 instead of four) which gives a balance of width and depth allowing input image sizes of 2677x1942 for CC views and 2974x1748 for MLO views. In this study, we build upon the proposal by Wu et al. (2019B), adapting it for use to classify BAC. Their publicly available trained network, including initial weights, was first exported from TensorFlow (Abadi et al., 2015) to an ONNX (Open Neural Network Exchange - https://github.com/onnx/onnx) model and imported into MATLAB. It was customised with two additional dropout layers to provide the 87-layer, 2-class BAC classification network shown in Figure 6.1. The architecture we used included the following key layers:

1. The input layer which accepts images of size 3372x4140.

2. Layers 2 to 80 are similar to those in the original model, comprising 5 residual blocks.

3. Two dropout layers were added to reduce overfitting. Dropout is an averaging technique based on stochastic sampling of neural networks. Neurons are randomly removed during training resulting in slightly different networks for each batch of training data. Weights of the trained network are then optimised to these multiple variations of the network. Dropout probability was set at 50%.

A layer-by-layer description of the network is shown in Appendix F.

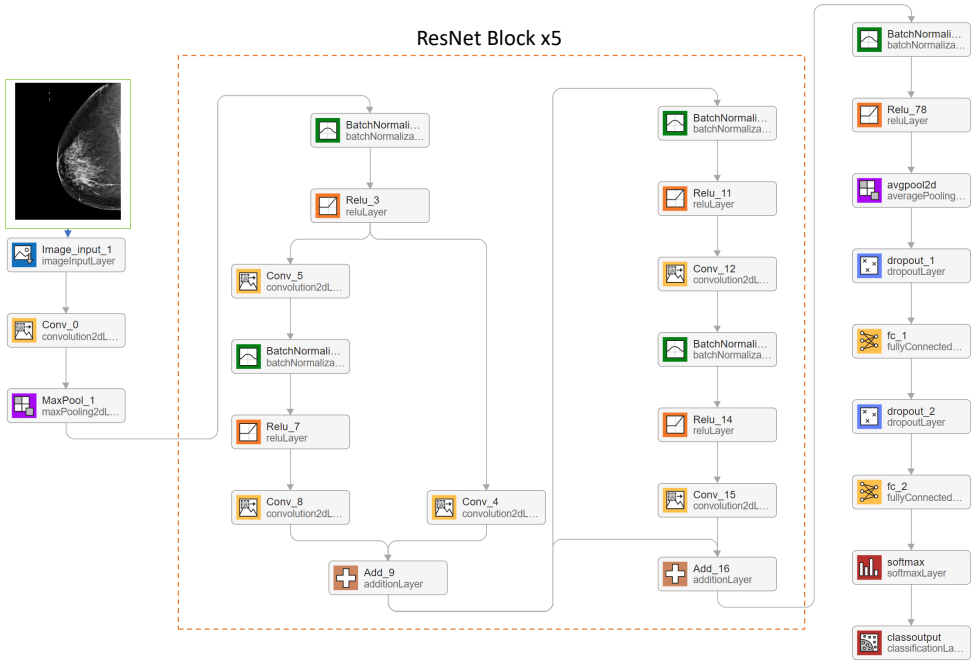


Figure 6.1: ResNet-22 BAC classification network.

6.2 Network Training

To evaluate the model, 10-fold cross validation was used. As shown in Table 6.1, for 10 iterations two folds were used as validation and test data and the rest were trained with augmented data for 20 epochs with a train:validate:test ratio of 5216:82:82 images (50% BAC/50% non-BAC). Data was split based on the case so that no two sets had images from the same patient. Training was undertaken using zerocenter, zscore, rescale-zero-one and rescale-symmetric normalization with initial weights seeded from Wu et al. (2019)'s study. Zerocenter normalization subtracts the mean while zscore subtracts the mean and divides

by the standard deviation. Rescale-symmetric normalization rescales the input to be in the range [-1,1] while rescale-zero-one rescales to the range [0,1].

Table 6.1: 10-fold cross validation training sequence.

Iteration	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Fold 1	**Test**	Train	Train	Train	Train	Train	Train	Train	Train	**Val.**
Fold 2	**Val.**	**Test**	Train	Train	Train	Train	Train	Train	Train	Train
Fold 3	Train	**Val.**	**Test**	Train	Train	Train	Train	Train	Train	Train
Fold 4	Train	Train	**Val.**	**Test**	Train	Train	Train	Train	Train	Train
Fold 5	Train	Train	Train	**Val.**	**Test**	Train	Train	Train	Train	Train
Fold 6	Train	Train	Train	Train	**Val.**	**Test**	Train	Train	Train	Train
Fold 7	Train	Train	Train	Train	Train	**Val.**	**Test**	Train	Train	Train
Fold 8	Train	Train	Train	Train	Train	Train	**Val.**	**Test**	Train	Train
Fold 9	Train	Train	Train	Train	Train	Train	Train	**Val.**	**Test**	Train
Fold 10	Train	Train	Train	Train	Train	Train	Train	Train	**Val.**	**Test**

6.2.1 Optimisation

Three optimisers were initially investigated - stochastic gradient descent with momentum, RMSProp and Adam. Szeliski (2022) notes that regular gradient descent is prone to stalling when the current solution reaches a “flat spot” and this can be addressed by using *momentum* where an exponentially decaying running average of gradients is accumulated and used to update the direction. This prevents updates jumping wildly back and forth. RMSProp keeps a history of the size of the gradient and uses this to scale the learning rate for each parameter (G. E. Hinton, 2012). Adam (Adaptive moments estimation) scales the learning rate for each parameter (like RMSProp) but also uses momentum to smooth out the updates (Kingma & Ba, 2017). Goodfellow et al. (2016) note that RMSProp lacks the correction factor of Adam and may have high bias early in training. They state that Adam is generally robust to the choice of hyperparameters although the learning rate may need to be changed from the suggested default.

After experimenting with several values (0.01, 0.001 and 0.0001), a learning rate of 0.0001 was used with an Adam optimiser. Batch size was limited to 4 as larger sizes led

to “out of memory” errors on the GPU as shown in Figure 6.2. This increased the training time. It took more than 37 hours to train one fold at full size (3372x4140). Images were subsequently scaled to 70% size (2360x2898) leading to a training time of 60 hours for the full ten folds.

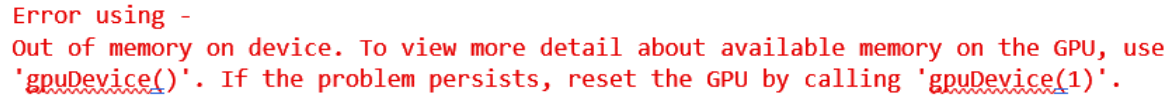


Figure 6.2: GPU out of memory error.

6.2.2 Metrics

The BAC classification model outputs “BAC” or “NON_BAC” based on the input image. Network performance was evaluated using the following metrics: test accuracy, precision, recall and F1 score. These are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = \frac{2 * Precision * Recall}{(Precision + Recall)}$$

where TP is the true positive, TN is the true negative, FP is the false positive and FN is the false negative.

Layer activations and feature maps further provided a visual representation of the model’s performance and are shown in Section 6.3.

6.3 Results

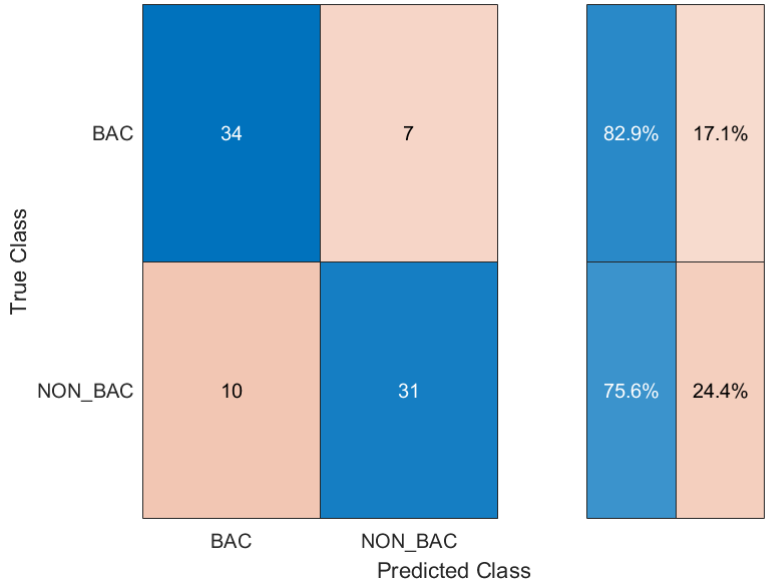


Figure 6.3: Rounded aggregated confusion matrix for 10-fold classification.

The results from 10-fold cross validation for BAC classification using our ResNet22 model with data augmentation are shown in Table 6.2. Using rescale-symmetric normalization resulted in the highest mean accuracy (0.80) and recall (0.79). Zscore normalization provided the highest mean precision (0.84) while both types attained identical mean F1 scores (0.81).

Initially, 10-fold cross validation was performed without images being divided into training, test and validation sets based on the patient case and also without the drop-out layers. Mean accuracy was initially found to be 81%. When the two dropout layers were added, this increased to 82.25%. When images were subsequently split by case, the results shown in Table 6.2 were achieved. The rounded aggregated confusion matrix for 10-fold classification indicating class-specific accuracy for both classes is shown in Figure 6.3 above.

Table 6.2: 10-Fold Cross Validation BAC Classification Results by Normalisation Type.

Normalisation Type	Mean Accuracy	Mean Precision	Mean Recall	Mean F1 Score
rescale-symmetric	**0.80**	0.83	**0.79**	**0.81**
zscore	0.79	**0.84**	0.78	**0.81**
rescale-zero-one	0.77	0.81	0.77	0.79
zerocenter	0.76	0.79	0.76	0.77

Table 6.3 below compares the performance of CLAHE and non-CLAHE pre-processed images with CLAHE-applied training sets achieving lower scores (averaging approximately 15.3% lower) across all metrics.

Table 6.3: BAC Classification Results Comparison of CLAHE to non-CLAHE.

Method	Mean Accuracy	Mean Precision	Mean Recall	Mean F1 Score
Non-CLAHE	0.80	0.83	0.79	0.81
CLAHE	0.63	0.66	0.64	0.65

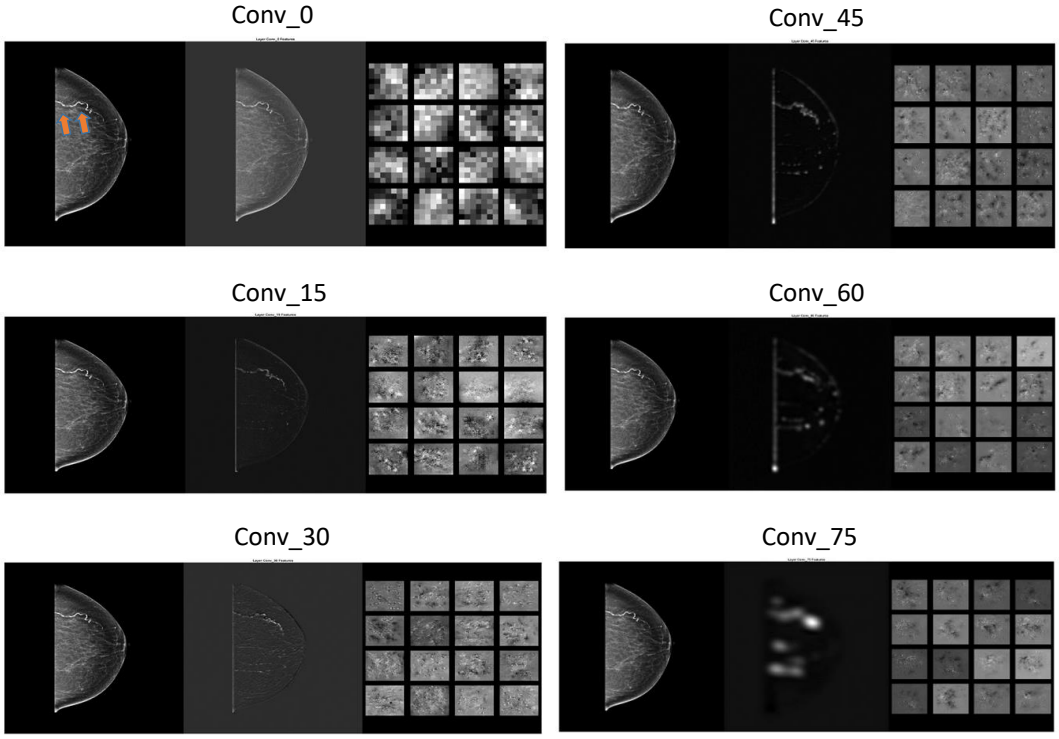


Figure 6.4: Layer Activations and their Associated Feature Maps. Each section contains the original image on the left. The middle image is the strongest activation for that layer applied to the image. The feature maps for that layer are shown on the right.

Layer activations with their associated feature maps were recorded at 6 layers (Conv_0, Conv_15, Conv_30, Conv_45, Conv_60 and Conv_75). A single image's progression through

the layers is shown in Figure 6.4. No feature maps examined provided clear representations of any linear or curvi-linear structures although the model clearly focuses on regions containing BAC.

6.4 Discussion

Only one other paper (Mobini et al., 2023) addresses image-level BAC classification for the presence or absence of BAC making comparison difficult for this method. The authors claimed 97% accuracy although only 10% of their test set were BAC-positive images. Answering the classification aspect of our main research question, i.e. how well do our models perform?, is difficult, therefore. Despite this, with our ResNet22 network achieving a test accuracy of 80%, we believe that this approach has the potential of being used as a simple flag for BAC to alert humans and/or other algorithms of its presence on mammograms.

Surprisingly, CLAHE had a detrimental effect on the model's performance. Multiple clip limits, tile numbers and distributions were investigated but the test accuracy never reached 70%. We would not advocate using CLAHE in any BAC classification deep learning studies.

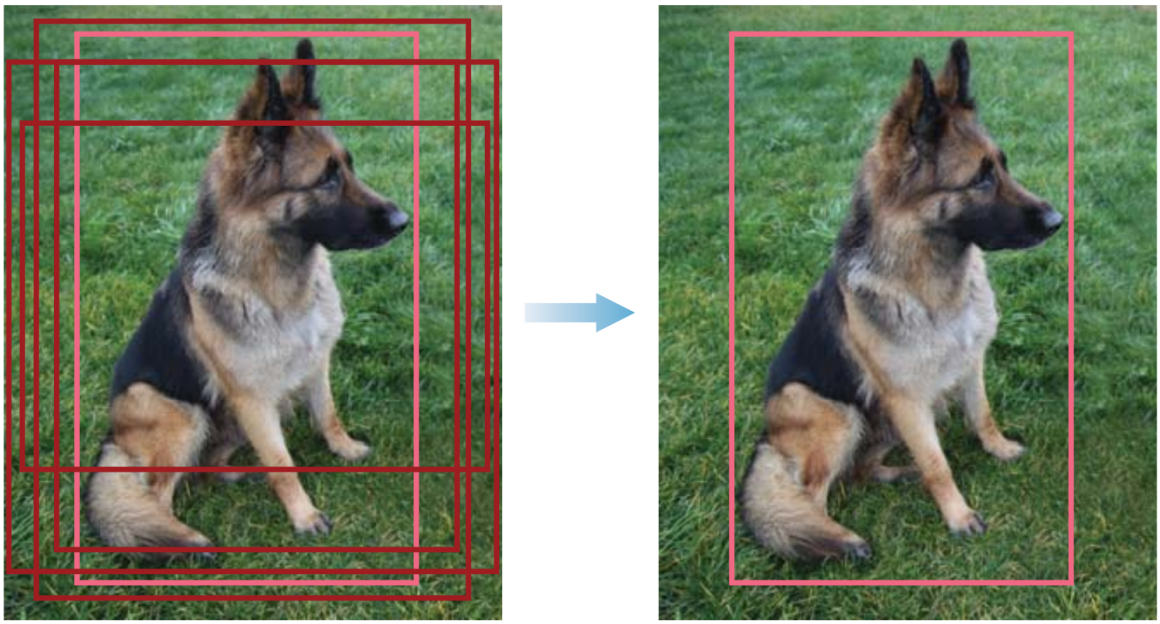
6.5 Summary

We have described a model to classify FFDM images for the presence or absence of BAC using transfer learning built on a custom ResNet-22 network with weights originally used to classify breast cancer lesions. Despite the application of CLAHE having a negative effect on performance, our network shows promise as a simple flag for image-level BAC classification.

The next chapter describes using object detection techniques in order to extend the above approach and predict the location of BAC on those images.

7 | BAC Object Detection Model

As we have seen in the previous chapter, the goal of image classification is to predict the type or class of an object in an image whereas *object detection* involves predicting the location of objects in an image using bounding boxes and the classes of the located objects (Elgendy, 2020). Object detectors operate by first proposing a number of plausible rectangular regions then classifiying each detection while also producing a confidence score. Regions undergo a non-maximum suppression stage (NMS) where a single entity is selected from many overlapping entities (Szeliski, 2022). The application of NMS is shown in Figure 7.1.



Predictions before NMS

After applying non-maximum suppression

Figure 7.1: NMS identifes the box with the maximum prediction probability and discards the rest. From Elgendy (2020).

7.1 Multi-stage Detectors

Girshick et al. (2014) introduced region-based convolutional neural networks (R-CNNs) which consisted of four main components: extracting region proposals, a feature extraction module, a classification module and a localisation module. R-CNNs extracted about 2,000 region proposals which were then rescaled to a 224 square image and passed through an AlexNet or VGG neural network with a support vector machine (SVM) final classifier (Szeliski, 2022).

Fast R-CNN (Girshick, 2015) interchanged the CNN and region extraction stages of R-CNN and replaced the SVM with some fully connected layers leading to much faster training and test times (2 seconds vs 50 seconds) as well as dramatically better accuracy.

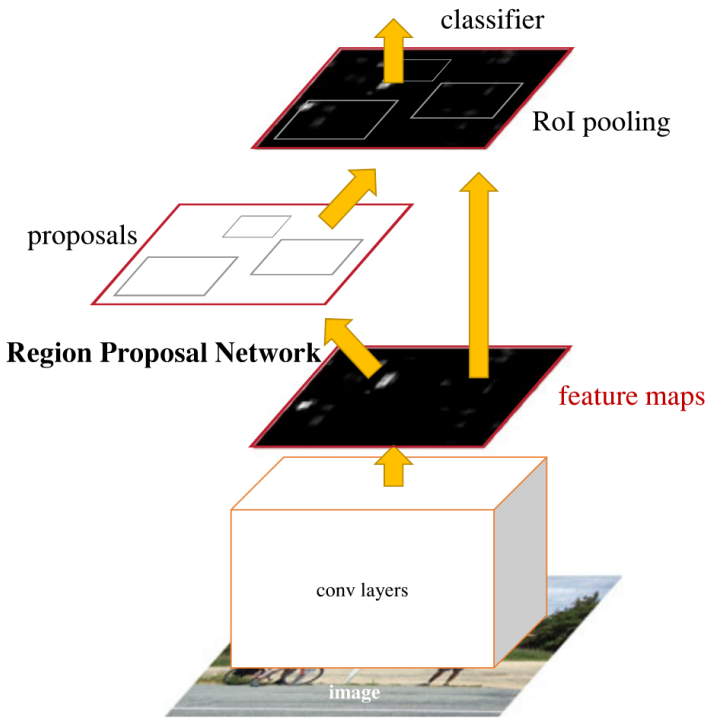


Figure 7.2: Faster R-CNN replaces the selective search method of Fast R-CNN with a region proposal network making the algorithm ten times faster. From Ren et al. (2017).

Faster R-CNN (Ren et al., 2017), shown in Figure 7.2 above, builds on R-CNN and Fast R-CNN by utilising a CNN to extract features and obtain regions of interest using a region proposal network (RPN). An RPN is a fully convolutional network that simultaneously predicts anchor boxes of varying shapes and sizes and objectness scores at each position using an instance of a Fast R-CNN head. Full-image convolutional features are shared with the

detection network, thus enabling nearly cost-free region proposals. The detections are then ranked and merged using non-maximum suppression. Faster R-CNN was found to be ten times faster in testing images than Fast R-CNN (Elgendy, 2020).

7.2 Single-stage Detectors

Another type of object detection network is a *single-stage network* which uses a single neural network to output detections at numerous locations (Szeliski, 2022). The YOLO ("You Only Look Once") family of object detectors is a series of end-to-end deep learning models designed for fast object detection (Elgendy, 2020). The first version, YOLOv1, unified the two main components of a detector, i.e. the object detector and the class predictor (Redmon, Divvala, Girshick, & Farhadi, 2016). Subsequent iterations improved test times and accuracy through larger models (Redmon & Farhadi, 2017, 2018).

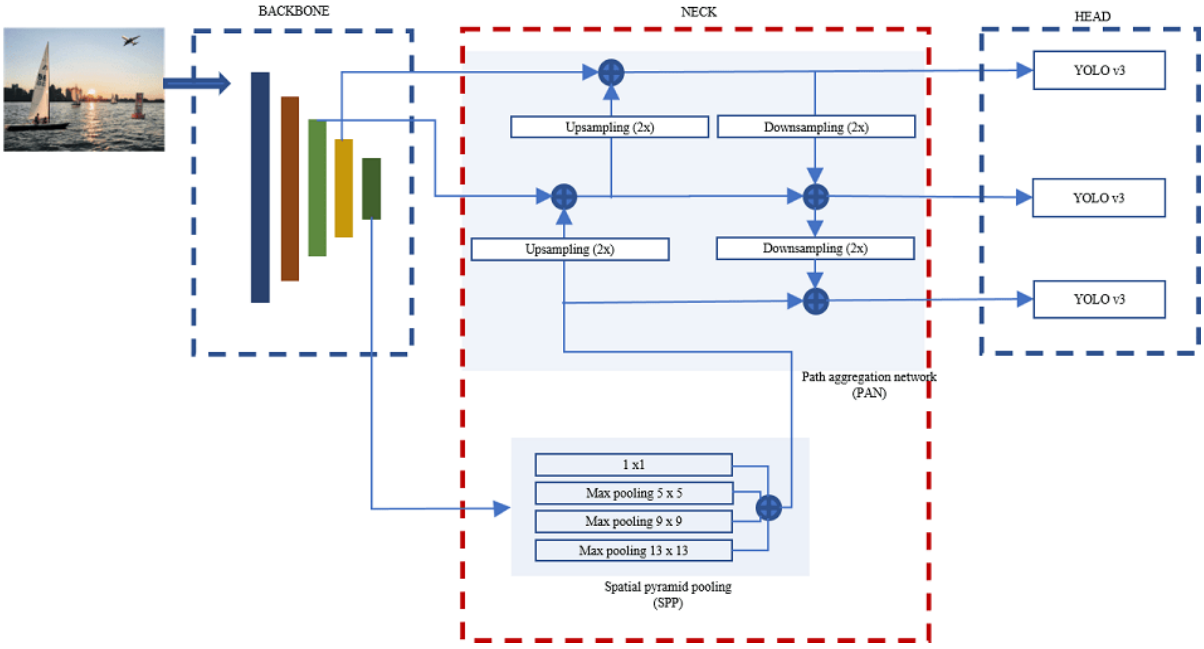


Figure 7.3: YOLOv4 architecture. From Mathworks (2023b).

YOLOv4 (Bochkovskiy et al., 2020), shown in Figure 7.3 above, comprises three parts: a backbone, a neck and a head. The backbone acts as a feature extraction network that creates feature maps from input images. The neck is composed of a spatial pyramid pooling (SPP) module and a path aggregation network (PAN). It performs top-down feature enhance-

ment and concatenates feature maps output by the backbone and passes these as inputs to the head. The latter processes the concatenated features, predicting boundary boxes, objectness scores and classification scores. This step uses one-stage object detectors, in this case YOLOv3 (Redmon & Farhadi, 2018), as detection heads.

Only two YOLO models have been implemented in MATLAB since YOLOv4, namely YOLOX (Z. Ge, Liu, Wang, Li, & Sun, 2021) and YOLOv8. Unfortunately, they were not available at the time of training. YOLOX forgoes using anchors to predict bounding box dimensions, instead the network divides the input image into a grid of three different scales, and uses the grid points as the top-left offsets of the bounding boxes. YOLOv8, which has no paper published, was released by the company Ultralytics in January 2023. It uses the same backbone as YOLOv4 and is also anchor-free and avails of mosaic augmentation during training (Terven & Cordova-Esparza, 2023). Improved accuracy and better speed and efficiency have led it to be considered the new state-of-the-art in object detection (Reis, Kupec, Hong, & Daoudi, 2023).

7.3 Network Training

Both multi-stage (Faster R-CNN) and single-stage detectors were used in training the object detection models. Multi-stage detectors offer greater accuracy while single-stage models provide greater speed (Szeliski, 2022).

7.3.1 Faster R-CNN

The 96-layer Faster R-CNN with a ResNet-22 feature extraction network is shown in Figure 7.4. A layer-by-layer description of the network is provided in Appendix G. Faster R-CNN adds a region proposal network to generate region proposals directly in the network instead of using an external algorithm, i.e. it is 'faster' within the network (Ren et al., 2017).

As mentioned in Section 5.2, we collected and prepared 408 BAC-postive images. For the object detection experiments, these were divided into training and test sets with a ratio of 80:20, giving 326:82 images respectively. The training images were augmented using the

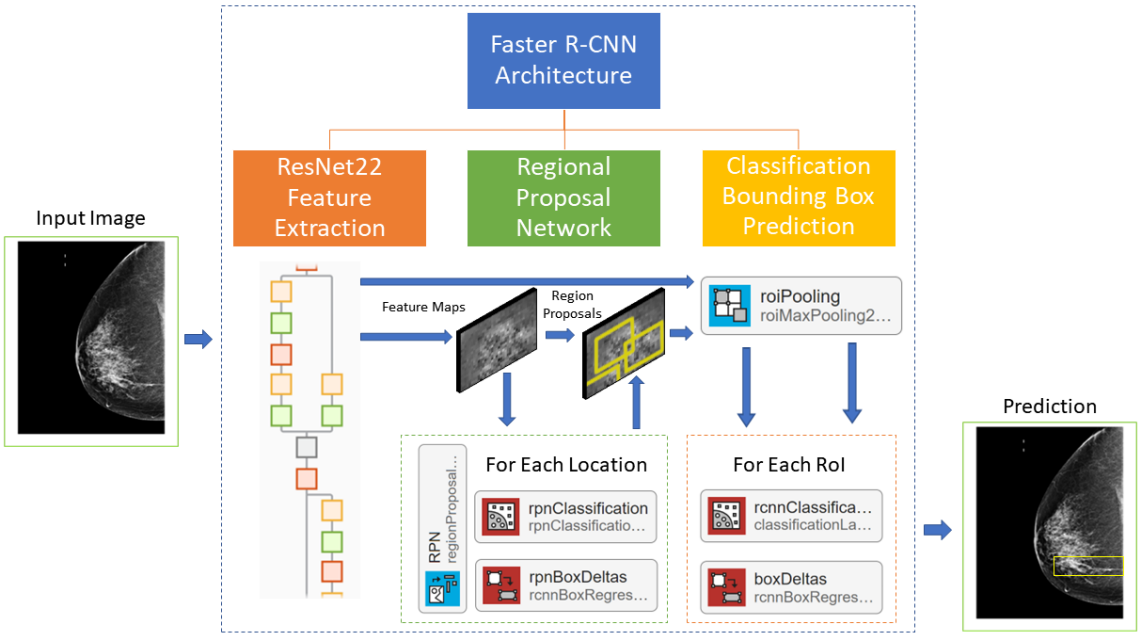


Figure 7.4: Faster R-CNN network with modified ResNet-22 feature extraction network.

techniques described in Section 5.4.3 bringing the number of training images up to 2604 (4 augmented images were discarded due to invalid bounding box positions). Data was split based on the case so that both sets did not have images from the same patient. Validation data were not used as, at the time of training, the MATLAB function to train the network, `trainFasterRCNNObjectDetector`, did not support it. The network was assembled using the modified ResNet-22 network used for BAC classification as the feature extraction with layer 65, "ReLu_63", being used as the feature layer.

Another MATLAB method, `estimateAnchorBoxes`, was used to predict anchor boxes based on the size of bounding boxes in the training data. This was done after plotting the number of anchors against mean IoU on the training data as shown in Figure 7.5. This showed significant increases in mean IoU between 6 and 9 and 9 and 15 boxes respectively, so anchor boxes were estimated for these three amounts. Estimated box sizes ranged from 148x132 to 1599x1014. More boxes increase the mean IoU but are accompanied by a greater computational overhead and longer training times.

The network was trained for 20 epochs with images at 70% of the original size, i.e. 2898x2360. Data was pre-processed to resize images and rescale bounding boxes to the target size. After experimenting with several values (0.01, 0.001, 0.0001), a learning rate of

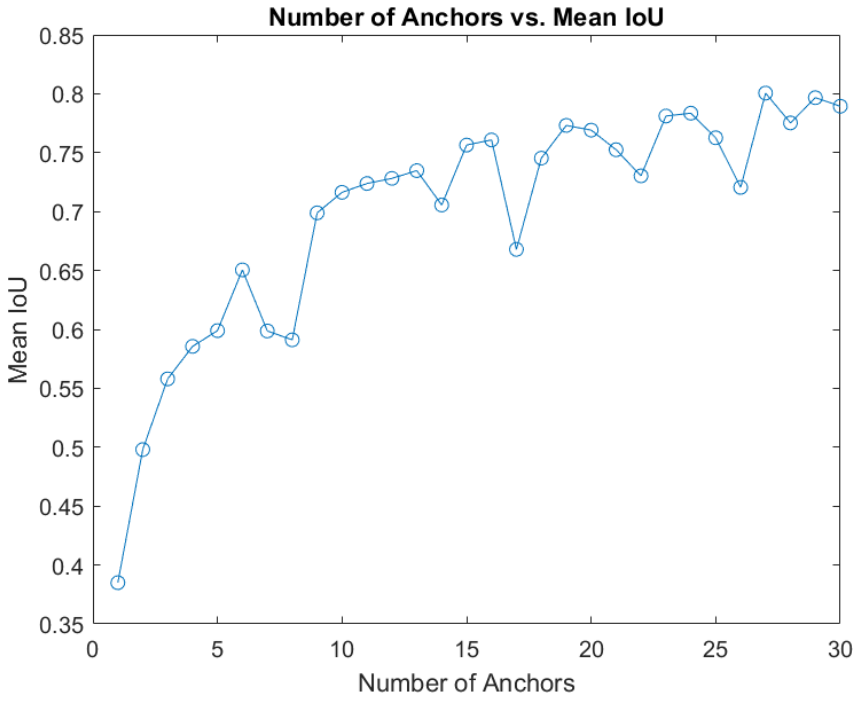


Figure 7.5: Number of anchor boxes versus mean IoU in the BAC-positive dataset.

0.001 was used with stochastic gradient descent optimisation.

We were unable to avail of parallel processing options in MATLAB to simultaneously train the region proposal and region classification subnetworks as this resulted in "Out of Memory" errors like those previously shown in Figure 6.2. These errors were also minimised by setting the mini-batch size to 4. Each subnetwork was then trained sequentially in four steps.

7.3.2 YOLOv4

Only one previous paper, K. Wang et al. (2019), has used object detection techniques to detect BAC. They found a mean IoU (Intersection over Union) of zero when using a YOLOv2 (Redmon & Farhadi, 2017) network. This project aimed to improve on that by using an updated YOLOv4 (Bochkovskiy et al., 2020) network utilising our customised ResNet-22 model as the feature extraction backbone. The network is shown in Figure 7.6 with a layer-by-layer description available in Appendix H.

As before, BAC-positive images were divided into training and test sets with an additional validation set. The training:validation:test ratio was 80:10:10. With augmented training data this led to 2604 training images and 41 unaugmented validation and test images each. Data

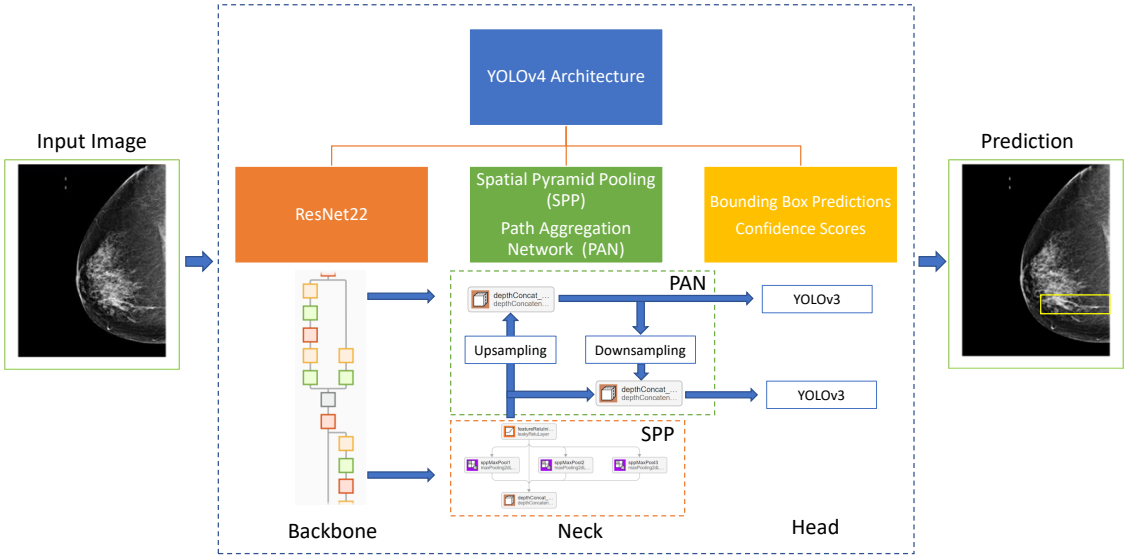


Figure 7.6: YOLOv4 BAC object detection network comprising a backbone (ResNet22 feature extraction network), a neck (Spatial pyramid pooling and a path aggregation network) and a head (Two YOLOv3 detectors).

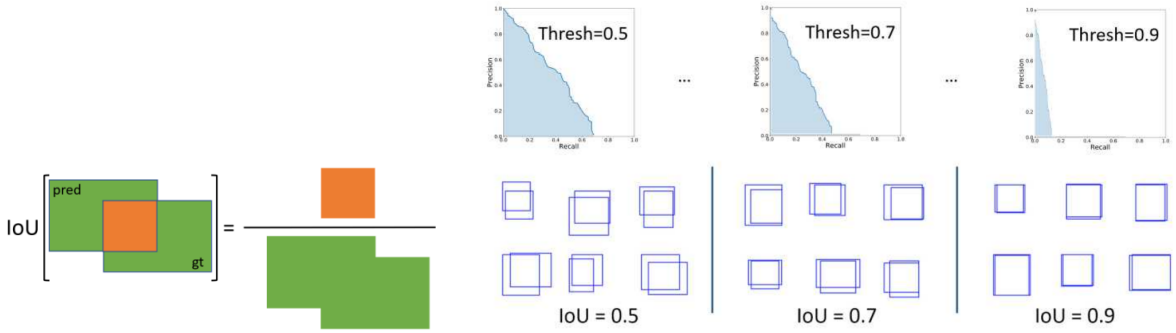
was split based on the case so that no two sets had images from the same patient. The network was assembled using the modified ResNet-22 network used for BAC classification as the feature extraction module with layers 46 and 65, “ReLu_44” and “ReLu_63” respectively, being used as the feature extraction layers.

The 101-layer network was trained for 70 epochs with BAC-positive images at 70% of the original size (2898x2360). Data was pre-processed to resize images and rescale bounding boxes to the target size. The resulting images, bounding boxes and labels were also validated. Images were checked that they were non-empty and had 2 dimensions. Bounding boxes required positive integer values stored in a $M \times 4$ matrix and labels had to be categorical and also non-empty.

After experimenting with several values (0.01, 0.001, 0.0001), a learning rate of 0.001 was used with an Adam optimiser. Batch size was again limited to 4 to curtail “Out of Memory” errors on the GPU. Six anchor boxes were used, the sizes of which were determined by the MATLAB function `estimateAnchorBoxes` and were based on the training data. Training YOLOv4 with 9 or 15 anchor boxes resulted in “Maximum variable size allowed on the device is exceeded” errors therefore training was limited to 6.

7.3.3 Metrics

Elgendy (2020) notes that the main tasks in object detection are to predict the coordinates of bounding boxes around all the objects of interest and to correctly label the objects. The most common metric is *intersection over union* (IoU), previously described in Section 5.2.1 and shown schematically in Figure 7.7a.



(a) (b)

Figure 7.7: (a) Schematic formula of IoU (b) Average precision averaged over several IoU thresholds (from less to more accurate). From Szeliski (2022). ©2020 Ross Girshick.

In order to evaluate an object detector, we need to traverse all detections from most confident to least and classify them as true or false positive (Szeliski, 2022). The former reflects a correct label and sufficiently high IoU whereas the latter may indicate an incorrect label or a ground truth object that has already been matched. Precision and recall can be determined at each IoU threshold level giving us a precision-recall curve like those in Figure 7.7b. *Average precision* (AP) is the area under this curve and this provides a single number that incorporates the ability of the detector to make correct classifications (precision) and the ability of the detector to find all relevant objects (recall). The ideal precision is 1 at all recall levels (Mathworks, 2023c).

Average precision was used to evaluate how successful our models were at detecting BAC.

7.4 Results

7.4.1 Faster R-CNN

Our ResNet-22 BAC classification network was used as a feature extraction backbone to a Faster R-CNN network trained with augmented data for BAC object detection. It was tested on 82 images. The average precision results for Faster R-CNN are shown in Table 7.1. At an IoU threshold of 0.5, the maximum AP is 0.0579. This is quite low and was obtained by the model being trained with 9 anchor boxes. Setting the threshold at 0.1 increases the AP marginally.

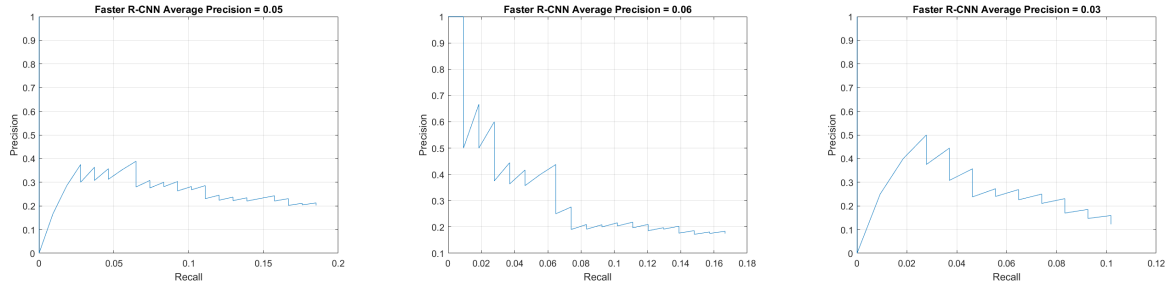
The precision-recall curves for a threshold of 0.5 are shown in Figure 7.8. At this threshold and with 9 anchor boxes, 17% (n=14) of the BAC-positive test set images were found to have no BAC present. Despite this, for some images the models did well in identifying BAC positive areas as shown in Figure 7.9.

Table 7.1: Faster R-CNN Average Precision Results.

No. of Anchor Boxes	IoU Threshold	Average Precision
6	0.5	0.0463
9	0.5	**0.0579**
15	0.5	0.0307
6	0.1	0.0516
9	0.1	0.0584
15	0.1	0.0319

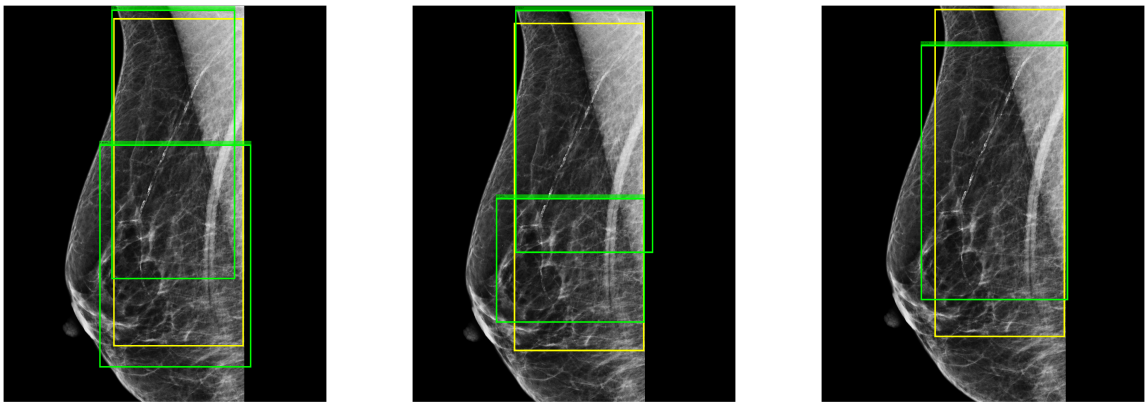
7.4.2 YOLOv4

Our ResNet-22 BAC-classification network was again used as a feature extraction backbone to a YOLOv4 network trained with augmented data for BAC object detection. It was tested on 41 images. The results for average precision for two IoU thresholds are shown in Table 7.2. An average precision of 0.02 was found for a threshold of 0.5 with a slight increase to 0.03 when it was lowered to 0.1.



(a) 6 anchor boxes. (b) 9 anchor boxes. (c) 15 anchor boxes.

Figure 7.8: Precision-recall curves for BAC object detection using Faster R-CNN with several anchor boxes at an IoU threshold of 0.5.



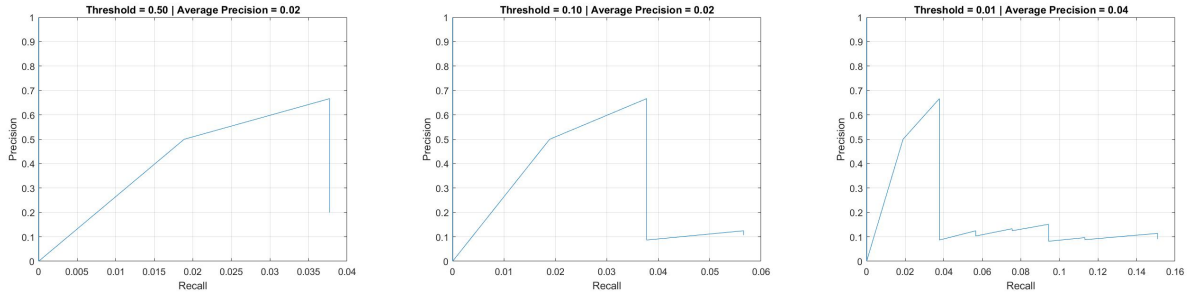
(a) 6 anchor boxes. (b) 9 anchor boxes. (c) 15 anchor boxes.

Figure 7.9: BAC-positive locations predicted by Faster R-CNN models trained with 6, 9 and 15 anchor boxes respectively. The yellow boxes are ground truth and the predictions are green.

Table 7.2: YOLOv4 Average Precision Results.

No. of Anchor Boxes	**IoU Threshold**	**Average Precision**
6	0.5	**0.0189**
6	0.1	0.027

Figure 7.10a overleaf shows the precision recall curve for BAC object detection using YOLOv4 at an IoU threshold of 0.5. An average precision of 0.02 is even lower than that achieved by the worst performing Faster R-CNN model. Table 7.3 shows the percentage of images reaching detection confidence scores for BAC object detection, i.e. the probability that the object in the box is BAC. 9.7% of test images (n = 4) attained confidence scores of greater than 0.9 while 19.5% (n = 8) reached 0.5.



(a) Threshold = 0.5, Average Precision = 0.02

(b) Threshold=0.1,Average Precision=0.02

(c) Threshold = 0.01, Average Precision = 0.04

Figure 7.10: Average Precision for BAC Object Detection at several thresholds.

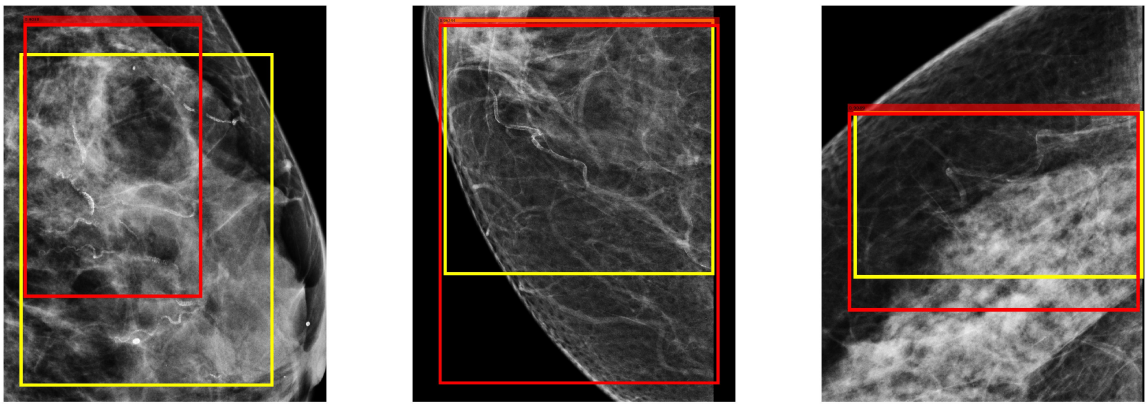
Table 7.3: Percentage of Test Images Reaching Detection Confidence Scores.

Confidence Score	No. of Test Images Reaching Score	% of Test Images Reaching Score
0.9	4	9.7%
0.8	5	12.1%
0.7	6	14.6%
0.6	7	17.1%
0.5	8	19.5%
0.4	9	22.0%
0.3	13	31.7%
0.2	17	41.5%
0.1	19	46.3%

Figure 7.11 shows a sample comparison of ground truth and predicted objects for test images that attained a detection confidence score of greater than 0.9. The predicted boxes correspond well to ground truth BAC regions. A sample comparison of ground truth and predicted objects for test images that attained a detection confidence score of between 0.1 and 0.2 is shown in Figure 7.12. Despite these low scores, some predicted areas correspond visually satisfactorily to ground truth BAC regions.

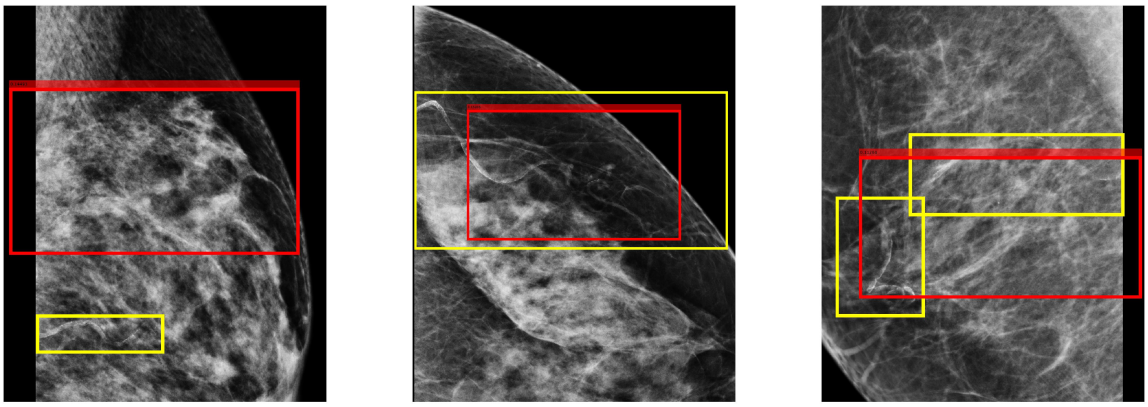
7.4.3 YOLOv4/Faster R-CNN Comparison

As we were unable to use validation data for the Faster R-CNN model, it was difficult to compare the latter to YOLOv4 as the test set had twice as many images. We evaluated the YOLOv4 test set, which makes up half of the Faster R-CNN test set, using two of the Faster R-CNN models and the results are shown in Table 7.4. Both Faster R-CNN models improved



(a) Confidence score: 0.9038 (b) Confidence score: 0.96244 (c) Confidence score: 0.9089

Figure 7.11: Comparison of ground truth and predicted objects for test images that attained a detection confidence score of greater than 0.9. Ground truth is in yellow with the predicted bounding box in red.



(a) Confidence score: 0.14493 (b) Confidence score: 0.15305 (c) Confidence score: 0.11261

Figure 7.12: Comparison of ground truth and predicted objects for test images that attained a detection confidence score between 0.1 and 0.2. Ground truth is in yellow with the predicted bounding box in red.

on the YOLOv4 trained model with the 6 anchor network more than doubling the AP.

Table 7.4: YOLOv4 vs Faster R-CNN Average Precision Results.

Model	No. of Anchor Boxes	IoU Threshold	Average Precision
YOLOv4	6	0.5	0.0189
Faster R-CNN	6	0.5	**0.0692**
Faster R-CNN	9	0.5	0.0270

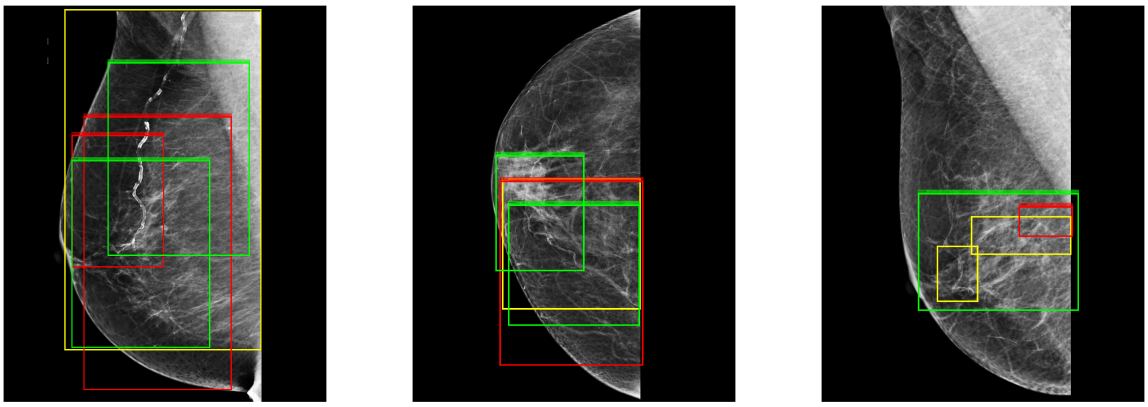


Figure 7.13: BAC-positive locations predicted by a Faster R-CNN and a YOLOv4 model trained with 6 anchor boxes respectively. The yellow boxes are ground truth and the predictions are green for Faster R-CNN and red for YOLOv4.

7.5 Discussion

Our Faster R-CNN object detection models performed poorly in identifying BAC on screening mammograms achieving a maximum average precision of 0.0579. Our YOLOv4 object detection network performed similarly poorly with an AP of just 0.0189. 83% of images reached the 0.5 IoU threshold using Faster R-CNN while only approximately 20% did so using YOLOv4. Despite this, the latter model obtained detection confidence scores of over 90% for four of the test images, three of which are shown in Figure 7.11. Furthermore, some bounding boxes with low confidence detection scores appear in or near locations where ground truth bounding boxes are situated as shown in Figure 7.12. Figure 7.9 also shows satisfactory performance in detecting BAC using Faster R-CNN.

The low average precision obtained may be due insufficient training data although data augmentation was used to increase the training set to 2604 images. Furthermore, the BAC "object" represents very little of the bounding box area (in most cases less than 1%) which may make training and detection difficult. Hyperparameter tuning was also curtailed in some instances by the limits of the training hardware such as in the case of mini-batch size.

Another study using Faster R-CNN and MATLAB for breast cancer detection (Reiazi, Paydar, Ardakani, & Etedadialiabadi, 2018) found similar results, i.e. an average precision of 0.1. Our aim was to improve on the previous research by K. Wang et al. (2019), which

was achieved. At higher IoU thresholds there may be some potential to use this approach as an initial screen. The low metrics obtained overall, however, present a major obstacle to its adoption for this task.

7.6 Summary

This chapter looked at BAC object detection using both multi-stage (Faster R-CNN) and single-stage (YOLOv4) detectors. Models were developed using our ResNet-22 BAC classification network as a feature extraction module. Despite showing some promise at higher IoU thresholds, the models performed poorly with a maximum average precision of approximately 0.06. Our results indicate that object detection techniques may not be suitable for BAC detection.

Chapter 8 will examine segmentation using deep learning as a potentially more promising method for automatic detection of BAC. After preliminary testing of several network architectures trained on a subset of images, we choose a model most suitable for BAC segmentation. This model will then be trained and evaluated using 220 images manually segmented for BAC.

8 | BAC Segmentation Model

Gonzalez et al. (2020) note that the segmentation of non-trivial images is one of the most challenging tasks in image processing where accuracy determines the success or failure of image analysis procedures. They describe this challenge as a “pixel-labelling” problem in which pixels are assigned to labeled classes, the union of which constitutes an image. This is also known as *semantic segmentation* (Minaee et al., 2022) with *instance segmentation* denoting the accurate delineation of each object.

Deep learning models have outperformed and replaced traditional segmentation methods such as thresholding (Otsu, 1979), k-means clustering (Dhanachandra, Manglem, & Chanu, 2015) and even advanced algorithms such as active contours (Kass, Witkin, & Terzopoulos, 1988). This has resulted in a paradigm shift in the field (Minaee et al., 2022).

Per-pixel semantic labeling using a single neural network was enabled by fully convolutional networks (FCNs) which combined semantic information from deep, coarse layers and appearance information from shallow, fine layers (Long, Shelhamer, & Darrell, 2015). Szeliski (2022) states that FCN accuracy and resolution were then improved by the addition of conditional random fields (CRFs) at a final stage such as those used in DeepLab (L.-C. Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2018), deconvolutional upsampling (Noh, Hong, & Han, 2015) and fine-level connections in U-nets (Ronneberger et al., 2015).

The next section first introduces architectures considered for BAC segmentation and then describes the preliminary training undertaken to find the most promising model for the task.

8.1 Preliminary Training

U-Net (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017) and the DeepLab family of networks (L.-C. Chen, Papandreou, et al., 2018) are popular segmentation networks. U-Nets consist of two parts, a contracting path to capture context and a symmetric expanding path that enables precise localisation. The original U-net study used a small number of light-transmitted microscopy images (n=30) along with data augmentation to produce their results. The images only contained two intensities, black and white.

SegNet built on the deconvolutional upsampling of Noh et al. (2015)'s network decoder by using pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear up-sampling (Minaee et al., 2022).

Early iterations of DeepLab networks incorporated three features that made them suitable for segmentation tasks (L.-C. Chen, Papandreou, et al., 2018):

- Dilated, or "atrous", convolutions, shown in Figure 8.1, that reversed the adverse effects on resolution of max-pooling and striding. A 3x3 kernel with a dilation rate of 2 will have the same size receptive field as a 5x5 kernel while only using 9 parameters and at no additional computational cost (Minaee et al., 2022).

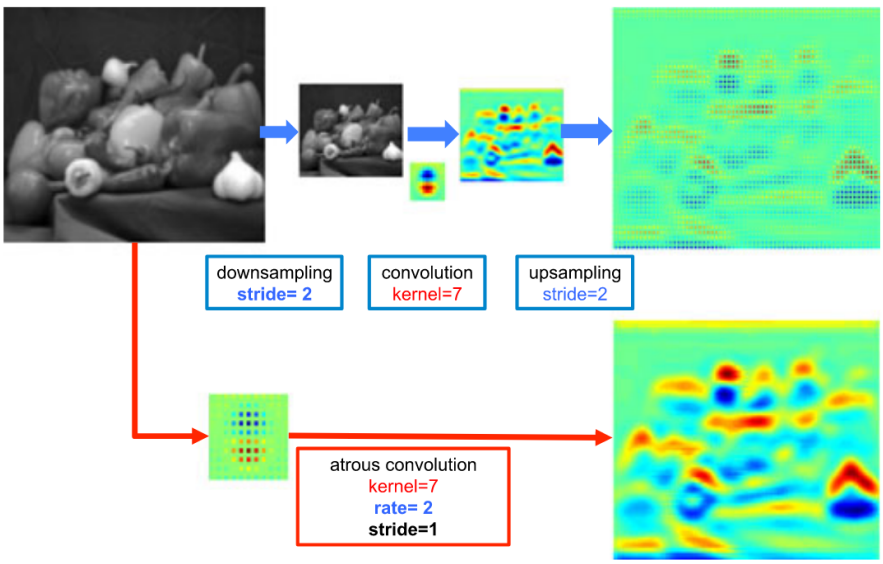


Figure 8.1: Illustration of atrous convolution in 2-D. From L.-C. Chen, Papandreou, et al. (2018).

- Atrous spatial pyramid pooling (ASPP), shown in Figure 8.1 which allows the capture

of both objects and multi-scale image context due to the ability to filter the incoming feature layer at multiple sampling rates.

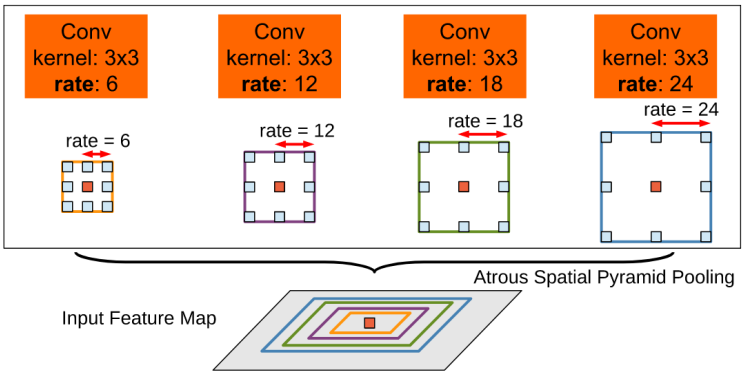


Figure 8.2: Atrous spatial pyramid pooling (ASPP). From L.-C. Chen, Papandreou, et al. (2018).

- Combining methods such as fully convolutional networks and CRFs to improve the localisation of object boundaries.

DeepLabv3 (L.-C. Chen, Papandreou, Schroff, & Adam, 2017) combined cascaded and parallel convolutions which allowed multiple dilation rates in order to further improve the capture of multi-scale context. DeepLabv3+ (L.-C. Chen, Papandreou, et al., 2018) uses DeepLabv3 as the encoder in a encoder-decoder architecture with the decoder module refining segmentations along object boundaries.

Given the various options, a number of networks, including U-Net, SegNet and DeepLabv3+ with several backbones, were initially trained on a subset of images in order to identify those more suitable for BAC segmentation. Unfortunately, MATLAB's pre-trained segmentation models do not support custom backbones. Backbones used with DeepLabv3+ included ResNet18 (He et al., 2016), Xception (Chollet, 2017) and Inception ResNetv2 (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017). ResNet18 is closest to the ResNet22 network we used for classification and object detection, minus one residual block. Inception ResNetv2 builds on the Inception (Szegedy et al., 2015) architecture by replacing its filter concatenation stage with residual connections thus reaping the benefits of the residual approach while maintaining computational efficiency. The Xception model also extends the Inception approach by replacing Inception modules with depthwise separable convolutions.

Models were trained using 61, 169 and 200 images. Visual results from the first stage

are shown in Figure 8.3. Quantitative preliminary results are shown in Table 8.1. The DeepLabv3+ models performed best at the first stage with DeepLabv3+ ResNet18 achieving a BFScore of 0.67. The Inception ResNetv2 backbone network was eliminated as it took 60 hours to complete training with the rest of the models taking 6 to 8 hours. At the next two stages, DeepLabv3+ ResNet18 attained better results than DeepLabv3+ Xception in terms of accuracy and BFScore and was chosen as our BAC segmentation model.

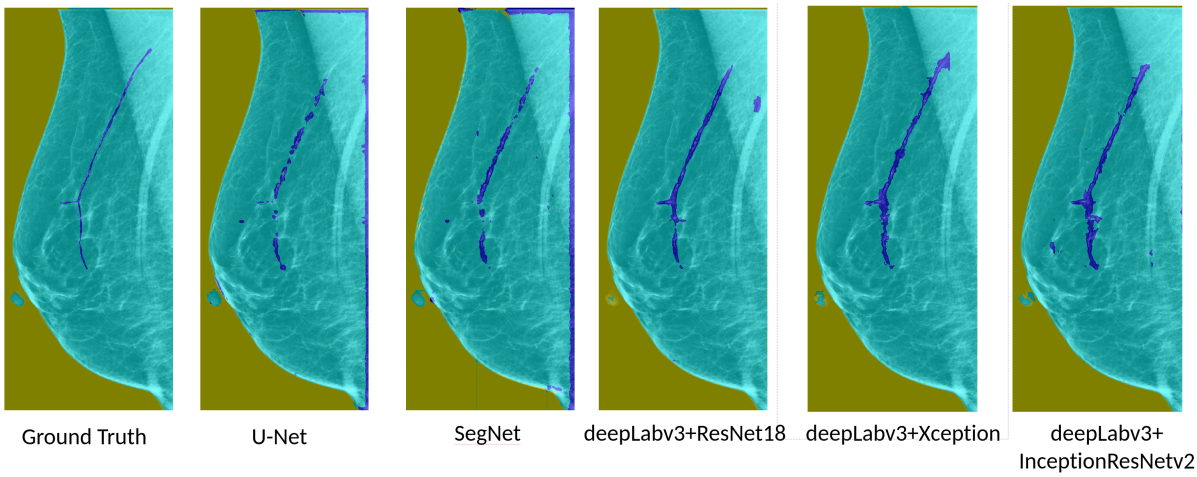


Figure 8.3: Visual segmentation results after training each model with 61 images.

Table 8.1: Preliminary Segmentation Network Results.

Model	**Layers**	**Images**	**Patches**	**Acc**	**IoU**	**BFScore**
U-Net	58	61	7808	0.30	0.10	0.26
SegNet	59	61	7808	0.42	0.05	0.29
DeepLabv3+ ResNet18	100	61	7808	0.54	0.30	0.67
DeepLabv3+ Xception	205	61	7808	0.45	0.26	0.64
DeepLabv3+ Inception ResNetv2	853	61	7808	0.56	0.23	0.52
DeepLabv3+ ResNet18	100	169	21632	0.48	0.37	0.63
DeepLabv3+ Xception	205	169	21632	0.25	0.217	0.41
DeepLabv3+ ResNet18	100	200	25600	0.57	0.33	0.604
DeepLabv3+ Xception	205	200	25600	0.373	0.29	0.59

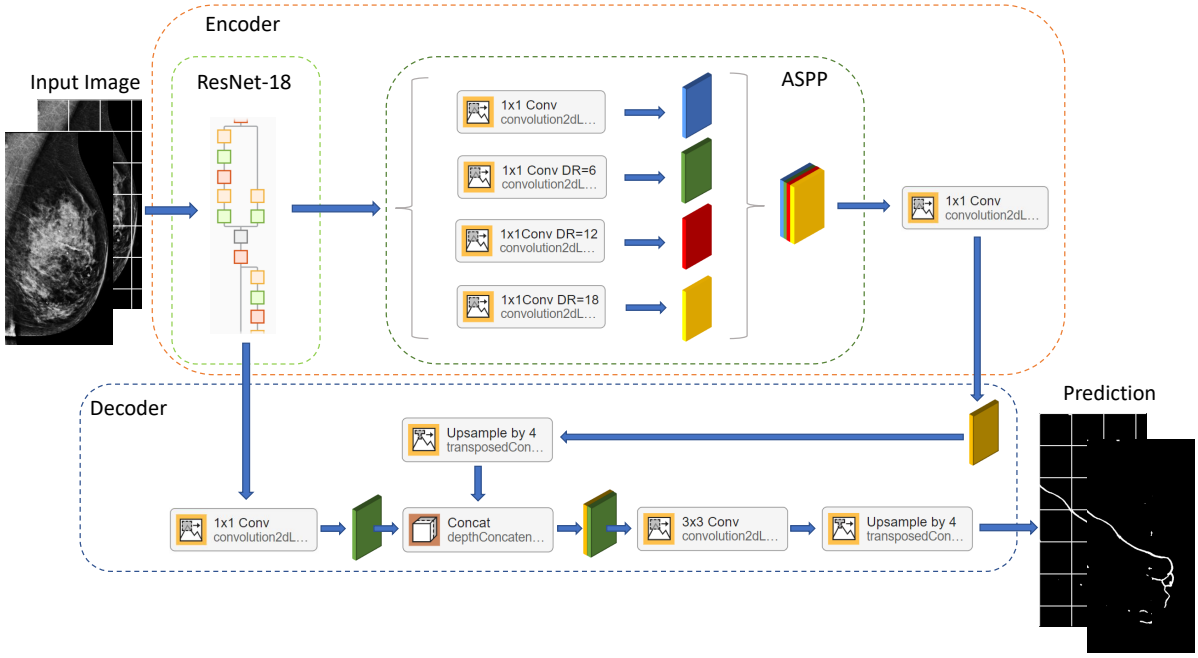


Figure 8.4: DeepLabv3+-ResNet18 BAC segmentation network. N.B. MATLAB's DeepLabv3+ implementation does not include a global average pooling layer in the ASPP.

8.2 Network Training

Given the above preliminary investigation, the DeepLabv3+-ResNet18 model shown in Figure 8.4, was used for BAC segmentation. A layer-by-layer description is available in Appendix I. MATLAB's implementation of DeepLabv3+-ResNet18 makes use of extracting random patches from two image-based datastores containing ground truth images and pixel label data and can handle multiple image sizes as input. 220 manually-segmented ground truth images were cropped to the breast and converted to RGB (required by ResNet-18) with the smallest and largest being 687x1904 and 2968x3944 respectively. The number of patches per image was set to 128 with a size of 512x512 providing 28160 patches in total.

Image pixels had been already annotated for three classes during pre-processing: breast, background and BAC. Class weights using inverse frequency were calculated in order to increase the weight given to BAC. The training:validation:test ratio was 80:10:10 leading to a 22528:2816:2816 patch ratio. With an initial learning rate of 0.001 and a mini-batch size of 16, the 100-layer model was trained for 70 epochs using stochastic gradient descent with a momentum of 0.9. Training lasted 78 hours as shown in Figure 8.5.

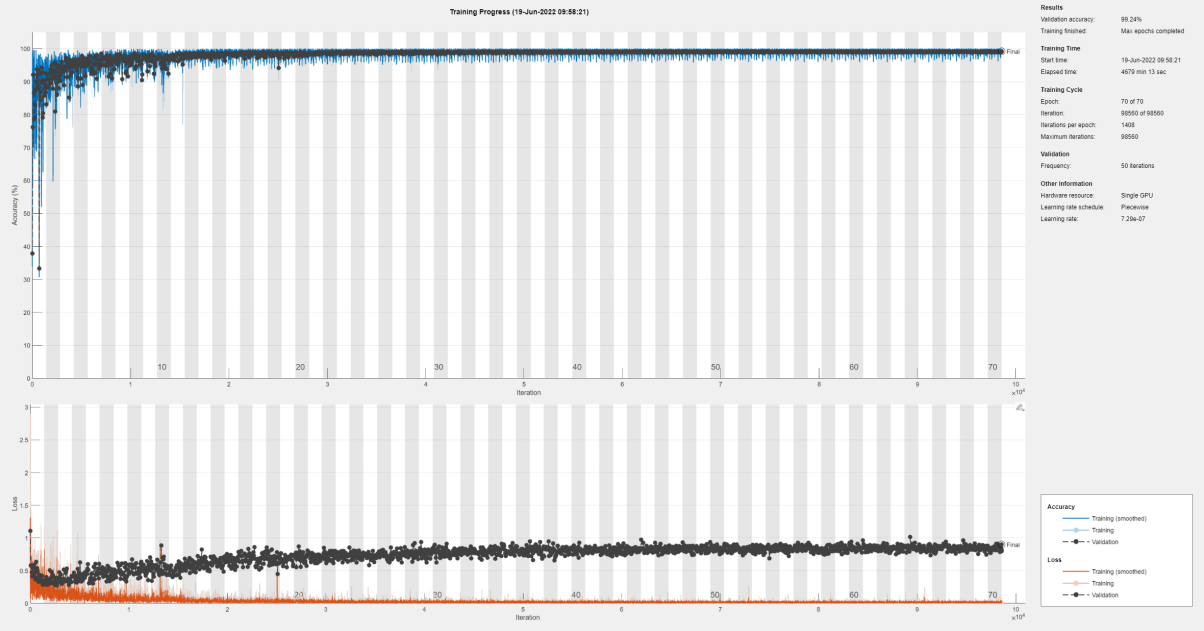


Figure 8.5: Training progress visualisation for DeepLabv3+ ResNet18 model over 70 epochs.

8.2.1 Metrics

Test images were segmented patch-wise and performance of the model for BAC segmentation was evaluated globally and class-wise in terms of accuracy, IoU (Intersection over Union) and mean BF (Boundary F1) score. The latter metric is the harmonic mean of precision and recall with a distance error tolerance to decide whether a point on the predicted boundary matches one on the ground truth boundary or not (Csurka, Larlus, & Perronnin, 2013). The metrics are defined as follows (accuracy is restated for ease of reference):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

$$BFScore = \frac{2 * Precision * Recall}{(Precision + Recall)}$$

where TP is the true positive, TN is the true negative, FP is the false positive and FN is the false negative.

8.3 Results

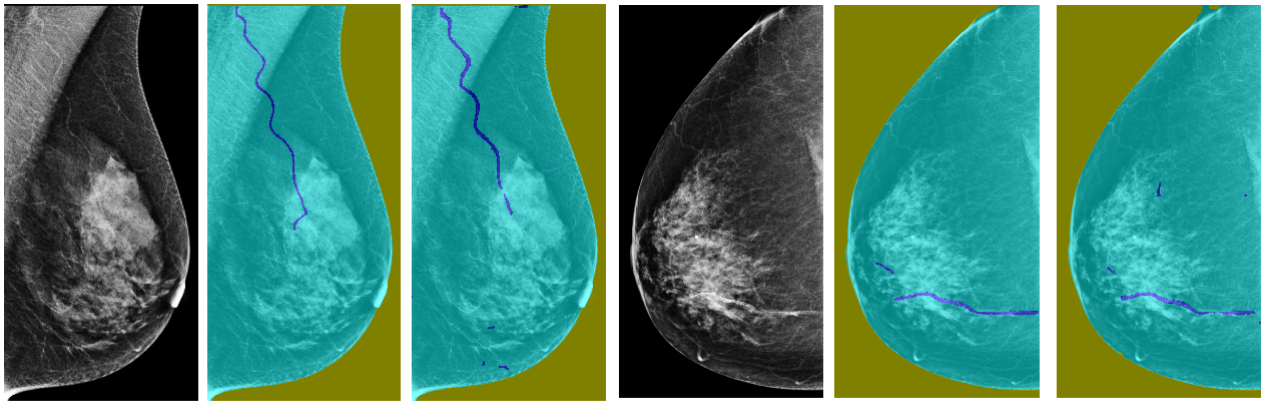


Figure 8.6: Visual comparison of BAC ground truth and predicted BAC on two images. BAC is denoted as blue, with background and breast tissue olive green and turquoise respectively.

Figure 8.6 shows a visual comparison of BAC ground truth and predicted BAC on two images. Background and breast tissue were accurately segmented with predicted BAC being satisfactory in terms of location on the whole although some small areas outside ground truth were classified as BAC.

Table 8.2 shows semantic segmentation metrics aggregated over the entire dataset. Global accuracy, a measure of correctly classified pixels regardless of class (background, BAC or breast), was found to be 0.9937 whereas the mean accuracy, the ratio of correctly classified pixels in each class to total pixels, was found to be 0.8585.

Table 8.2: Segmentation Global Results.

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BFScore
0.9937	0.8585	0.8044	0.9886	0.8636

The class-level results are shown in Table 8.3 with BAC segmentation accuracy lower than that of both background and breast at 0.5834. IoU and BFScores for BAC were found to be 0.4283 and 0.7019 respectively, again lower than the other two classes. The normalized confusion matrix shown in Figure 8.7 indicates that many more BAC pixels were wrongly classified as breast tissue (41.66%) compared to background (0.003162%).

Table 8.3: Segmentation Class-level Results.

Class	Accuracy	IoU	Mean BFScore
background	0.9961	0.9937	0.9683
BAC	0.5834	0.4283	0.7019
breast	0.9961	0.9911	0.9206

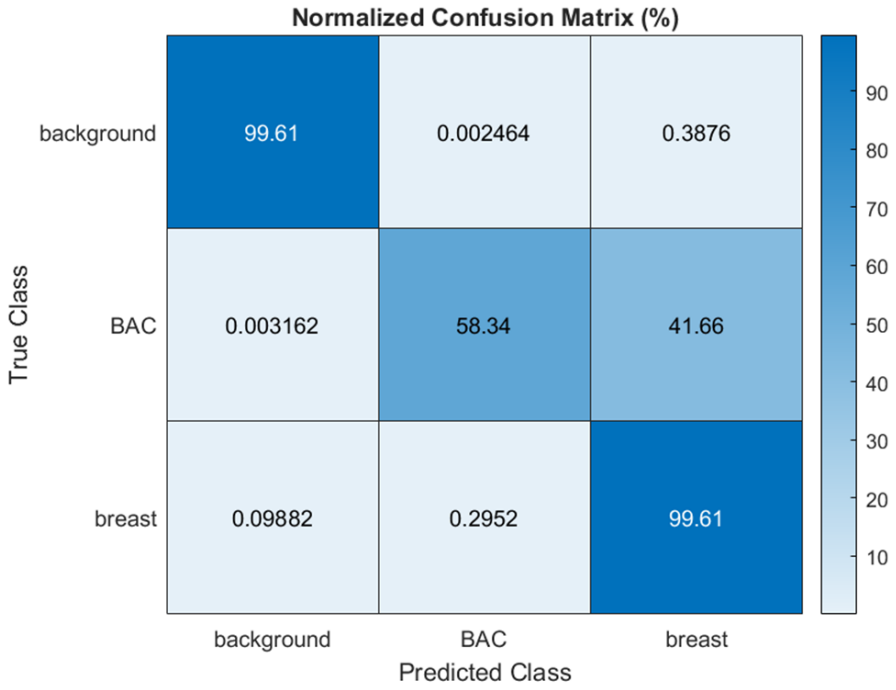


Figure 8.7: Normalized segmentation confusion matrix.

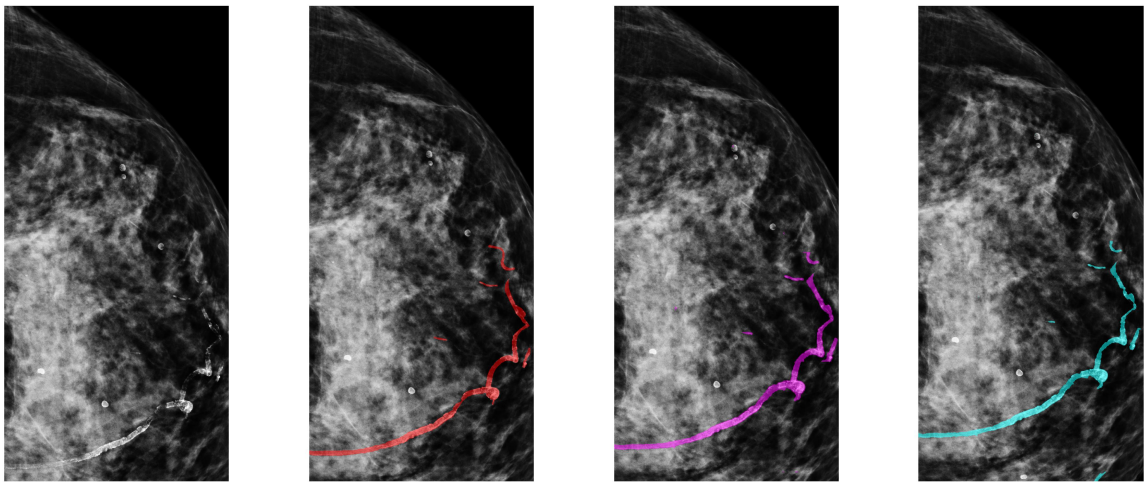
8.3.1 Post-processing

Minaee et al. (2022) note that some researchers have used active contours as a simple post-processor to the output of segmentation models. We applied the edge-based MATLAB `activecontour` function to the masks generated through training in order to ascertain whether the predicted boundaries could be improved in relation to ground truth. Table 8.4 shows the effect of using active contouring post-processing for BAC segmentation. IoU increased by 3% to 0.4415 whereas accuracy decreased by 9% to 0.5307.

Table 8.4: BAC segmentation class-level result comparison.

	Accuracy	IoU		Mean BFScor
No post-processing	0.5834	0.4283	0.7019Class weights using inverse frequency were calculate	
Post-processing	0.5307	0.4415		0.7222

Figure 8.8 shows the stages of BAC segmentation from the original image to ground truth, from predicted BAC to the post-processed revised prediction.



(a) Original image. (b) BAC ground truth. (c) Predicted BAC. (d) Revised prediction.

Figure 8.8: An example showing BAC Segmentation Stages.

8.4 Discussion

BAC segmentation is probably the most important pre-requisite for automating BAC grading as it forms the quantitative basis of any calcification measurement and provides a more granular location assessment than that of bounding boxes. Our DeepLabv3+ with ResNet18 backbone model doubled the best IoU score for BAC segmentation obtained by K. Wang et al. (2019) who also used a DeepLabv3+-based network.

Our global results for segmentation accuracy matched those in two more recent papers (Ghamdi et al., 2020; Guo et al., 2021) but were much lower specifically for the BAC class. Wrongly classed BAC pixels were mostly classed as breast tissue. Only 23 BAC pixels were wrongly classed as background. Predicted BAC segmentations were processed with an edge-based active contour function which increased the mean IoU but decreased accuracy. More research is needed to ascertain the effectiveness of this post-processing technique.

The BFScore measures how close the predicted boundary of an object matches the ground truth boundary and replicates human qualitative assessment more so than the IoU

metric (Csurka et al., 2013). The good BFScore we achieved indicates that our model shows promise for BAC segmentation.

8.5 Summary

This chapter examined BAC segmentation using a DeepLabv3+ model with a ResNet18 backbone after having initially training a range of models including U-Net and SegNet on a subset of the 220 hand-annotated FFDM image dataset. We improved IoU and matched global segmentation accuracy compared to other papers in the literature while also achieving good boundary predictions. This indicates that this model has the potential to be used in BAC segmentation tasks.

The next chapter will review the research objectives and draw conclusions on our approaches to BAC classification, detection and segmentation, outlining our contribution to the field. The limitations of our study will be described and recommendations for future research will be made.

9 | Conclusions and Future Work

9.1 Summary

In this study, we successfully achieved the research objectives set out in Chapter 1. We conducted a review of the literature regarding BAC and its detection, classification and segmentation using automatic, computerised methods, identifying the current state of the art. We obtained, pre-processed, augmented and annotated an anonymised dataset of mammography images in preparation for BAC model training. The latter annotation task was validated to a satisfactory standard by an observer reader study involving two consultant radiologists, one of whom also provided guidance for pixel-level manual segmentation.

We investigated a number of deep learning models to aid automatic classification, detection and segmentation of BAC in order to assist radiologists in determining the presence or absence of BAC, where it is located and how extensive it is. Our main research question sought to ascertain how well our models performed in these three tasks. For classification, we used a custom ResNet22 network initialised with weights from a recent breast cancer detection study. A test accuracy of 80% indicates that this method could be used as a simple flag for the presence or absence of BAC. The ResNet22 network developed was also used as a feature extraction network for Faster R-CNN and YOLOv4 BAC object detection models. These models performed poorly with very low average precision scores leading us to reject the hypothesis that they could be successfully used for region-level BAC object detection. More promising was a DeepLabv3+-based network for pixel-level BAC segmentation which obtained a BFScore of over 70% and doubled the IoU achieved by a study that used a similar model.

Based on the findings of this research, a two-step pipeline is recommended with our classifier triaging mammographic images for BAC and our segmentation model providing an indication of the extent of its presence.

In conclusion, the main contributions of this study are: (i) it presents a comprehensive and systematic review of BAC classification, object detection and segmentation approaches using deep learning, (ii) it creates a new dataset, validated by consultant radiologists, which has been annotated for the presence and location of BAC at an image, region and pixel level and (iii) deep learning models for BAC classification, object detection and segmentation have been developed and evaluated.

9.2 Study Limitations

The limitations of this study included the non-availability of any large public dataset with BAC-annotated images for training. It would also have been useful to have had another dataset to compare the performance of our models directly. Deep learning studies in retinal imaging can avail of numerous high quality benchmark datasets (S. M. Khan et al., 2021). A similar approach would be useful for BAC studies.

Ground truth for the presence, location and extent of BAC was provided by the author who, while having significant medical imaging domain knowledge, is not a breast radiologist. Ideally, ground truth would be provided by the latter practitioners but, as we have shown, this is currently a challenging task to do so on a large scale.

Training full-size mammography images proved to be computationally expensive and tested the memory limits of our GPU hardware and presented time challenges. This resulted in reducing the image sizes to 70% of the original for both the classification and object detection models. For the same reason, we also had to forego parallel processing features available in MATLAB during the object detection task.

9.3 Future Work

The results of this study show promise in relation to BAC classification and segmentation as a first step towards automatic BAC grading to predict cardiovascular risk in patients undergoing mammographic imaging and to direct women to further approved cardiac diagnostic testing and treatment programs. There are several other aspects which would progress this journey further and these include:

- The other papers examining BAC segmentation using deep learning had access to similar small datasets as our own study. Our metrics for BAC segmentation improved as we added more images. Many more images will be needed for future studies. Multi-centre curation of images via a federated learning platform, for example, could prove useful for training and minimising dataset bias.

- It would also be useful for a consensus to be reached on standardised BAC reporting and what exactly BAC grading should be, i.e. is it a quantitative measure of area or length?, and what subsequent clinical pathway actions need to happen based on its severity.

- Future research should have a more clinical focus with prospective studies that monitor patients over time. Our retrospective study used images from patients within a narrow age range, 65-70. This cohort could be enlarged to include all age ranges over forty. Clinical data such as BMI, blood pressure and other risk factor parameters could be added to existing model features to boost network performance. Practical clinical implementations deployed onsite would provide valuable feedback from clinicians on their use and from metrics on their accuracy in the real world.

- It would be beneficial to investigate other types of deep models such as generative adversarial networks as to their suitability for the above tasks. As we have shown, combining deep models with traditional image processing such as active contours could also prove a rewarding research pathway. All of the BAC papers used supervised

learning to train their models. Semi- and unsupervised learning methods could prove useful while also reducing the manual annotation overhead.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from https://www.tensorflow.org/ (Software available from tensorflow.org)

Abi Rafeh, N., Castellanos, M. R., Khoueiry, G., Meghani, M., El-Sayegh, S., Wetz, R. V., ... Kowalski, M. (2012). Association Between Coronary Artery Disease Diagnosed by Coronary Angiography and Breast Arterial Calcifications on Mammography: Meta-Analysis of the Data. *Journal of Women's Health*, *21*(10), 1053–1058. doi: 10.1089/jwh.2011.3388

Agatston, A. S., Janowitz, W. R., Hildner, F. J., Zusmer, N. R., Viamonte, M., & Detrano, R. (1990). Quantification of coronary artery calcium using ultrafast computed tomography. *Journal of the American College of Cardiology*, *15*(4), 827–832. doi: https://doi.org/10.1016/0735-1097(90)90282-T

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B. C., ... Asari, V. K. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. Retrieved from http://arxiv.org/abs/1803.01164 doi: 10.1016/S0011-9164(00)80105-8

Arbab-Zadeh, A. (2016). What Will it Take to Retire Invasive Coronary Angiography? *JACC: Cardiovascular Imaging*, *9*(5), 565–567. doi: 10.1016/j.jcmg.2015.09.013

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*, *39*(12), 2481-2495. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28060704 doi: 10.1109/TPAMI.2016.2644615

Bays, H. E., Kulkarni, A., German, C., Satish, P., Iluyomade, A., Dudum, R., ... Toth, P. P. (2022). Ten things to know about ten cardiovascular disease risk factors - 2022. *Am J Prev Cardiol*, *10*, 100342. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/35517870 (Edition: 20220406) doi: 10.1016/j.ajpc.2022.100342

Beam, C. A., Layde, P. M., & Sullivan, D. C. (1996). Variability in the Interpretation of Screening Mammograms by US Radiologists: Findings From a National Sample. *Archives of Internal Medicine*, *156*(2), 209–213. Retrieved from https://dx.doi.org/10.1001/archinte.1996.00440020119016 doi: 10.1001/archinte.1996.00440020119016

Bengio, Y., Courville, A., & Vincent, P. (2012). *Representation learning: A review and new perspectives* [Electronic Article]. (Retrieved from https://ui.adsabs.harvard.edu/abs/2012arXiv1206.5538B. Accessed December 22, 2022)

Bober, R. M., & Jahangir, E. (2015). What Is Ischemia and How Should This Be Defined Based on Modern Imaging? *Progress in Cardiovascular Diseases*, *57*(6), 537–554. Retrieved from http://dx.doi.org/10.1016/j.pcad.2015.02.001 doi: 10.1016/j.pcad.2015.02.001

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv*. (Retrieved from https://arxiv.org/abs/2004.10934. Accessed December 22, 2022)

Boudoulas, K. D., Triposciadis, F., Geleris, P., & Boudoulas, H. (2016). Coronary Atherosclerosis: Pathophysiologic Basis for Diagnosis and Management. *Progress in Cardiovascular Diseases*, *58*(6), 676–692. doi: 10.1016/j.pcad.2016.04.003

Boyer, B., Balleyguier, C., Granat, O., & Pharaboz, C. (2008). CAD in questions/answers. *European Journal of Radiology*, *69*(1), 24–33. doi: 10.1016/j.ejrad.2008.07.042

British Heart Foundation. (2018). UK CVD Factsheet. (October). Retrieved from https://www.bhf.org.uk/-/media/files/research/heart-statistics/bhf-cvd-statistics---uk-factsheet.pdf?la=en

British Heart Foundation. (2022). *UK CVD Factsheet*. London: British Heart Foundation.

Brown, A. L., Wahab, R. A., Zhang, B., Smetherman, D. H., & Mahoney, M. C. (2022, January). Reporting and Perceptions of Breast Arterial Calcification on Mammog-

raphy: A Survey of ACR Radiologists. *Academic Radiology*, *29*, S192–S198. Retrieved 2023-06-27, from https://www.sciencedirect.com/science/article/pii/S1076633221000362 doi: 10.1016/j.acra.2021.01.027

Bryson, A. E., & Ho, Y.-c. (1969). *Applied Optimal Control*. Waltham, MA: Blaisdell.

Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis*. London: Routledge. Retrieved from https://doi.org/10.4324/9781315242804

Chadashvili, T., Litmanovich, D., Hall, F., & Slanetz, P. J. (2016). Do breast arterial calcifications on mammography predict elevated risk of coronary artery disease? *European Journal of Radiology*, *85*(6), 1121–1124. doi: https://doi.org/10.1016/j.ejrad.2016.03.006

Chan, H.-P., Doi, K., Vybrony, C. J., Schmidt, R. A., Metz, C. E., & Lam, K. L. (1990). Improvement in Radiologists' Detection of Clustered Microcalcifications on Mammograms: The Potential of Computer-Aided Diagnosis. *Investigative Radiology*, *25*(10), 1102–1110. doi: 10.1097/00004424-199010000-00006

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide.* (Tech. Rep.). Various: SPSS. Retrieved 2024-11-02, from https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. Retrieved 2024-02-24, from http://ieeexplore.ieee.org/document/7913730/ doi: 10.1109/TPAMI.2017.2699184

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation.* arXiv. Retrieved 2024-02-24, from http://arxiv.org/abs/1706.05587 (arXiv:1706.05587 [cs])

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation [Journal Article]. *Lecture Notes in Computer Science (including subseries Lecture Notes in Ar-*

tificial Intelligence and Lecture Notes in Bioinformatics), *11211 LNCS*, 833-851. doi: 10.1007/978-3-030-01234-2_49

Chen, Y., Taib, A. G., Darker, I. T., & James, J. J. (2023). Performance of a breast cancer detection ai algorithm using the personal performance in mammographic screening scheme. *Radiology*, *308*(3), e223299. Retrieved from https://doi.org/10.1148/radiol.223299 (PMID: 37668522) doi: 10.1148/radiol.223299

Cheng, J. Z., Chen, C. M., Cole, E. B., Pisano, E. D., & Shen, D. (2012). Automated delineation of calcified vessels in mammography by tracking with uncertainty and graphical linking techniques. *IEEE Transactions on Medical Imaging*, *31*(11), 2143–2155. doi: 10.1109/TMI.2012.2215880

Cheng, J. Z., Chen, C. M., & Shen, D. (2012). Identification of breast vascular calcium deposition in digital mammography by linear structure analysis. *Proceedings - International Symposium on Biomedical Imaging*, 126-129. doi: 10.1109/ISBI.2012.6235500

Cheng, P. M., Montagnon, E., Yamashita, R., Pan, I., Cadrin-Chênevert, A., Perdigón Romero, F., ... Tang, A. (2021). Deep Learning: An Update for Radiologists. *RadioGraphics*, *41*(5), 1427–1445. Retrieved 2023-09-01, from http://pubs.rsna.org/doi/10.1148/rg.2021200210 doi: 10.1148/rg.2021200210

Chollet, F. (2017). *Xception: Deep Learning with Depthwise Separable Convolutions.* arXiv. Retrieved 2024-02-18, from http://arxiv.org/abs/1610.02357 (arXiv:1610.02357 [cs])

Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Computer Methods and Programs in Biomedicine*, *157*, 19-30. doi: 10.1016/j.cmpb.2018.01.011

Clark, A. (2020). *Pillow* [Computer Program]. (Retrieved from https://pillow.readthedocs.io/en/stable/#. Accessed January 31, 2022)

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. Retrieved 2024-03-13, from http://journals.sagepub.com/doi/10.1177/001316446002000104 doi: 10.1177/001316446002000104

Cole, E. B., Marques, H. S., Zhang, Z., Yaffe, M. J., Pisano, E. D., & Edward Hendrick, R. (2014). Impact of Computer-Aided Detection Systems on Radiologist Accuracy With Digital Mammography. *American Journal of Roentgenology*, *203*(4), 909–916. doi: 10.2214/ajr.12.10187

Csurka, G., Larlus, D., & Perronnin, F. (2013). What is a good evaluation measure for semantic segmentation? [Conference Paper]. In *BMVC 2013.* Bristol, UK.

Cunha Carneiro, P., Lemos Debs, C., Oliveira Andrade, A., & Patrocinio, A. C. (2019). Clahe parameters effects on the quantitative and visual assessment of dense breast mammograms. *IEEE Latin America Transactions*, *17*(05), 851-857. doi: 10.1109/TLA.2019.8891954

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). doi: 10.1109/CVPR.2009.5206848

Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. *Eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India*, *54*, 764–771. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050915014143 doi: 10.1016/j.procs.2015.06.090

Elgendy, M. (2020). *Deep Learning for Vision Systems*. Shelter Island, New York: Manning Publications.

Erbil, D., Onal, E. M., Demirel, C., Sag, A. A., Dagel, T., Afsar, B., ... Kanbay, M. (2018). Can Incidental Vascular Calcifications at Mammography be Used as a Screening Biomarker for Heart and Kidney Disease? *Angiology*, *70*(3), 210–219. doi: 10.1177/0003319718779322

Erickson, B. J. (2019). Deep Learning and Machine Learning in Imaging: Basic Principles. In E. R. Ranschaert, S. Morozov, & P. R. Algra (Eds.), *Artificial intelligence in medical*

imaging: Opportunities, applications and risks (pp. 39–46). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-94878-2{_}4 doi: 10.1007/978-3-319-94878-2_4

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The Pascal Visual Object Classes Challenge: A Retrospective. *International Journal of Computer Vision*, *111*(1), 98–136. Retrieved 2024-01-07, from http://link.springer.com/10.1007/s11263-014-0733-5 doi: 10.1007/s11263-014-0733-5

Galiano, N. G., Eiro, N., Martín, A., Fernández-Guinea, O., Martínez, C. D. B., & Vizoso, F. J. (2022). Relationship between Arterial Calcifications on Mammograms and Cardiovascular Events: A Twenty-Three Year Follow-Up Retrospective Cohort Study. *Biomedicines*, *10*(12), 3227. Retrieved 2023-05-21, from https://www.mdpi.com/2227-9059/10/12/3227 doi: 10.3390/biomedicines10123227

Ge, J., Chan, H. P., Sahiner, B., Zhou, C., Helvie, M. A., Wei, J., ... Shi, J. (2008). Automated detection of breast vascular calcification on full-field digital mammograms. *Medical Imaging 2008: Computer-Aided Diagnosis, Pts 1 and 2*, *6915*(January 2014), 91517. doi: Doi10.1117/12.773096

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). *YOLOX: Exceeding YOLO Series in 2021.* arXiv. Retrieved 2024-04-02, from http://arxiv.org/abs/2107.08430 (arXiv:2107.08430 [cs])

Getty, D. J., Pickett, R. M., D'Orsi, C. J., & Swets, J. A. (1988). Enhanced Interpretation of Diagnostic Images. *Investigative Radiology*, *23*(4).

Ghamdi, M. A., Abdel-Mottaleb, M., & Collado-Mesa, F. (2020). DU-Net: Convolutional Network for the Detection of Arterial Calcifications in Mammograms. *IEEE Transactions on Medical Imaging*, *39*(10), 1–1. doi: 10.1109/tmi.2020.2989737

Girshick, R. (2015). Fast R-CNN. , arXiv:1504.08083. Retrieved from https://ui.adsabs.harvard.edu/abs/2015arXiv150408083G

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the 2014 IEEE*

Conference on Computer Vision and Pattern Recognition (p. 580–587). USA: IEEE Computer Society. Retrieved from https://doi.org/10.1109/CVPR.2014.81 doi: 10.1109/CVPR.2014.81

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In T. Yee Whye & T. Mike (Eds.), (Vol. 9, pp. 249–256). Proceedings of Machine Learning Research: PMLR. Retrieved from http://proceedings.mlr.press/v9/glorot10a.html

Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing* (4th ed.). London: Pearson Education.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2020). *Digital Image Processing Using MATLAB* (3rd ed.). Knoxville: Gatesmark Publishing.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, Massachusetts: The MIT Press.

Gould, K. L. (2009). Does Coronary Flow Trump Coronary Anatomy? *JACC: Cardiovascular Imaging*, *2*(8), 1009–1023. Retrieved from http://dx.doi.org/10.1016/j.jcmg.2009.06.004 doi: 10.1016/j.jcmg.2009.06.004

Graham, R., Perriss, R., & Scarsbrook, A. (2005). DICOM demystified: A review of digital file formats and their use in radiological practice. *Clinical Radiology*, *60*(11), 1133–1140. Retrieved 2023-11-14, from https://linkinghub.elsevier.com/retrieve/pii/S0009926005002199 doi: 10.1016/j.crad.2005.07.003

Guo, X., O'Neill, W. C., Vey, B., Yang, T. C., Kim, T. J., Ghassemi, M., ... Banerjee, I. (2021). SCU-Net: A deep learning method for segmentation and quantification of breast arterial calcifications on mammograms. *Med Phys*, *48*(10), 5851-5861. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/34328661 doi: 10.1002/mp.15017

Hajar, R. (2017). Risk factors for coronary artery disease: Historical perspectives. *Heart Views*, *18*, 109. Retrieved from http://link.galegroup.com.salford.idm.oclc.org/apps/doc/A514275971/AONE?u=salcal2{&}sid=AONE{&}xid=68c3e997

Hall, J. (2023). Mammography News: FDA Clears AI Software for Breast Arterial Calcification Detection. *Diagnostic Imaging*. Retrieved 2023-10-

22, from https://www.diagnosticimaging.com/view/mammography-news-fda-clears-ai-software-for-breast-arterial-calcification-detection

Halling-Brown, M. D., Warren, L. M., Ward, D., Lewis, E., Mackenzie, A., Wallis, M. G., ... Young, K. C. (2021). OPTIMAM Mammography Image Database: A large-scale resource of mammography images and clinical data. *Radiology: Artificial Intelligence*, *3*(1), e200103. Retrieved from https://pubs.rsna.org/doi/abs/10.1148/ryai.2020200103 doi: 10.1148/ryai.2020200103

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., & Zwiggelaar, R. (2018). Deep learning in mammography and breast histology, an overview and future trends. *Medical Image Analysis*, *47*, 45–67. doi: 10.1016/j.media.2018.03.006

Harvey, H., Heindl, A., Khara, G., Korkinof, D., O'Neill, M., Yearsley, J., ... Forrai, G. (2019). Deep Learning in Breast Cancer Screening. In E. R. Ranschaert, S. Morozov, & P. R. Algra (Eds.), *Artificial intelligence in medical imaging: Opportunities, applications and risks* (pp. 187–215). Cham: Springer International Publishing. doi: 10.1007/978-3-319-94878-2_14

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). Las Vegas, NV, USA: IEEE. Retrieved 2023-02-25, from http://ieeexplore.ieee.org/document/7780459/ doi: 10.1109/CVPR.2016.90

Heaney, R. M., Zaki-Metias, K. M., McKee, H., Wang, H., Ogunde, B., Yong-Hing, C. J., ... Nguyen, E. T. (2022). Correlation Between Breast Arterial Calcifications and Higher Cardiovascular Risk: Awareness and Attitudes Amongst Canadian Radiologists Who Report Mammography. *Canadian Association of Radiologists Journal*, 084653712211403. Retrieved 2023-05-21, from http://journals.sagepub.com/doi/10.1177/08465371221140347 doi: 10.1177/08465371221140347

Heim, E., Roß, T., Seitel, A., März, K., Stieltjes, B., Eisenmann, M., ... Sommer, G. (2018). Large-scale medical image annotation with crowd-powered algorithms. *Journal of Medical Imaging*, *5*(03), 1. Retrieved 2024-01-04, from https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/

volume-5/issue-03/034002/Large-scale-medical-image-annotation-with-crowd-powered-algorithms/10.1117/1.JMI.5.3.034002.full doi: 10.1117/1.JMI.5.3.034002

Hendriks, E. J. E., De Jong, P. A., van der Graaf, Y., Mali, W. P. T. M., van der Schouw, Y. T., & Beulens, J. W. J. (2015). Breast arterial calcifications: A systematic review and meta-analysis of their determinants and their association with cardiovascular events. *Atherosclerosis*, *239*(1), 11-20. doi: 10.1016/j.atherosclerosis.2014.12.035

Henein, M. Y., Vancheri, S., Bajraktari, G., & Vancheri, F. (2020). Coronary Atherosclerosis Imaging. *Diagnostics (Basel)*, *10*(2). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/31991633 (Edition: 2020/01/30) doi: 10.3390/diagnostics10020065

Highnam, R., Brady, M., Yaffe, M. J., Karssemeijer, N., & Harvey, J. (2010). Robust breast composition measurement - Volpara™. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6136 LNCS*, 342-349. doi: 10.1007/978-3-642-13666-5_46

Hinton, G. (2018). Deep Learning—A Technology With the Potential to Transform Health Care. *JAMA*, *320*(11), 1101–1102. Retrieved from https://doi.org/10.1001/jama.2018.11100 doi: 10.1001/jama.2018.11100

Hinton, G. E. (2012). *Neural networks for machine learning lecture 6e. rmsprop: Divide the gradient by a running average of its recent magnitude.* [Lecture Presentation]. Retrieved 2023-01-14, from https://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput*, *18*(7), 1527–54. (Edition: 2006/06/13) doi: 10.1162/neco.2006.18.7.1527

Hippisley-Cox, J., & Coupland, C. (2017). Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ (Clinical research ed.)*, *359*(May), j5019. doi: 10.1136/bmj.j5019

Hogg, P., Kelly, J., & Mercer, C. (2015). *Digital Mammography: A Holistic Approach*. London: Springer.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv*. Retrieved 2024-01-14, from http://arxiv.org/abs/1602.07360 (arXiv:1602.07360 [cs])

Iivari, J., Hirschheim, R., & Klein, H. K. (1998). A Paradigmatic Analysis Contrasting Information Systems Development Approaches and Methodologies. *Information Systems Research*, *9*(2), 164–193. Retrieved 2023-06-11, from https://www.jstor.org/stable/23010925 (Publisher: INFORMS)

Information Commisioners Office. (2012). *Anonymisation: Managing data protection risk code of practice.* Chesire: Information Commisioners Office.

Iribarren, C., Chandra, M., Lee, C., Sanchez, G., Sam, D. L., Azamian, F. F., ... Molloi, S. (2022). Breast Arterial Calcification: a Novel Cardiovascular Risk Enhancer Among Postmenopausal Women. *Circ Cardiovasc Imaging*, *15*(3), e013526. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/35290077 (Edition: 20220315) doi: 10.1161/CIRCIMAGING.121.013526

Iribarren, C., Chandra, M., Parikh, R. V., Sanchez, G., Sam, D. L., Azamian, F. F., ... Go, A. S. (2023). Breast arterial calcification is associated with incident atrial fibrillation among older but not younger post-menopausal women. *European Heart Journal Open*, *3*(2), oead017. Retrieved 2023-05-21, from https://academic.oup.com/ehjopen/article/doi/10.1093/ehjopen/oead017/7060080 doi: 10.1093/ehjopen/oead017

Iribarren, C., Go, A. S., Tolstykh, I., Sidney, S., Johnston, S. C., & Spring, D. B. (2004). Breast Vascular Calcification and Risk of Coronary Heart Disease, Stroke, and Heart Failure. *Journal of Women's Health*, *13*(4), 381–389. doi: 10.1089/154099904323087060

Iribarren, C., & Molloi, S. (2013). Breast Arterial Calcification: A New Marker of Cardiovascular Risk? *Current Cardiovascular Risk Reports*, *7*(2), 126–135. doi: 10.1007/s12170-013-0290-4

Iribarren, C., Sanchez, G., Husson, G., Levine-Hall, T., Quesenberry, C., Sam, D. L., ... Molloi, S. (2017). MultIethNic Study of BrEast ARterial Calcium Gradation and Car-

dioVAscular Disease: Cohort recruitment and baseline characteristics. *Annals of Epidemiology*, *28*(1), 41–47.e12. doi: 10.1016/j.annepidem.2017.11.007

Jain, A., & Levy, D. (2017). *DeepMammo: Breast Mass Classification using Deep Convolutional Neural Networks* (Tech. Rep.). Stanford, California: Stanford University. Retrieved from http://cs231n.stanford.edu/reports/2016/pdfs/306{_}Report.pdf

Jiang, X., Clark, M., Singh, R. K., Juhn, A., & Schnatz, P. F. (2015). Association of breast arterial calcification with stroke and angiographically proven coronary artery disease: a meta-analysis. *Menopause*, *22*(2).

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, *1*(4), 321–331. Retrieved from https://doi.org/10.1007/BF00133570 doi: 10.1007/BF00133570

Katzen, J., & Dodelzon, K. (2018). A review of computer aided detection in mammography. *Clinical Imaging*, *52*(August), 305–309. Retrieved from https://doi.org/10.1016/j.clinimag.2018.08.014 doi: 10.1016/j.clinimag.2018.08.014

Kelly, B. S., Scanlon, E., Heneghan, H., Redmond, C. E., Healy, G. M., Mc Dermott, E., ... Mc Nally, S. (2018). Breast Arterial Calcification on screening mammography can predict significant Coronary Artery Disease in women. *Clinical Imaging*, *49*(October 2017), 48–53. doi: 10.1016/j.clinimag.2017.10.021

Khan, R., & Masala, G. L. (2023). Detecting Breast Arterial Calcifications in Mammograms with Transfer Learning. *Electronics*, *12*(1), 231. Retrieved 2023-05-14, from https://www.mdpi.com/2079-9292/12/1/231 doi: 10.3390/electronics12010231

Khan, S. M., Liu, X., Nath, S., Korot, E., Faes, L., Wagner, S. K., ... Denniston, A. K. (2021). A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *The Lancet Digital Health*, *3*(1), e51–e66. Retrieved 2024-04-03, from https://linkinghub.elsevier.com/retrieve/pii/S2589750020302405 doi: 10.1016/S2589-7500(20)30240-5

Khan, Z., Yahya, N., Alsaih, K., Ali, S. S. A., & Meriaudeau, F. (2020). Evaluation of Deep Neural Networks for Semantic Segmentation of Prostate in T2W MRI. *Sensors (Basel)*,

20(11). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/32503330 doi: 10.3390/s20113183

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization.* arXiv. Retrieved 2020-01-28, from https://arxiv.org/pdf/1412.6980.pdf

Kotre, C. J., & dos Reis, C. (2015). Mammography Equipment. In P. Hogg, J. Kelly, & C. Mercer (Eds.), *Digital mammography: A holistic approach* (pp. 125–141). London: Springer.

Krizhevsky, A., Sulskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information and Processing Systems (NIPS)*, *60*(6), 84-90. doi: 10.1145/3065386

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007

Krupinski, E. A. (2010). Softcopy Reading. In U. Bick & F. Diekmann (Eds.), *Digital mammography.* Heidelberg: Springer Verlag.

Lakhani, P., Prater, A. B., Hutson, R. K., Andriole, K. P., Dreyer, K. J., Morey, J., ... Hawkins, C. M. (2017). Machine Learning in Radiology : Applications Beyond Image Interpretation. *Journal of the American College of Radiology*, *15*(2), 350–359. doi: 10.1016/j.jacr.2017.09.044

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. doi: 10.1038/nature14539

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten Digit Recognition with a Back-Propagation Network. *Advances in Neural Information Processing Systems 2 (NIPS 1989)*. doi: 10.1111/dsu.12130

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2323. doi: 10.1109/5.726791

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., & Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis

research. *Scientific Data*, *4*(1), 170177. Retrieved 2023-10-22, from https://www.nature.com/articles/sdata2017177 doi: 10.1038/sdata.2017.177

Lee, S. C., Pirikahu, S., Fritschi, L., Boyle, T., Schultz, C., Wylie, E., & Stone, J. (2023). The association between breast arterial calcification and atherosclerotic cardiovascular disease in an Australian population-based breast cancer case–control study. *La radiologia medica*, *128*(4), 426–433. Retrieved 2023-05-21, from https://link.springer.com/10.1007/s11547-023-01611-y doi: 10.1007/s11547-023-01611-y

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*(December 2012), 60–88. doi: 10.1016/j.media.2017.07.005

Liu, B., Jiang, Q., Liu, W., Wang, M., Zhang, S., Zhang, X., ... Yue, Z. (2018). A vessel segmentation method for serialized cerebralvascular DSA images based on spatial feature point set of rotating coordinate system. *Computer Methods and Programs in Biomedicine*, *161*, 55–72. Retrieved from https://doi.org/10.1016/j.cmpb.2018.04.010 doi: 10.1016/j.cmpb.2018.04.010

Loizidou, K., Elia, R., & Pitris, C. (2023). Computer-aided breast cancer detection and classification in mammography: A comprehensive review. *Computers in Biology and Medicine*, *153*, 106554. Retrieved 2023-09-01, from https://linkinghub.elsevier.com/retrieve/pii/S0010482523000197 doi: 10.1016/j.compbiomed.2023.106554

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *arXiv*. Retrieved from https://arxiv.org/pdf/1411.4038.pdf

Looney, P. T., Young, K. C., & Halling-Brown, M. D. (2016). Medxviewer: Providing a web-enabled workstation environment for collaborative and remote medical imaging viewing, perception studies and reader training. *Radiat Prot Dosimetry*, *169*(1-4), 32-7. doi: 10.1093/rpd/ncv482

Lotter, W., Rahman Diab, A., Haslam, B., Kim, J. G., Grisot, G., Wu, E., ... Sorensen, A. G. (2019). *Robust breast cancer detection in mammography and digital breast tomosynthesis using annotation-efficient deep learning approach* [Electronic Article].

Retrieved from https://ui.adsabs.harvard.edu/abs/2019arXiv191211027L

Lunit Inc. (n.d.). Retrieved 2023-10-19, from https://www.lunit.io/en/products/mmg

Maas, A. H. E. M., van der Schouw, Y. T., Atsma, F., Beijerinck, D., Deurenberg, J. J. M., Mali, W. P. T. M., & van der Graaf, Y. (2007). Breast arterial calcifications are correlated with subsequent development of coronary artery calcifications, but their aetiology is predominantly different. *European Journal of Radiology*, *63*(3), 396–400. doi: 10.1016/j.ejrad.2007.02.009

Magni, V., Capra, D., Cozzi, A., Monti, C. B., Mobini, N., Colarieti, A., & Sardanelli, F. (2023). Mammography biomarkers of cardiovascular and musculoskeletal health: A review. *Maturitas*, *167*, 75-81. doi: 10.1016/j.maturitas.2022.10.001

Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*, *383*(9921), 999–1008. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24084292https://www.ncbi.nlm.nih.gov/pmc/PMC4159698/ doi: 10.1016/S0140-6736(13)61752-3

Mantas, D., & Markopoulos, C. (2016). Screening mammography: Usefulness beyond early detection of breast cancer. *Atherosclerosis*, *248*(February), 1. Retrieved from http://dx.doi.org/10.1016/j.atherosclerosis.2016.02.019 doi: 10.1016/j.atherosclerosis.2016.02.019

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

Margolies, L., Salvatore, M., Hecht, H. S., Kotkin, S., Yip, R., Baber, U., ... Henschke, C. (2016). Digital Mammography and Screening for Coronary Artery Disease. *JACC: Cardiovascular Imaging*, *9*(4), 350–360. doi: https://doi.org/10.1016/j.jcmg.2015.10.022

Mason, D., scaramallion, rhaxton, mrbean bremen, Suever, J., Vanessasaurus, & Lemaitre, G. (2020). *pydicom* [Computer Program]. (Retrieved from https://pydicom.github.io/. Accessed January 31, 2022)

Mathworks. (2023a). *DICOM Support in Image Processing Toolbox - MAT-*

LAB & Simulink - MathWorks United Kingdom. Retrieved 2023-11-22, from https://uk.mathworks.com/help/images/dicom-support-in-the-image-processing-toolbox.html

Mathworks. (2023b). *Getting Started with YOLO v4.* Retrieved 2024-01-23, from https://uk.mathworks.com/help/vision/ug/getting-started-with-yolo-v4.html

Mathworks. (2023c). *Object Detection Using YOLO v4 Deep Learning - MATLAB & Simulink - MathWorks United Kingdom.* Retrieved 2024-02-08, from https://uk.mathworks.com/help/vision/ug/object-detection-using-yolov4-deep-learning.html

MATLAB. (2022). *R2022a (9.12.0)*. Natick, Massachusetts: The MathWorks Inc.

Matsumura, M. E., Maksimik, C., Martinez, M. W., Weiss, M., Newcomb, J., Harris, K., & Rossi, M. A. (2013). Breast artery calcium noted on screening mammography is predictive of high risk coronary calcium in asymptomatic women: a case control study. *Vasa*, *42*(6), 429–433. Retrieved from https://econtent.hogrefe.com/doi/10.1024/0301-1526/a000312 doi: 10.1024/0301-1526/a000312

Mazidi, N., Roobottom, C., & Masala, G. (2019). Automatic Quantification of Breast Arterial Calcification on Mammographic Images. In Y.-W. Chen, A. Zimmermann, R. J. Howlett, & L. C. Jain (Eds.), *Innovation in medicine and healthcare systems, and multimedia - proceedings of kes-inmed-19 and kes-iimss-19 conferences* (pp. 283–292). Singapore: Springer Singapore. Retrieved from http://dx.doi.org/10.1007/978-981-13-8566-7{_}36 doi: 10.1007/978-981-13-8566-7

Mazurowski, M. A., Buda, M., Saha, A., & Bashir, M. R. (2018). Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *Journal of Magnetic Resonance Imaging*. doi: 10.1002/jmri.26534

McBee, M. P., Awan, O. A., Colucci, A. T., Ghobadi, C. W., Kadom, N., Kansagra, A. P., ... Auffermann, W. F. (2018). Deep Learning in Radiology. *Academic Radiology*, *25*(11), 1472–1480. Retrieved from https://doi.org/10.1016/j.acra.2018.02.018 doi: 10.1016/j.acra.2018.02.018

McClintic, B. R., McClintic, J. I., Bisognano, J. D., & Block, R. C. (2010). The relationship

between retinal microvascular abnormalities and coronary heart disease: a review. *Am J Med*, *123*(4), 374 e1-7. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20362758 doi: 10.1016/j.amjmed.2009.05.030

McGuire, K. P. (2016). Breast Anatomy and Physiology. In A. Aydiner, A. Igci, & A. Soran (Eds.), *Breast Disease* (2nd ed., pp. 1–14). Cham, Switzerland. (Section: Chapter 1) doi: 10.1007/978-3-319-22843-3_1

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... Karthikesalingam, A. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*. Retrieved from http://dx.doi.org/10.1038/s41586-019-1799-6 doi: 10.1038/s41586-019-1799-6

Mehta, P., Sandfort, V., Gheysens, D., Braeckevelt, G.-J., Berte, J., & Summers, R. M. (2019). Segmenting The Kidney On CT Scans Via Crowdsourcing. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (pp. 829–832). Venice, Italy: IEEE. Retrieved 2024-01-04, from https://ieeexplore.ieee.org/document/8759240/ doi: 10.1109/ISBI.2019.8759240

Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J. G., & Shapiro, L. (2018). Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018* (pp. 1–8). Retrieved from http://link.springer.com/10.1007/978-3-319-46723-8 doi: 10.1007/978-3-319-46723-8

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image Segmentation Using Deep Learning: A Survey. *IEEE Trans Pattern Anal Mach Intell*, *44*(7), 3523–3542. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33596172 (Edition: 20220603) doi: 10.1109/TPAMI.2021.3059968

Mingers, J. (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, *12*(3), 240–259. Retrieved 2023-06-11, from https://www.jstor.org/stable/23011015 (Publisher: INFORMS)

Minssen, L., Dao, T. H., Quang, A. V., Martin, L., Andureau, E., Luciani, A., ... Deux, J. F. (2022). Breast arterial calcifications on mammography: a new marker of cardiovascular

risk in asymptomatic middle age women? *Eur Radiol*. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/35147775 (Edition: 20220211) doi: 10.1007/s00330-022-08571-3

Mobini, N., Codari, M., Riva, F., Ienco, M. G., Capra, D., Cozzi, A., ... Sardanelli, F. (2023). Detection and quantification of breast arterial calcifications on mammograms: a deep learning approach. *European Radiology*. Retrieved 2023-05-14, from https://link.springer.com/10.1007/s00330-023-09668-z doi: 10.1007/s00330-023-09668-z

Mohamed, A. A., Berg, W. A., Peng, H., Luo, Y., Jankowitz, R. C., & Wu, S. (2018). A deep learning method for classifying mammographic breast density categories. *Medical Physics*, *45*(1), 314–321. doi: 10.1002/mp.12683

Molloi, S., Mehraien, T., Iribarren, C., Smith, C., Ducote, J. L., & Feig, S. A. (2009). Reproducibility of Breast Arterial Calcium Mass Quantification Using Digital Mammography. *Academic Radiology*, *16*(3), 275–282. doi: 10.1016/j.acra.2008.08.011

Molloi, S., Xu, T., Ducote, J., & Iribarren, C. (2008). Quantification of breast arterial calcification using full field digital mammography. *Medical Physics*, *35*(4), 1428–1439. doi: 10.1118/1.2868756

Montagnon, E., Cerny, M., Cadrin-Chênevert, A., Hamilton, V., Derennes, T., Ilinca, A., ... Tang, A. (2020). Deep learning workflow in radiology: a primer. *Insights into Imaging*, *11*(1). doi: 10.1186/s13244-019-0832-5

Montgomery, G. H., Schnur, J. B., Erblich, J., Narula, J., Benck, K., & Margolies, L. (2022). Breast arterial calcification rates in a diverse, urban group of screening mammography patients. *Ann Epidemiol*, *75*, 16-20. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/36031094 doi: 10.1016/j.annepidem.2022.08.040

Moradi, M., Adibi, A., & Abedi, M. (2014). Relationship between breast arterial calcification on mammography with CT Calcium scoring and coronary CT angiography results. *Advanced Biomedical Research*, *3*, 79. doi: 10.4103/2277-9175.127992

Mordang, J.-J. (2018). *Towards an independent observer of screening mammograms: detection of calcifications* (Unpublished doctoral dissertation). Radboud University.

Mordang, J.-J., Gubern-Mérida, A., den Heeten, G., & Karssemeijer, N. (2016). Reducing false positives of microcalcification detection systems by removal of breast arterial calcifications. *Medical Physics*, *43*(4), 1676–1687. doi: 10.1118/1.4943376

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast: toward a full-field digital mammographic database. *Acad Radiol*, *19*(2), 236–48. (Edition: 2011/11/15) doi: 10.1016/j.acra.2011.09.014

Mostafavi, L., Marfori, W., Arellano, C., Tognolini, A., Speier, W., Adibi, A., & Ruehm, S. G. (2015). Prevalence of coronary artery disease assessed by dual source CT angiography in women with mammographically detected breast arterial calcifications. *PLoS ONE*, *10*(4), 1–12. doi: 10.5061/dryad.pr11m

Mukhlif, A. A., Al-Khateeb, B., & Mohammed, M. A. (2022). An extensive review of state-of-the-art transfer learning techniques used in medical imaging: Open issues and challenges. *Journal of Intelligent Systems*, *31*(1), 1085–1111. Retrieved 2023-09-01, from https://www.degruyter.com/document/doi/10.1515/jisys-2022-0198/html doi: 10.1515/jisys-2022-0198

Namora, & Jan Everhard Riwurohi. (2022, April). Prediction of Water Levels on Peatland using Deep Learning. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *6*(2), 234–239. Retrieved 2024-11-02, from https://www.jurnal.iaii.or.id/index.php/RESTI/article/view/3919 doi: 10.29207/resti.v6i2.3919

Nasir, K., & McEvoy, J. W. (2016). Recognizing Breast Arterial Calcification as Atherosclerotic CVD Risk Equivalent from Evidence to Action. *JACC: Cardiovascular Imaging*, *9*(4), 361–363. doi: 10.1016/j.jcmg.2015.09.017

Neumann, F.-J., Sousa-Uva, M., Ahlsson, A., Alfonso, F., Banning, A. P., Benedetto, U., ... Henderson, R. (2018). 2018 ESC/EACTS Guidelines on myocardial revascularization. *European Heart Journal*, 87–165. Retrieved from https://academic.oup.com/eurheartj/advance-article/doi/10.1093/eurheartj/ehy394/5079120 doi: 10.1093/eurheartj/ehy394

Neves, P. O., Andrade, J., & Monção, H. (2017). Coronary artery calcium score: current status. *Radiologia Brasileira*, *50*(3), 182–189. Retrieved from

http://www.scielo.br/scielo.php?script=sci{_}arttext{&}pid=S0100-39842017000300182{&}lng=en{&}tlng=en doi: 10.1590/0100-3984.2015.0235

Niakšu, O. (2015). CRISP Data Mining Methodology Extension for Medical Domain.

Nishikawa, R. M. (2010). Computer-aided Detection and Diagnosis. In U. Bick & F. Diekmann (Eds.), *Digital mammography* (pp. 85–106). Heidelberg: Springer Verlag.

Noh, H., Hong, S., & Han, B. (2015). Learning Deconvolution Network for Semantic Segmentation. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 1520–1528). Santiago, Chile: IEEE. Retrieved 2024-02-18, from http://ieeexplore.ieee.org/document/7410535/ doi: 10.1109/ICCV.2015.178

Nudy, M., Asmaro, R., Jiang, X., & Schnatz, P. F. (2022). The association between incidentally found breast arterial calcification on routine screening mammography and the development of coronary artery disease and stroke: results of a 10-year prospective study. *Menopause*, *29*(12), 1375-1380. Retrieved from https://journals.lww.com/menopausejournal/Fulltext/2022/12000/The_association_between_incidentally_found_breast.6.aspx doi: 10.1097/gme.0000000000002088

Oates, B. J., Griffiths, M., & McLean, R. (2022). *Researching Information Systems and Computing*. London: SAGE Publications Ltd.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. Retrieved 2024-02-24, from http://ieeexplore.ieee.org/document/4310076/ doi: 10.1109/TSMC.1979.4310076

Pan, I., Chu, A. J., Wang, Y., Merck, D., & Lourenco, A. (2018). Predicting Breast Nodule Malignancy with Efficient Convolutional Neural Networks. In *2018 SIIM Conference on Machine Intelligence in Medical Imaging* (pp. 1–3). Retrieved from https://cdn.ymaws.com/siim.org/resource/resmgr/mimi18/presentations/18cmimi_ps3-chu.pdf&https://cdn.ymaws.com/siim.org/resource/resmgr/mimi18/abstracts/18paper3-Pan.pdf

Petersen, S. E., Sanghvi, M. M., Aung, N., Cooper, J. A., Paiva, J. M., Zemrak, F., ...

Neubauer, S. (2017). The impact of cardiovascular risk factors on cardiac structure and function: Insights from the UK Biobank imaging enhancement study. *PLoS ONE*, *12*(10), 1–14. doi: 10.1371/journal.pone.0185114

Pianykh, O. S. (2008). *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide*. Berlin: Springer.

Pisano, E. D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J. K., Acharyya, S., ... Rebner, M. (2005). Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *New England Journal of Medicine*, *353*(17), 1773–1783. doi: 10.1056/NEJMoa052911

Polonsky, T. S., & Greenland, P. (2017). Breast arterial calcification: Expanding the reach of cardiovascular prevention. *Circulation*, *135*(6), 499-501. doi: 10.1161/CIRCULATIONAHA.116.025277

Reddy, J., Son, H., Smith, S. J., Paultre, F., & Mosca, L. (2005). Prevalence of breast arterial calcifications in an ethnically diverse population of women. *Annals of Epidemiology*, *15*(5), 344-350. doi: 10.1016/j.annepidem.2004.11.006

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788). Las Vegas, NV, USA: IEEE. Retrieved 2024-01-28, from http://ieeexplore.ieee.org/document/7780460/ doi: 10.1109/CVPR.2016.91

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6517–6525). Honolulu, HI: IEEE. Retrieved 2024-01-28, from http://ieeexplore.ieee.org/document/8100173/ doi: 10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement.* arXiv. Retrieved 2024-01-28, from http://arxiv.org/abs/1804.02767 (arXiv:1804.02767 [cs])

Reiazi, R., Paydar, R., Ardakani, A. A., & Etedadialiabadi, M. (2018). Mammography Lesion Detection Using Faster R-CNN Detector. In (pp. 111–115). doi: 10.5121/csit.2018.80212

Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). *Real-Time Flying Object Detection with YOLOv8.* arXiv. Retrieved 2024-04-02, from http://arxiv.org/abs/2305.09972 (arXiv:2305.09972 [cs])

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234-241. doi: 10.1007/978-3-319-24574-4_28

Royal College of Radiologists. (2017). *Guidance on the use of patient images obtained as part of standard care for teaching, training and research*. London: Royal College of Radiologists.

Royal College of Radiologists. (2019). *Picture archiving and communication systems (PACS) and guidelines on diagnostic display devices.* (3rd ed.). London: Royal College of Radiologists.

Russell, S., & Norvig, P. (2013). *Artificial Intelligence: A Modern Approach*. doi: 10.1017/S0269888900007724

Ružičić, D., Dobrić, M., Vuković, M., Hrnčić, D., Đorđević, S., Ružičić, M., ... Beleslin, B. (2018a). The correlation of SYNTAX score by coronary angiography with breast arterial calcification by digital mammography. *Clinical Radiology*, *73*(5), 454–459. doi: 10.1016/j.crad.2017.12.002

Ružičić, D., Dobrić, M., Vuković, M., Hrnčić, D., Đorđević, S., Ružičić, M., ... Beleslin, B. (2018b). Novel assessment tool for coronary artery disease severity during screening mammography. *Health Care for Women International*(May), 1–15. doi: 10.1080/07399332.2018.1463226

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 526–534. Retrieved 2024-11-02, from https://linkinghub.elsevier.com/retrieve/pii/

S1877050921002416 doi: 10.1016/j.procs.2021.01.199

Sechopoulos, I., & Sá dos Reis, C. (2022). Mammography Equipment. In C. Mercer, P. Hogg, & J. Kelly (Eds.), *Digital Mammography: A Holistic Approach* (2nd ed., pp. 199–216). Cham, Switzerland: Springer Nature Switzerland.

Seeram, E. (2019). Full-Field Digital Mammography. In *Digital Radiography* (pp. 111–123). Singapore: Springer Singapore. Retrieved 2023-09-10, from http://link.springer.com/10.1007/978-981-13-3244-9_7 doi: 10.1007/978-981-13-3244-9_7

Selvikvåg Lundervold, A., & Lundervold, A. (2018). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift fur Medizinische Physik*(Mmiv). Retrieved from https://doi.org/10.1016/j.zemedi.2018.11.002 doi: 10.1016/j.zemedi.2018.11.002

Shah, N., Chainani, V., Delafontaine, P., Abdo, A., Lafferty, J., & Rafeh, N. A. (2014). Mammographically detectable breast arterial calcification and atherosclerosis. *Cardiology in Review*, *22*(2), 69–78. doi: 10.1097/CRD.0b013e318295e029

Shams, S., Platania, R., Zhang, J., Kim, J., & Lee, K. (2018). Deep Generative Breast Cancer Screening and Diagnosis. In *Medical image computing and computer-assisted intervention – miccai 2018* (Vol. 8150). Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-642-40763-5 doi: 10.1007/978-3-642-40763-5

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. , 1–14. Retrieved from http://arxiv.org/abs/1409.1556 doi: 10.1016/j.infsof.2008.09.005

Singh, V. K., Romani, S., Rashwan, H. A., Akram, F., Pandey, N., Sarker, M. M. K., ... Puig, D. (2018). Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11071 LNCS*, 833–840. doi: 10.1007/978-3-030-00934-2_92

Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., & Klang, E. (2019). Convolutional neural networks for radiologic images: A radiologist's guide. *Radiology*,

290(3), 590-606. doi: 10.1148/radiol.2018180547

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929-1958. doi: 10.1214/12-AOS1000

Su, R., Zhang, D., Liu, J., & Cheng, C. (2021). MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation. *Front Genet*, *12*, 639930. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33679900 doi: 10.3389/fgene.2021.639930

Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological Physics and Technology*, *10*(3), 257–273. doi: 10.1007/s12194-017-0406-5

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). Retrieved 2024-02-18, from https://ojs.aaai.org/index.php/AAAI/article/view/11231 doi: 10.1609/aaai.v31i1.11231

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 1–9. doi: 10.1109/CVPR.2015.7298594

Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*. Cham: Springer International Publishing. Retrieved 2023-08-01, from https://link.springer.com/10.1007/978-3-030-34372-9 doi: 10.1007/978-3-030-34372-9

Taylor, P., Champness, J., Given-Wilson, R., Johnston, K., & Potts, H. (2005). Impact of computer-aided detection prompts on screening mammography. *Health Technology Assessment*, *9*(6). doi: 10.3310/hta9060

ter Haar Romeny, B. M. (2019). A Deeper Understanding of Deep Learning. In E. R. Ranschaert, S. Morozov, & P. R. Algra (Eds.), *Artificial intelligence in medical imaging: Opportunities, applications and risks* (pp. 25–38). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-94878-2{_}3 doi: 10.1007/978-3-319-94878-2_3

Terven, J., & Cordova-Esparza, D. (2023). A Comprehensive Review of YOLO Architectures

in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, *5*(4), 1680–1716. Retrieved 2024-04-02, from http://arxiv.org/abs/2304.00501 (arXiv:2304.00501 [cs]) doi: 10.3390/make5040083

Trimboli, R. M., Capra, D., Codari, M., Cozzi, A., Di Leo, G., & Sardanelli, F. (2021). Breast arterial calcifications as a biomarker of cardiovascular risk: radiologists' awareness, reporting, and action. A survey among the EUSOBI members. *European Radiology*, *31*(2), 958–966. Retrieved 2023-06-27, from https://link.springer.com/10.1007/s00330-020-07136-6 doi: 10.1007/s00330-020-07136-6

Wang, J., Ding, H., Bidgoli, F. A., Zhou, B., Iribarren, C., Molloi, S., & Baldi, P. (2017). Detecting cardiovascular disease from mammograms with deep learning. *IEEE Transactions on Medical Imaging*, *36*(5), 1172-1181. doi: 10.1109/TMI.2017.2655486

Wang, K., Hill, M., Knowles-Barley, S., Tikhonov, A., Litchfield, L., & Bare, J. C. (2023). Improving Segmentation of Breast Arterial Calcifications from Digital Mammography: Good Annotation is All You Need. In Y. Zheng, H. Y. Keleş, & P. Koniusz (Eds.), *Computer Vision – ACCV 2022 Workshops* (Vol. 13848, pp. 134–150). Cham: Springer Nature Switzerland. Retrieved 2023-09-03, from https://link.springer.com/10.1007/978-3-031-27066-6_10 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-031-27066-6_10

Wang, K., Khan, N., & Highnam, R. (2019). Automated Segmentation of Breast Arterial Calcifications from Digital Mammography. In *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)* (pp. 1–6). doi: 10.1109/IVCNZ48456.2019.8960956

Warren Burhenne, L. J., Wood, S. A., D'Orsi, C. J., Feig, S. A., Kopans, D. B., O'Shaughnessy, K. F., ... Castellino, R. A. (2000). Potential Contribution of Computer-aided Detection to the Sensitivity of Screening Mammography. *Radiology*, *215*(2), 554–562. Retrieved from https://doi.org/10.1148/radiology.215.2.r00ma15554 doi: 10.1148/radiology.215.2.r00ma15554

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., ... Lungren, M. P. (2020). Preparing Medical Imaging Data for Machine Learning. *Radiology*,

295(1), 4–15. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/32068507 (Edition: 2020/02/19) doi: 10.1148/radiol.2020192224

Winsberg, F., Elkin, M., Macy, J., Bordaz, V., & Weymouth, W. (1967). Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. *Radiology*, *89*(2), 211–215. Retrieved from https://doi.org/10.1148/89.2.211 doi: 10.1148/89.2.211

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 4, pp. 29–39).

Wooldridge, M. (2021). *The Road to Conscious Machines: The Story of AI*. London: Pelican.

World Heart Foundation. (2017). *Risk Factors.* Retrieved 2019-01-06, from https://www.world-heart-federation.org/resources/risk-factors/

Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., ... Geras, K. J. (2019). Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Transactions on Medical Imaging*, 1-1. doi: 10.1109/tmi.2019.2945514

Yaffe, M. J. (2009). Basic Physics of Digital Mammography. In U. Bick & F. Diekmann (Eds.), *Digital Mammography.* Berlin, Heidelberg, GERMANY: Springer Berlin / Heidelberg. Retrieved 2023-09-10, from http://ebookcentral.proquest.com/lib/salford/detail.action?docID=690948

Yi, P. H., Lin, A., Wei, J., Sair, H. I., Hui, F. K., & Harvey, S. C. (2018). Deep-Learning Based Semantic Labeling for 2D Mammography & Comparison of Computational Complexity for Machine Learning Tasks. In *2018 siim conference on machine intelligence in medical imaging* (pp. 1–2). Retrieved from https://cdn.ymaws.com/siim.org/resource/resmgr/mimi18/abstracts/18paper1-Yi.pdf

Yoon, Y. E., Kim, K. M., Han, J. S., Kang, S.-H., Chun, E. J., Ahn, S., ... Suh, J.-W. (2018). Prediction of Subclinical Coronary Artery Disease With Breast Arterial Calcification and Low Bone Mass in Asymptomatic Women: Registry for the Women Health Cohort for Breast, Bone, and Coronary Artery Disease Study. *JACC: Cardiovascular Imaging*. doi: https://doi.org/10.1016/j.jcmg.2018.07.004

Zeiser, F. A., da Costa, C. A., Zonta, T., Marques, N. M., Roehe, A. V., Moreno, M., & da Rosa Righi, R. (2020). Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning. *Journal of Digital Imaging*, 858–868. doi: 10.1007/s10278-020-00330-4

A | Ethical Approval

Research, Enterprise and Engagement
Ethical Approval Panel

Doctoral & Research Support
Research and Knowledge Exchange,
Room 827, Maxwell Building,
University of Salford,
Manchester
M5 4WT

T +44(0)161 295 2280

www.salford.ac.uk

29 January 2019

Dear Dominic,

RE: ETHICS APPLICATION–HSR1819-034 – 'The application of machine learning and computer-aided detection (CADe) to identify women at risk of cardiovascular disease from breast-screening mammography.'

Based on the information that you have provided, I am pleased to inform you that ethics application HSR1819-034 has been approved.

If there are any changes to the project and/or its methodology, then please inform the Panel as soon as possible by contacting Health-ResearchEthics@salford.ac.uk

Yours sincerely,

Professor Sue McAndrew
Chair of the Research Ethics Panel

Amendment Notification Form

Title of Project:		
The application of machine learning and computer-aided detection to identify women at risk of cardiovascular disease from breast screening mammography.		
Name of Lead Applicant:	**School:**	
Dominic Maguire	Health & Society	
Are you the original Principal Investigator (PI) for this study?		No
If you have selected 'NO', please explain why you are applying for the amendment: *I am a part-time PhD student based in Ireland. Recent circumstances relating to the pandemic do not allow me to travel to Salford. My original plan was to have two observer studies on campus. I wish to amend my ethics application to hold these locally within my hospital group.*		
Date original approval obtained:	**Reference No:**	**Externally funded project?**
14/01/2019	**HSR1819-034**	No

Please outline the proposed changes to the project. *NB. If the changes require any amendments to the PIS, Consent Form(s) or recruitment material, then please submit these with this form highlighting where the changes have been made:*

- **Recruitment of participants to include clinical radiology colleagues in my hospital group.**
  - **Please see attached my ethics application, page 13.**
- **Location of observer studies to be at a suitable, local viewing facility.**
  - **Please see attached the PIS and Consent Forms, page 1.**

Please say whether the proposed changes present any new ethical issues or changes to ethical issues that were identified in the original ethics review, and provide details of how these will be addressed:

- No new ethical issues anticipated.

Amendment Approved:	**YES**	***Date of Approval:***	**17/08/2021**

Chair's Signature:

Once completed you should submit this form and any additional documentation to the RKE Ethics Team at ethics@salford.ac.uk

B | Dataset Licence Agreement

THIS AGREEMENT is made the 26th November **2018**

BETWEEN:

(1) **CANCER RESEARCH TECHNOLOGY LIMITED**, a company registered in England and Wales under number 1626049 with registered office at Angel Building, 407 St John Street, London, EC1V 4AD, England ("**CRT**"); and

(2) **UNIVERSITY OF SALFORD**, a charitable body incorporated by Royal Charter with registration no. RC000666 whose administrative office is at The Crescent, Salford M5 4WT, England ("**Institution**"); and

(3) **ROYAL SURREY COUNTY HOSPITAL**, having its principal administrative offices at Egerton Road, Guildford, Surrey, GU2 7XX, UK ("**RSCH**").

RECITALS

(A) CRT is an oncology focused technology transfer and development company.

(B) CRT has taken assignment of, and possesses the rights in the Source Database (as defined below) which is currently held and managed by the Medical Physics Department at RSCH.

(C) Institution is a place of academic teaching and learning which also conducts research in a variety of disciplines.

(D) Institution wishes to take a licence to access the Database Images (as defined below) and the Related Data (as defined below) to conduct the Study (as defined below) as a step towards achieving the goal stated in Recital C.

(E) CRT agrees to license the Database Images and Related Data for use in the Study on the following terms and conditions.

OPERATIVE PROVISIONS

NAME: DOMINIC MAGUIRE
SIGNATURE: Dominic Maguire
DATE: 06/06/19

C | Annotation Validation Reader Study Forms

University of Salford MANCHESTER

Appendix [2]

CONSENT FORM

Title of study: The application of machine learning and computer-aided detection (CADe) to identify women at risk of cardiovascular disease from breast-screening mammography.

Name of Researcher: Dominic Maguire

Please complete and sign this form **after** you have read and understood the study information sheet. Read the following statements, and select 'Yes' or 'No' in the box on the right hand side.

1.	I confirm that I have read and understand the study information sheet version 3.0, dated 08/08/2021, for the above study. I have had the opportunity to consider the information and to ask questions Which have been answered satisfactorily.	Yes/No
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my rights being affected.	Yes/No
3.	If I do decide to withdraw I understand that the information I have given, up to the point of withdrawal, will be used in the research.	Yes/No
4.	I agree to participate by annotating anonymised mammogram images from the OPTIMAM dataset for the existence and extent of breast arterial calcification.	Yes/No
5.	I understand that my personal details will be kept confidential and will not be revealed to people outside the research.	Yes/No
6.	I understand that my anonymised data will be used in the researcher's thesis, other academic publications, conference papers and presentations.	Yes/No
7.	I agree to take part in the study:	Yes/No

AISLINN O'REILLY	09/12/21	M. O'Reilly.
Name of participant	**Date**	**Signature**
DOMINIC MAGUIRE	09/12/21	Dominic Maguire
Name of person taking consent	**Date**	**Signature**

Version [3] Date [08/08/2021]

University of Salford MANCHESTER

Appendix [2]

CONSENT FORM

Title of study: The application of machine learning and computer-aided detection (CADe) to identify women at risk of cardiovascular disease from breast-screening mammography.

Name of Researcher: Dominic Maguire

Please complete and sign this form **after** you have read and understood the study information sheet. Read the following statements, and select 'Yes' or 'No' in the box on the right hand side.

1.	I confirm that I have read and understand the study information sheet version 3.0, dated 08/08/2021, for the above study. I have had the opportunity to consider the information and to ask questions Which have been answered satisfactorily.	Yes/No
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my rights being affected.	Yes/No
3.	If I do decide to withdraw I understand that the information I have given, up to the point of withdrawal, will be used in the research.	Yes/No
4.	I agree to participate by annotating anonymised mammogram images from the OPTIMAM dataset for the existence and extent of breast arterial calcification.	Yes/No
5.	I understand that my personal details will be kept confidential and will not be revealed to people outside the research.	Yes/No
6.	I understand that my anonymised data will be used in the researcher's thesis, other academic publications, conference papers and presentations.	Yes/No
7.	I agree to take part in the study:	Yes/No

John Hyland	19/1/22	
Name of participant	Date	Signature
DOMINIC MAGUIRE	19/01/22	
Name of person taking consent	Date	Signature

Version [3] Date [08/08/2021]

University of Salford MANCHESTER

Appendix [2]

PARTICIPANT INFORMATION SHEET

Title of study: The application of machine learning and computer-aided detection (CADe) to identify women at risk of cardiovascular disease from breast-screening mammography.

Name of Researcher: xxx

1. Invitation paragraph

I would like to invite you to take part in a research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read is not clear or you would like more information. Take time to decide whether or not to take part.

2. What is the purpose of the study?

The study aims to develop a software tool to automatically identify breast arterial calcification on screening mammograms and to grade the level of calcification present in order to predict the risk of cardiovascular disease in women. The research is part of a PhD in machine learning I am doing at the University of Salford.

3. Why have I been invited to take part?

Your clinical experience and skills in radiology/mammography reporting would be useful to the project as part of a human reader study for annotating images on the existence and extent of breast arterial calcification.

4. Do I have to take part?

It is up to you to decide. We will describe the study and go through the information sheet which we will give to you. We will then ask you to sign a consent form to show that you have agreed to take part. You are free to withdraw at any time, without giving a reason.

5. What will happen to me if I take part?

You will be asked to annotate a number of anonymised cases (n ≈ 60) from the OPTIMAM mammography dataset for the existence and extent of breast arterial calcification. These will be used to provide ground-truth information that will allow me to train machine learning algorithms. Each session will last around an hour and will take place in a suitable, local viewing facility.

6. Expenses and payments?

You will receive no compensation/payment for your participation in the research study.

7. What are the possible disadvantages and risks of taking part?

There is no anticipated risk associated with participating in the study although the work could involve a considerable time commitment. The risk of Covid will be mitigated as much as possible.

8. What are the possible benefits of taking part?

The information we get from the study will help progress automatic identification of patients at risk from cardiovascular disease, the number one cause of premature death in women.

9. What if there is a problem?

If you have a concern about any aspect of this study, I can be contacted at **xxx**.
If you remain unhappy and wish to complain formally you can do this by contacting my supervisors via email at **xxx** and **xxx**.

However, if you remain dissatisfied and wish to complain formally, please forward your concerns to Professor Andrew Clark, Chair of the Health Research Ethical Approval Panel, Allerton Building, Frederick Road Campus, University of Salford, Salford, M6 6PU. Tel: 0161 295 4109. E: a.clark@salford.ac.uk

10. Will my taking part in the study be kept confidential?

All information which is collected from you during the course of the research will be anonymised. Your data will be stored safely and used only for the purposes of the study. Your data will be accessible only by authorised persons such as researchers within the team and supervisors.

Depending on the OPTIMAM licence agreement, your data will be retained for a maximum of 3 years before being disposed of securely.

11. What will happen if I don't carry on with the study?

You are free to withdraw from the study at any time without notice or explanation. If you withdraw from the study we will use the data collected up to your withdrawal.

12. What will happen to the results of the research study?

Findings will be disseminated through journals and conference papers. You will not be identified in any report/publication unless you have given consent.

13. Who is organising or sponsoring the research?

University of Salford.

14. Further information and contact details:

Xxx

University of Salford MANCHESTER

Appendix [2]

CONSENT FORM

Title of study: The application of machine learning and computer-aided detection (CADe) to identify women at risk of cardiovascular disease from breast-screening mammography.

Name of Researcher: Dominic Maguire

Please complete and sign this form **after** you have read and understood the study information sheet. Read the following statements, and select 'Yes' or 'No' in the box on the right hand side.

1.	I confirm that I have read and understand the study information sheet version 3.0, dated 08/08/2021, for the above study. I have had the opportunity to consider the information and to ask questions Which have been answered satisfactorily.	Yes/No
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my rights being affected.	Yes/No
3.	If I do decide to withdraw I understand that the information I have given, up to the point of withdrawal, will be used in the research.	Yes/No
4.	I agree to participate by annotating anonymised mammogram images from the OPTIMAM dataset for the existence and extent of breast arterial calcification.	Yes/No
5.	I understand that my personal details will be kept confidential and will not be revealed to people outside the research.	Yes/No
6.	I understand that my anonymised data will be used in the researcher's thesis, other academic publications, conference papers and presentations.	Yes/No
7.	I agree to take part in the study:	Yes/No

_____	_____	_____
Name of participant	**Date**	**Signature**

_____	_____	_____
Name of person taking consent	**Date**	**Signature**

Risk Assessment Form

ALL projects MUST include a risk assessment. If this summary assessment of the risk proves insignificant, i.e. you answer 'no' to all of the questions, then no further action is necessary. However, if you identify any risks then you must identify the precautions you will put in place to control these.

1. What is the title of the project?

The application of machine learning and computer-aided detection (CADe) to identify women at risk of cardiovascular disease from breast-screening mammography.

2. Is the project purely literature based? NO

If YES, please go to the bottom of the assessment and sign where indicated. If NO, then please complete section 3 and list your proposed controls.

3. Please highlight the risk(s) which applies to your study:

Hazards	Risks	If yes, consider what precautions will be taken to minimise risk and discuss with your Supervisor
Use of ionising or non-ionising radiation	*Exposure to radiation* ***NO***	
Use of hazardous substances	*Exposure to harmful substances* ***NO***	
Use of face-to-face interviews *Interviewees could be upset by interview and become aggressive or violent toward researcher*	*Interviewing …* ***NO***	
Use of face-to-face interviews *Participants or interviewees could become upset by*	***NO***	

interview and suffer psychological effects		
Sensitive data	*Exposure to data or information which may cause upset or distress to the researcher* ***NO***	
Physical activity	*Exposure to levels of exertion unsuitable for an individual's level of fitness* ***NO***	
Equipment	*Exposure to faulty or unfamiliar equipment.* ***NO***	
Sensitive issues i.e. Gender/Cultural e.g. when observing or dealing with undressed members of the opposite sex	*Exposure to vulnerable situations/ sensitive issues that may cause distress to interviewer or interviewee* ***NO***	
Children	***NO***	
Manual handling activities	*Exposure to an activity that could result in injury* ***NO***	

If you have answered 'YES' to any of the hazards in section 3, then please list the proposed precautions below:

Signature of student …….. Date ……..

Signature of Supervisor …… Date ……

D | MedXViewer Data Output

```
{
  "session_id": "122",
  "set_data": {
    "Group 1": {
      "caseList": [
        "demd126320",
        "demd120047",
        "demd113666",
        "demd140609",
        "demd71486",
        "demd140972",
        "demd37105",
        "demd139399",
        "demd125989",
        "demd139860"
      ],
      "caseDataMap": {
        "demd126320": {
          "roiMap": {
            "1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.222.0": {
              "1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.241.0": [
                {
                  "coords": [
                    10.125760649087223,
                    13.501014198782963,
                    700.3651115618662,
                    1604.9330628803248
                  ],
                  "type": "java.awt.geom.Rectangle2D$Double",
                  "unixTime": 1642508798,
                  "sliceNo": 0.0
                }
              ],
              "1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.238.0": [
                {
                  "coords": [
                    11.813387423935092,
                    401.65517241379314,
                    722.3042596348885,
                    826.9371196754564
                  ],
                  "type": "java.awt.geom.Rectangle2D$Double",
                  "unixTime": 1642508774,
                  "sliceNo": 0.0
                }
              ],
              "1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.225.0": [
                {
                  "coords": [
                    1576.2434077079108,
                    20.251521298174445,
                    725.6795131845843,
                    1631.9350912778907
                  ],
                  "type": "java.awt.geom.Rectangle2D$Double",
                  "unixTime": 1642508718,
                  "sliceNo": 0.0
                }
              ],
              "1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.235.0": [
                {
                  "coords": [
                    2026.8397565922924,
                    671.6754563894524,
                    528.2271805273834,
```

```
                    528.2271805273834,
                    997.3874239350914
                  ],
                  "type": "java.awt.geom.Rectangle2D$Double",
                  "unixTime": 1642508734,
                  "sliceNo": 0.0
                }
              ]
            }
          },
          "caseAnswers": {
            "childrenQuestion": [
              {
                "q": {
                  "questionID": 1.0,
                  "questionString": "Is breast arterial calcification present in this case?",
                  "questionType": "org.nccpm.medXViewer.question.model.CheckBoxListQuestionModel",
                  "questionChoices": [
                    {
                      "allowedChoices": [],
                      "string": "Yes"
                    },
                    {
                      "allowedChoices": [],
                      "string": "No"
                    },
                    {
                      "allowedChoices": [],
                      "string": "Indeterminate"
                    }
                  ],
                  "questionAnswer": [
                    "Yes"
                  ]
```

E | Annotation IoU Results

Case	Image	IoU
'demd126320'	'1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.241.0.dcm'	0.839895
'demd126320'	'1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.238.0.dcm'	0.768616
'demd126320'	'1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.225.0.dcm'	0.556538
'demd126320'	'1.2.826.0.1.3680043.9.3218.1.1.41220200.1549.1546015312104.235.0.dcm'	0.633927
'demd139399'	'1.2.826.0.1.3680043.9.3218.1.1.14623017.5998.1547650561117.332.0.dcm'	0.388307
'demd139399'	'1.2.826.0.1.3680043.9.3218.1.1.14623017.5998.1547650561117.343.0.dcm'	0.114656
'demd139399'	'1.2.826.0.1.3680043.9.3218.1.1.14623017.5998.1547650561117.347.0.dcm'	0.279519
'demd139399'	'1.2.826.0.1.3680043.9.3218.1.1.14623017.5998.1547650561117.345.0.dcm'	0
'demd125989'	'1.2.826.0.1.3680043.9.3218.1.1.40564071.1740.1545949699164.190.0.dcm'	0.155205
'demd125989'	'1.2.826.0.1.3680043.9.3218.1.1.40564071.1740.1545949699164.186.0.dcm'	0.129342
'demd125989'	'1.2.826.0.1.3680043.9.3218.1.1.40564071.1740.1545949699164.196.0.dcm'	0.093888
'demd125989'	'1.2.826.0.1.3680043.9.3218.1.1.40564071.1740.1545949699164.196.0.dcm'	0.081708
'demd125989'	'1.2.826.0.1.3680043.9.3218.1.1.40564071.1740.1545949699164.184.0.dcm'	0.196424
'demd113666'	'1.2.826.0.1.3680043.9.3218.1.1.13306350.3444.1543223927131.320.0.dcm'	0.087163
'demd113666'	'1.2.826.0.1.3680043.9.3218.1.1.13306350.3444.1543223927131.324.0.dcm'	0
'demd113666'	'1.2.826.0.1.3680043.9.3218.1.1.13306350.3444.1543223927131.324.0.dcm'	0.468282
'demd113666'	'1.2.826.0.1.3680043.9.3218.1.1.13306350.3444.1543223927131.314.0.dcm'	0.717132
'demd113666'	'1.2.826.0.1.3680043.9.3218.1.1.13306350.3444.1543223927131.327.0.dcm'	0.749743
'demd120047'	'1.2.826.0.1.3680043.9.3218.1.1.22181111.5677.1548406370479.215.0.dcm'	0.492763
'demd120047'	'1.2.826.0.1.3680043.9.3218.1.1.22181111.5677.1548406370479.225.0.dcm'	0.757246
'demd120047'	'1.2.826.0.1.3680043.9.3218.1.1.22181111.5677.1548406370479.211.0.dcm'	0.883592
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.1916.1529288190893.16.0.dcm'	0.085768
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.1916.1529288190893.12.0.dcm'	0.289758
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.1916.1529288190893.12.0.dcm'	0
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.1916.1529288190893.12.0.dcm'	0
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.1916.1529288190893.12.0.dcm'	0.838449
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.19166.1529288190893.8.0.dcm'	0.055771
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.19166.1529288190893.8.0.dcm'	0
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.19166.1529288190893.4.0.dcm'	0
'demd9697'	'1.2.826.0.1.3680043.9.3218.1.1.279800749.19166.1529288190893.4.0.dcm'	0.398554
'demd57707'	'1.2.826.0.1.3680043.9.3218.1.1.83447471.1466.1537681772207.199.0.dcm'	0.440928
'demd57707'	'1.2.826.0.1.3680043.9.3218.1.1.83447471.1466.1537681772207.197.0.dcm'	0.622183
'demd57707'	'1.2.826.0.1.3680043.9.3218.1.1.83447471.1466.1537681772207.201.0.dcm'	0.769106
'demd57707'	'1.2.826.0.1.3680043.9.3218.1.1.83447471.1466.1537681772207.205.0.dcm'	0.348069
'demd57707'	'1.2.826.0.1.3680043.9.3218.1.1.83447471.1466.1537681772207.205.0.dcm'	0.374961
'demd127928'	'1.2.826.0.1.3680043.9.3218.1.1.181979402.1279.1546370238730.99.0.dcm'	0.693953
'demd127928'	'1.2.826.0.1.3680043.9.3218.1.1.181979402.1279.1546370238730.85.0.dcm'	0.964305
'demd134625'	'1.2.826.0.1.3680043.9.3218.1.1.97068472.2953.1547158944053.315.0.dcm'	0.558332
'demd134625'	'1.2.826.0.1.3680043.9.3218.1.1.97068472.2953.1547158944053.306.0.dcm'	0.804708
'demd134625'	'1.2.826.0.1.3680043.9.3218.1.1.97068472.2953.1547158944053.297.0.dcm'	0.564337
'demd134625'	'1.2.826.0.1.3680043.9.3218.1.1.97068472.2953.1547158944053.312.0.dcm'	0.799325
'demd103365'	'1.2.826.0.1.3680043.9.3218.1.1.72939361.9942.1541966231393.377.0.dcm'	0.537556
'demd103365'	'1.2.826.0.1.3680043.9.3218.1.1.72939361.9942.1541966231393.386.0.dcm'	0.446475
'demd103365'	'1.2.826.0.1.3680043.9.3218.1.1.72939361.9942.1541966231393.386.0.dcm'	0.016643
'demd103365'	'1.2.826.0.1.3680043.9.3218.1.1.72939361.9942.1541966231393.389.0.dcm'	0.677619
'demd103365'	'1.2.826.0.1.3680043.9.3218.1.1.72939361.9942.1541966231393.389.0.dcm'	0
'demd103365'	'1.2.826.0.1.3680043.9.3218.1.1.72939361.9942.1541966231393.380.0.dcm'	0.743044

F | ResNet-22 BAC Classification Network

	Layer Name	Layer Type	Description
1	'Image_input_1'	Image Input	2898×2360×1 images with 'rescale-symmetric' normalization
2	'Conv_0'	2-D Convolution	16 7×7×1 convolutions with stride [2 2] and padding [0 0 0 0]
3	'MaxPool_1'	2-D Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
4	'BatchNormalization_2'	Batch Normalization	Batch normalization with 16 channels
5	'Relu_3'	ReLU	ReLU
6	'Conv_5'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
7	'BatchNormalization_6'	Batch Normalization	Batch normalization with 16 channels
8	'Relu_7'	ReLU	ReLU
9	'Conv_8'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
10	'Conv_4'	2-D Convolution	16 1×1×16 convolutions with stride [1 1] and padding [0 0 0 0]
11	'Add_9'	Addition	Element-wise addition of 2 inputs
12	'BatchNormalization_10'	Batch Normalization	Batch normalization with 16 channels
13	'Relu_11'	ReLU	ReLU
14	'Conv_12'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
15	'BatchNormalization_13'	Batch Normalization	Batch normalization with 16 channels
16	'Relu_14'	ReLU	ReLU
17	'Conv_15'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
18	'Add_16'	Addition	Element-wise addition of 2 inputs
19	'BatchNormalization_17'	Batch Normalization	Batch normalization with 16 channels
20	'Relu_18'	ReLU	ReLU
21	'Conv_20'	2-D Convolution	32 3×3×16 convolutions with stride [2 2] and padding [1 1 1 1]

22	'BatchNormalization_21'	Batch Normalization	Batch normalization with 32 channels
23	'Relu_22'	ReLU	ReLU
24	'Conv_23'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
25	'Conv_19'	2-D Convolution	32 1×1×16 convolutions with stride [2 2] and padding [0 0 0 0]
26	'Add_24'	Addition	Element-wise addition of 2 inputs
27	'BatchNormalization_25'	Batch Normalization	Batch normalization with 32 channels
28	'Relu_26'	ReLU	ReLU
29	'Conv_27'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
30	'BatchNormalization_28'	Batch Normalization	Batch normalization with 32 channels
31	'Relu_29'	ReLU	ReLU
32	'Conv_30'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
33	'Add_31'	Addition	Element-wise addition of 2 inputs
34	'BatchNormalization_32'	Batch Normalization	Batch normalization with 32 channels
35	'Relu_33'	ReLU	ReLU
36	'Conv_34'	2-D Convolution	64 1×1×32 convolutions with stride [2 2] and padding [0 0 0 0]
37	'Conv_35'	2-D Convolution	64 3×3×32 convolutions with stride [2 2] and padding [1 1 1 1]
38	'BatchNormalization_36'	Batch Normalization	Batch normalization with 64 channels
39	'Relu_37'	ReLU	ReLU
40	'Conv_38'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
41	'Add_39'	Addition	Element-wise addition of 2 inputs
42	'BatchNormalization_40'	Batch Normalization	Batch normalization with 64 channels
43	'Relu_41'	ReLU	ReLU

44	'Conv_42'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
45	'BatchNormalization_43'	Batch Normalization	Batch normalization with 64 channels
46	'Relu_44'	ReLU	ReLU
47	'Conv_45'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
48	'Add_46'	Addition	Element-wise addition of 2 inputs
49	'BatchNormalization_47'	Batch Normalization	Batch normalization with 64 channels
50	'Relu_48'	ReLU	ReLU
51	'Conv_50'	2-D Convolution	128 3×3×64 convolutions with stride [2 2] and padding [1 1 1 1]
52	'BatchNormalization_51'	Batch Normalization	Batch normalization with 128 channels
53	'Relu_52'	ReLU	ReLU
54	'Conv_53'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
55	'Conv_49'	2-D Convolution	128 1×1×64 convolutions with stride [2 2] and padding [0 0 0 0]
56	'Add_54'	Addition	Element-wise addition of 2 inputs
57	'BatchNormalization_55'	Batch Normalization	Batch normalization with 128 channels
58	'Relu_56'	ReLU	ReLU
59	'Conv_57'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
60	'BatchNormalization_58'	Batch Normalization	Batch normalization with 128 channels
61	'Relu_59'	ReLU	ReLU
62	'Conv_60'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
63	'Add_61'	Addition	Element-wise addition of 2 inputs
64	'BatchNormalization_62'	Batch Normalization	Batch normalization with 128 channels
65	'Relu_63'	ReLU	ReLU

66	'Conv_64'	2-D Convolution	256 1×1×128 convolutions with stride [2 2] and padding [0 0 0 0]
67	'Conv_65'	2-D Convolution	256 3×3×128 convolutions with stride [2 2] and padding [1 1 1 1]
68	'BatchNormalization_66'	Batch Normalization	Batch normalization with 256 channels
69	'Relu_67'	ReLU	ReLU
70	'Conv_68'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
71	'Add_69'	Addition	Element-wise addition of 2 inputs
72	'BatchNormalization_70'	Batch Normalization	Batch normalization with 256 channels
73	'Relu_71'	ReLU	ReLU
74	'Conv_72'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
75	'BatchNormalization_73'	Batch Normalization	Batch normalization with 256 channels
76	'Relu_74'	ReLU	ReLU
77	'Conv_75'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
78	'Add_76'	Addition	Element-wise addition of 2 inputs
79	'BatchNormalization_77'	Batch Normalization	Batch normalization with 256 channels
80	'Relu_78'	ReLU	ReLU
81	'avgpool2d'	2-D Average Pooling	5×5 average pooling with stride [1 1] and padding 'same'
82	'dropout_1'	Dropout	50% dropout
83	'fc_1'	Fully Connected	256 fully connected layer
84	'dropout_2'	Dropout	50% dropout
85	'fc_2'	Fully Connected	2 fully connected layer
86	'softmax'	Softmax	softmax
87	'classoutput'	Classification Output	crossentropyex

G | Faster R-CNN Object Detection Network with ResNet-22 Feature Extraction Module

	Layer Name	Layer Type	Layer Description
1	'Image_input_1'	Image Input	2898×2360×1 images with 'zerocenter' normalization
2	'Conv_0'	2-D Convolution	16 7×7×1 convolutions with stride [2 2] and padding [0 0 0 0]
3	'MaxPool_1'	2-D Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
4	'BatchNormalization_2'	Batch Normalization	Batch normalization with 16 channels
5	'Relu_3'	ReLU	ReLU
6	'Conv_5'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
7	'BatchNormalization_6'	Batch Normalization	Batch normalization with 16 channels
8	'Relu_7'	ReLU	ReLU
9	'Conv_8'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
10	'Conv_4'	2-D Convolution	16 1×1×16 convolutions with stride [1 1] and padding [0 0 0 0]
11	'Add_9'	Addition	Element-wise addition of 2 inputs
12	'BatchNormalization_10'	Batch Normalization	Batch normalization with 16 channels
13	'Relu_11'	ReLU	ReLU
14	'Conv_12'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
15	'BatchNormalization_13'	Batch Normalization	Batch normalization with 16 channels
16	'Relu_14'	ReLU	ReLU
17	'Conv_15'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
18	'Add_16'	Addition	Element-wise addition of 2 inputs
19	'BatchNormalization_17'	Batch Normalization	Batch normalization with 16 channels
20	'Relu_18'	ReLU	ReLU
21	'Conv_20'	2-D Convolution	32 3×3×16 convolutions with stride [2 2] and padding [1 1 1 1]

22	'Conv_19'	2-D Convolution	32 1×1×16 convolutions with stride [2 2] and padding [0 0 0 0]
23	'BatchNormalization_21'	Batch Normalization	Batch normalization with 32 channels
24	'Relu_22'	ReLU	ReLU
25	'Conv_23'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
26	'Add_24'	Addition	Element-wise addition of 2 inputs
27	'BatchNormalization_25'	Batch Normalization	Batch normalization with 32 channels
28	'Relu_26'	ReLU	ReLU
29	'Conv_27'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
30	'BatchNormalization_28'	Batch Normalization	Batch normalization with 32 channels
31	'Relu_29'	ReLU	ReLU
32	'Conv_30'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
33	'Add_31'	Addition	Element-wise addition of 2 inputs
34	'BatchNormalization_32'	Batch Normalization	Batch normalization with 32 channels
35	'Relu_33'	ReLU	ReLU
36	'Conv_35'	2-D Convolution	64 3×3×32 convolutions with stride [2 2] and padding [1 1 1 1]
37	'BatchNormalization_36'	Batch Normalization	Batch normalization with 64 channels
38	'Relu_37'	ReLU	ReLU
39	'Conv_38'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
40	'Conv_34'	2-D Convolution	64 1×1×32 convolutions with stride [2 2] and padding [0 0 0 0]
41	'Add_39'	Addition	Element-wise addition of 2 inputs
42	'BatchNormalization_40'	Batch Normalization	Batch normalization with 64 channels
43	'Relu_41'	ReLU	ReLU

44	'Conv_42'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
45	'BatchNormalization_43'	Batch Normalization	Batch normalization with 64 channels
46	'Relu_44'	ReLU	ReLU
47	'Conv_45'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
48	'Add_46'	Addition	Element-wise addition of 2 inputs
49	'BatchNormalization_47'	Batch Normalization	Batch normalization with 64 channels
50	'Relu_48'	ReLU	ReLU
51	'Conv_49'	2-D Convolution	128 1×1×64 convolutions with stride [2 2] and padding [0 0 0 0]
52	'Conv_50'	2-D Convolution	128 3×3×64 convolutions with stride [2 2] and padding [1 1 1 1]
53	'BatchNormalization_51'	Batch Normalization	Batch normalization with 128 channels
54	'Relu_52'	ReLU	ReLU
55	'Conv_53'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
56	'Add_54'	Addition	Element-wise addition of 2 inputs
57	'BatchNormalization_55'	Batch Normalization	Batch normalization with 128 channels
58	'Relu_56'	ReLU	ReLU
59	'Conv_57'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
60	'BatchNormalization_58'	Batch Normalization	Batch normalization with 128 channels
61	'Relu_59'	ReLU	ReLU
62	'Conv_60'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
63	'Add_61'	Addition	Element-wise addition of 2 inputs
64	'BatchNormalization_62'	Batch Normalization	Batch normalization with 128 channels
65	'Relu_63'	ReLU	ReLU

66	'rpnConv3x3'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
67	'rpnRelu'	ReLU	ReLU
68	'rpnConv1x1BoxDeltas'	2-D Convolution	24 1×1×128 convolutions with stride [1 1] and padding [0 0 0 0]
69	'rpnBoxDeltas'	Box Regression Output	smooth-l1 loss
70	'rpnConv1x1ClsScores'	2-D Convolution	12 1×1×128 convolutions with stride [1 1] and padding [0 0 0 0]
71	'rpnSoftmax'	RPN Softmax	rpn softmax
72	'rpnClassification'	RPN Classification Output	cross-entropy loss with 'object' and 'background' classes
73	'regionProposal'	Region Proposal	region proposal with 6 anchor boxes
74	'roiPooling'	ROI Max Pooling	ROI Max Pooling with pooled output size [91 74]
75	'Conv_64'	2-D Convolution	256 1×1×128 convolutions with stride [2 2] and padding [0 0 0 0]
76	'Conv_65'	2-D Convolution	256 3×3×128 convolutions with stride [2 2] and padding [1 1 1 1]
77	'BatchNormalization_66'	Batch Normalization	Batch normalization with 256 channels
78	'Relu_67'	ReLU	ReLU
79	'Conv_68'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
80	'Add_69'	Addition	Element-wise addition of 2 inputs
81	'BatchNormalization_70'	Batch Normalization	Batch normalization with 256 channels
82	'Relu_71'	ReLU	ReLU
83	'Conv_72'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
84	'BatchNormalization_73'	Batch Normalization	Batch normalization with 256 channels
85	'Relu_74'	ReLU	ReLU
86	'Conv_75'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
87	'Add_76'	Addition	Element-wise addition of 2 inputs

88	'BatchNormalization_77'	Batch Normalization	Batch normalization with 256 channels
89	'Relu_78'	ReLU	ReLU
90	'avgpool2d'	2-D Average Pooling	5×5 average pooling with stride [1 1] and padding 'same'
91	'fc_1'	Fully Connected	256 fully connected layer
92	'rcnnFC'	Fully Connected	2 fully connected layer
93	'rcnnSoftmax'	Softmax	softmax
94	'rcnnClassification'	Classification Output	crossentropyex with classes 'BAC' and 'Background'
95	'fcBoxDeltas'	Fully Connected	4 fully connected layer
96	'boxDeltas'	Box Regression Output	smooth-l1 loss

H | YOLOv4 Object Detection Network with ResNet-22 Feature Extraction Module

	Layer Name	Layer Type	Description
1	'image_input.1'	Image Input	2898×2360×1 images
2	'Conv_0'	2-D Convolution	16 7×7×1 convolutions with stride [2 2] and padding [0 0 0 0]
3	'MaxPool_1'	2-D Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
4	'BatchNormalization_2'	Batch Normalization	Batch normalization with 16 channels
5	'Relu_3'	ReLU	ReLU
6	'Conv_5'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
7	'BatchNormalization_6'	Batch Normalization	Batch normalization with 16 channels
8	'Relu_7'	ReLU	ReLU
9	'Conv_8'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
10	'Conv_4'	2-D Convolution	16 1×1×16 convolutions with stride [1 1] and padding [0 0 0 0]
11	'Add_9'	Addition	Element-wise addition of 2 inputs
12	'BatchNormalization_10'	Batch Normalization	Batch normalization with 16 channels
13	'Relu_11'	ReLU	ReLU
14	'Conv_12'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
15	'BatchNormalization_13'	Batch Normalization	Batch normalization with 16 channels
16	'Relu_14'	ReLU	ReLU
17	'Conv_15'	2-D Convolution	16 3×3×16 convolutions with stride [1 1] and padding [1 1 1 1]
18	'Add_16'	Addition	Element-wise addition of 2 inputs
19	'BatchNormalization_17'	Batch Normalization	Batch normalization with 16 channels
20	'Relu_18'	ReLU	ReLU
21	'Conv_20'	2-D Convolution	32 3×3×16 convolutions with stride [2 2] and padding [1 1 1 1]

22	'BatchNormalization_21'	Batch Normalization	Batch normalization with 32 channels
23	'Relu_22'	ReLU	ReLU
24	'Conv_23'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
25	'Conv_19'	2-D Convolution	32 1×1×16 convolutions with stride [2 2] and padding [0 0 0 0]
26	'Add_24'	Addition	Element-wise addition of 2 inputs
27	'BatchNormalization_25'	Batch Normalization	Batch normalization with 32 channels
28	'Relu_26'	ReLU	ReLU
29	'Conv_27'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
30	'BatchNormalization_28'	Batch Normalization	Batch normalization with 32 channels
31	'Relu_29'	ReLU	ReLU
32	'Conv_30'	2-D Convolution	32 3×3×32 convolutions with stride [1 1] and padding [1 1 1 1]
33	'Add_31'	Addition	Element-wise addition of 2 inputs
34	'BatchNormalization_32'	Batch Normalization	Batch normalization with 32 channels
35	'Relu_33'	ReLU	ReLU
36	'Conv_34'	2-D Convolution	64 1×1×32 convolutions with stride [2 2] and padding [0 0 0 0]
37	'Conv_35'	2-D Convolution	64 3×3×32 convolutions with stride [2 2] and padding [1 1 1 1]
38	'BatchNormalization_36'	Batch Normalization	Batch normalization with 64 channels
39	'Relu_37'	ReLU	ReLU
40	'Conv_38'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
41	'Add_39'	Addition	Element-wise addition of 2 inputs
42	'BatchNormalization_40'	Batch Normalization	Batch normalization with 64 channels
43	'Relu_41'	ReLU	ReLU

44	'Conv_42'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
45	'BatchNormalization_43'	Batch Normalization	Batch normalization with 64 channels
46	'Relu_44'	ReLU	ReLU
47	'Conv_45'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
48	'Add_46'	Addition	Element-wise addition of 2 inputs
49	'BatchNormalization_47'	Batch Normalization	Batch normalization with 64 channels
50	'Relu_48'	ReLU	ReLU
51	'Conv_50'	2-D Convolution	128 3×3×64 convolutions with stride [2 2] and padding [1 1 1 1]
52	'BatchNormalization_51'	Batch Normalization	Batch normalization with 128 channels
53	'Relu_52'	ReLU	ReLU
54	'Conv_53'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
55	'Conv_49'	2-D Convolution	128 1×1×64 convolutions with stride [2 2] and padding [0 0 0 0]
56	'Add_54'	Addition	Element-wise addition of 2 inputs
57	'BatchNormalization_55'	Batch Normalization	Batch normalization with 128 channels
58	'Relu_56'	ReLU	ReLU
59	'Conv_57'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
60	'BatchNormalization_58'	Batch Normalization	Batch normalization with 128 channels
61	'Relu_59'	ReLU	ReLU
62	'Conv_60'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
63	'Add_61'	Addition	Element-wise addition of 2 inputs
64	'BatchNormalization_62'	Batch Normalization	Batch normalization with 128 channels
65	'Relu_63'	ReLU	ReLU

66	'featureConvInitial1'	2-D Convolution	64 1×1×128 convolutions with stride [1 1] and padding 'same'
67	'featureBatchNormInitial1'	Batch Normalization	Batch normalization with 64 channels
68	'featureReluInitial1'	Leaky ReLU	Leaky ReLU with scale 0.1
69	'sppMaxPool1'	2-D Max Pooling	5×5 max pooling with stride [1 1] and padding 'same'
70	'sppMaxPool2'	2-D Max Pooling	9×9 max pooling with stride [1 1] and padding 'same'
71	'sppMaxPool3'	2-D Max Pooling	13×13 max pooling with stride [1 1] and padding 'same'
72	'depthConcat_spp_1'	Depth concatenation	Depth concatenation of 4 inputs
73	'featureConvSPP1'	2-D Convolution	64 1×1×256 convolutions with stride [1 1] and padding 'same'
74	'featureBatchNormSPP1'	Batch Normalization	Batch normalization with 64 channels
75	'featureReluSPP1'	Leaky ReLU	Leaky ReLU with scale 0.1
76	'depthConcat1'	Depth concatenation	Depth concatenation of 2 inputs
77	'featureConv_1_1'	2-D Convolution	64 1×1×96 convolutions with stride [1 1] and padding 'same'
78	'featureBatchNorm_1_1'	Batch Normalization	Batch normalization with 64 channels
79	'featureRelu_1_1'	Leaky ReLU	Leaky ReLU with scale 0.1
80	'customConv1'	2-D Convolution	128 3×3×64 convolutions with stride [1 1] and padding 'same'
81	'customBatchNorm1'	Batch Normalization	Batch normalization with 128 channels
82	'customRelu1'	Leaky ReLU	Leaky ReLU with scale 0.1
83	'customOutputConv1'	2-D Convolution	18 1×1×128 convolutions with stride [1 1] and padding 'same'
84	'featureConvInitial2'	2-D Convolution	32 1×1×64 convolutions with stride [1 1] and padding 'same'
85	'featureBatchNormInitial2'	Batch Normalization	Batch normalization with 32 channels
86	'featureReluInitial2'	Leaky ReLU	Leaky ReLU with scale 0.1
87	'featureConv_1_2'	2-D Convolution	32 1×1×64 convolutions with stride [1 1] and padding 'same'

88	'featureBatchNorm_1_2'	Batch Normalization	Batch normalization with 32 channels
89	'featureRelu_1_2'	Leaky ReLU	Leaky ReLU with scale 0.1
90	'featureResize2'	Resize	nnet.cnn.layer.Resize2DLayer
91	'depthConcat2'	Depth concatenation	Depth concatenation of 2 inputs
92	'featureConv_2_2'	2-D Convolution	32 1×1×64 convolutions with stride [1 1] and padding 'same'
93	'featureBatchNorm_2_2'	Batch Normalization	Batch normalization with 32 channels
94	'featureRelu_2_2'	Leaky ReLU	Leaky ReLU with scale 0.1
95	'customConv2'	2-D Convolution	64 3×3×32 convolutions with stride [1 1] and padding 'same'
96	'customBatchNorm2'	Batch Normalization	Batch normalization with 64 channels
97	'customRelu2'	Leaky ReLU	Leaky ReLU with scale 0.1
98	'customOutputConv2'	2-D Convolution	18 1×1×64 convolutions with stride [1 1] and padding 'same'
99	'featureConv_bottom_2'	2-D Convolution	32 1×1×32 convolutions with stride [2 2] and padding 'same'
100	'featureBatchNorm_bottom_2'	Batch Normalization	Batch normalization with 32 channels
101	'featureRelu_bottom_2'	Leaky ReLU	Leaky ReLU with scale 0.1

I DeepLabv3+ Segmentation Network with ResNet-18 Feature Extraction Module

	Layer Name	Layer Type	Description
1	'Image_input_1'	Image Input	512×512×3 images with 'rescale-symmetric' normalization
2	'conv1'	2-D Convolution	64 7×7×3 convolutions with stride [2 2] and padding [3 3 3 3]
3	'bn_conv1'	Batch Normalization	Batch normalization with 64 channels
4	'conv1_relu'	ReLU	ReLU
5	'pool1'	2-D Max Pooling	3×3 max pooling with stride [2 2] and padding [1 1 1 1]
6	'res2a_branch2a'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
7	'bn2a_branch2a'	Batch Normalization	Batch normalization with 64 channels
8	'res2a_branch2a_relu'	ReLU	ReLU
9	'res2a_branch2b'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
10	'bn2a_branch2b'	Batch Normalization	Batch normalization with 64 channels
11	'res2a'	Addition	Element-wise addition of 2 inputs
12	'res2a_relu'	ReLU	ReLU
13	'res2b_branch2a'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
14	'bn2b_branch2a'	Batch Normalization	Batch normalization with 64 channels
15	'res2b_branch2a_relu'	ReLU	ReLU
16	'res2b_branch2b'	2-D Convolution	64 3×3×64 convolutions with stride [1 1] and padding [1 1 1 1]
17	'bn2b_branch2b'	Batch Normalization	Batch normalization with 64 channels
18	'res2b'	Addition	Element-wise addition of 2 inputs
19	'res2b_relu'	ReLU	ReLU
20	'res3a_branch2a'	2-D Convolution	128 3×3×64 convolutions with stride [2 2] and padding [1 1 1 1]
21	'bn3a_branch2a'	Batch Normalization	Batch normalization with 128 channels

22	'res3a_branch2a_relu'	ReLU	ReLU
23	'res3a_branch2b'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
24	'bn3a_branch2b'	Batch Normalization	Batch normalization with 128 channels
25	'res3a'	Addition	Element-wise addition of 2 inputs
26	'res3a_relu'	ReLU	ReLU
27	'res3a_branch1'	2-D Convolution	128 1×1×64 convolutions with stride [2 2] and padding [0 0 0 0]
28	'bn3a_branch1'	Batch Normalization	Batch normalization with 128 channels
29	'res3b_branch2a'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
30	'bn3b_branch2a'	Batch Normalization	Batch normalization with 128 channels
31	'res3b_branch2a_relu'	ReLU	ReLU
32	'res3b_branch2b'	2-D Convolution	128 3×3×128 convolutions with stride [1 1] and padding [1 1 1 1]
33	'bn3b_branch2b'	Batch Normalization	Batch normalization with 128 channels
34	'res3b'	Addition	Element-wise addition of 2 inputs
35	'res3b_relu'	ReLU	ReLU
36	'res4a_branch2a'	2-D Convolution	256 3×3×128 convolutions with stride [2 2] and padding 'same'
37	'bn4a_branch2a'	Batch Normalization	Batch normalization with 256 channels
38	'res4a_branch2a_relu'	ReLU	ReLU
39	'res4a_branch2b'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
40	'bn4a_branch2b'	Batch Normalization	Batch normalization with 256 channels
41	'res4a'	Addition	Element-wise addition of 2 inputs
42	'res4a_relu'	ReLU	ReLU
43	'res4a_branch1'	2-D Convolution	256 1×1×128 convolutions with stride [2 2] and padding 'same'

44	'bn4a_branch1'	Batch Normalization	Batch normalization with 256 channels
45	'res4b_branch2a'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
46	'bn4b_branch2a'	Batch Normalization	Batch normalization with 256 channels
47	'res4b_branch2a_relu'	ReLU	ReLU
48	'res4b_branch2b'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]
49	'bn4b_branch2b'	Batch Normalization	Batch normalization with 256 channels
50	'res4b'	Addition	Element-wise addition of 2 inputs
51	'res4b_relu'	ReLU	ReLU
52	'res5a_branch2a'	2-D Convolution	512 3×3×256 convolutions with stride [1 1] and padding 'same'
53	'bn5a_branch2a'	Batch Normalization	Batch normalization with 512 channels
54	'res5a_branch2a_relu'	ReLU	ReLU
55	'res5a_branch2b'	2-D Convolution	512 3×3×512 convolutions with stride [1 1], dilation factor [2 2], and padding 'same'
56	'bn5a_branch2b'	Batch Normalization	Batch normalization with 512 channels
57	'res5a'	Addition	Element-wise addition of 2 inputs
58	'res5a_relu'	ReLU	ReLU
59	'res5a_branch1'	2-D Convolution	512 1×1×256 convolutions with stride [1 1] and padding [0 0 0 0]
60	'bn5a_branch1'	Batch Normalization	Batch normalization with 512 channels
61	'res5b_branch2a'	2-D Convolution	512 3×3×512 convolutions with stride [1 1], dilation factor [2 2], and padding 'same'
62	'bn5b_branch2a'	Batch Normalization	Batch normalization with 512 channels
63	'res5b_branch2a_relu'	ReLU	ReLU
64	'res5b_branch2b'	2-D Convolution	512 3×3×512 convolutions with stride [1 1], dilation factor [2 2], and padding 'same'
65	'bn5b_branch2b'	Batch Normalization	Batch normalization with 512 channels

66	'res5b'	Addition	Element-wise addition of 2 inputs
67	'res5b_relu'	ReLU	ReLU
68	'catAspp'	Depth concatenation	Depth concatenation of 4 inputs
69	'aspp_Conv_1'	2-D Convolution	256 1×1×512 convolutions with stride [1 1] and padding 'same'
70	'aspp_BatchNorm_1'	Batch Normalization	Batch normalization with 256 channels
71	'aspp_Relu_1'	ReLU	ReLU
72	'aspp_Conv_2'	2-D Convolution	256 3×3×512 convolutions with stride [1 1], dilation factor [6 6], and padding 'same'
73	'aspp_BatchNorm_2'	Batch Normalization	Batch normalization with 256 channels
74	'aspp_Relu_2'	ReLU	ReLU
75	'aspp_Conv_3'	2-D Convolution	256 3×3×512 convolutions with stride [1 1], dilation factor [12 12], and padding 'same'
76	'aspp_BatchNorm_3'	Batch Normalization	Batch normalization with 256 channels
77	'aspp_Relu_3'	ReLU	ReLU
78	'aspp_Conv_4'	2-D Convolution	256 3×3×512 convolutions with stride [1 1], dilation factor [18 18], and padding 'same'
79	'aspp_BatchNorm_4'	Batch Normalization	Batch normalization with 256 channels
80	'aspp_Relu_4'	ReLU	ReLU
81	'dec_c1'	2-D Convolution	256 1×1×1024 convolutions with stride [1 1] and padding [0 0 0 0]
82	'dec_bn1'	Batch Normalization	Batch normalization with 256 channels
83	'dec_relu1'	ReLU	ReLU
84	'dec_upsample1'	2-D Transposed Convolution	256 8×8×256 transposed convolutions with stride [4 4] and cropping [2 2 2 2]
85	'dec_crop1'	Crop 2D	center crop
86	'dec_c2'	2-D Convolution	48 1×1×64 convolutions with stride [1 1] and padding [0 0 0 0]
87	'dec_bn2'	Batch Normalization	Batch normalization with 48 channels

88	'dec_relu2'	ReLU	ReLU
89	'dec_cat1'	Depth concatenation	Depth concatenation of 2 inputs
90	'dec_c3'	2-D Convolution	256 3×3×304 convolutions with stride [1 1] and padding 'same'
91	'dec_bn3'	Batch Normalization	Batch normalization with 256 channels
92	'dec_relu3'	ReLU	ReLU
93	'dec_c4'	2-D Convolution	256 3×3×256 convolutions with stride [1 1] and padding 'same'
94	'dec_bn4'	Batch Normalization	Batch normalization with 256 channels
95	'dec_relu4'	ReLU	ReLU
96	'scorer'	2-D Convolution	3 1×1×256 convolutions with stride [1 1] and padding [0 0 0 0]
97	'dec_upsample2'	2-D Transposed Convolution	3 8×8×3 transposed convolutions with stride [4 4] and cropping [2 2 2 2]
98	'dec_crop2'	Crop 2D	center crop
99	'softmax-out'	Softmax	softmax
100	'Seg-Layer'	Pixel Classification Layer	Class weighted cross-entropy loss with 'background', 'BAC', and 1 other classes