Coventry University

Faculty of Engineering, Environment and Computing
School of Computing, Electronics and Maths

Breast Cancer Diagnosis on Screening Mammography via Deep Learning Classification

Antonio Mpembe Franco

(SID:) 9870776

Supervisors: Reda Al Bodour, Carey Pridgeon

Submitted in partial fulfilment of the requirements for the Degree of Master of Science in:
Data Science and Computational Intelligence

Fri, 27th Aug 2021

Declaration of Originality

I declare that this project is all my work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet, etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists. I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

Statement of Copyright

Statement of Ethical Engagement

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website https://ethics.coventry.ac.uk/ and that the application number is **P122889**

Name: Antonio Mpembe Franco

Signature: 

Date: Fri, 27th Aug 2021

Abstract

This research investigates the application of deep learning models for breast cancer diagnosis from mammography images, focusing on the challenges posed by limited and complex datasets. The primary objective was to develop an end-to-end pipeline for tumor detection and classification by leveraging state-of-the-art architectures for computer-aided diagnosis Computer-aided detection/diagnosis (CAD) in breast cancer mammography by investigating the state-of-the-art while taking into consideration of medical data scarcity issues and model robustness to environmental changes by having not only test but different data sources to simulate how the model would work in a change of environment. The proposed methodology centered on a teacher-student framework, where a teacher model and student model which are both (EfficientNet-U-Net) using a novel semi-supervised learning approach Meta Pseudo Labels, with a data pipeline designed to leverage a small set of labeled images alongside a larger pool of unlabeled, using labels and unlabeled data. This to determine if it could enhance the model performance and mitigate database bias by data scarcity for tumor segmentation.

The project's execution revealed a critical dependency: the success of the Meta Pseudo Labels framework is predicated on the student model's ability to first achieve a reasonable baseline performance on the labeled data alone. In this study, the student model failed to converge even on the supervised segmentation task, achieving a Dice coefficient near zero. indicating no meaningful overlap between the predicted masks and the ground-truth masks. The model was effectively producing random noise or empty masks, when it came to the segmentation section. Consequently, this thesis concludes that while Meta Pseudo Labels presents a theoretically powerful solution to data scarcity, its practical application requires a foundational, supervised model that can first learn a basic representation of the task. The study highlights the profound impact of data limitations and recommends that future work focus on establishing a robust baseline model through techniques such as aggressive augmentation, transfer learning from medical domains.

Due to the challenges in achieving anything in segmentation, an alternative approach of direct image classification using a standalone EfficientNet-B0 was also investigated. While the classification model demonstrated high efficacy on a standard benchmark dataset (Oxford Flowers, achieving 100% accuracy), dataset used for sanity purposes, both the segmentation and classification models failed to converge on the available mammography data. The reasons pointing data complexity, scarcity, imbalance, and lack of computational power and storage. It concludes with a comprehensive set of recommendations for future work, including advanced data augmentation techniques, the use of transfer learning from domain-specific pretrained models, computational power and storage.

Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms

Acknowledgement

Throughout the writing, coding, and comprehension of this dissertation, I got a significant amount of help and support.

I'd want to use this page to thank everyone of my teachers, family, and friends who have helped me along the way.

I'd want to express my gratitude to everyone who has assisted me along the journey, as well as the open source communities. This work would be more difficult or impossible to do without them, notably Github, Cornell University arXiv, and Wikipedia. I'd also like to thank all the instructors that taught me everything and gave me the knowledge that helped me to get to where I am now. Without their encouragement, support, and guidance, none of this would be possible.

Introduction

1.1 Background

Breast cancer is one of the leading causes of mortality in women worldwide, World Health Organisation (WHO) estimates that 23% of cancer-related cases and 14% of cancer-related deaths among women are due to breast cancer. It is one of the most prevalent cancers in the United Kingdom, most commonly diagnosed in women over the age of 50, although it can also strike younger women. Breast cancer affects about 1 in every 8 women at some point in their lives (NHS, 2019). According to Cancer Research UK, breast cancer is the second most frequent cause of cancer mortality in women in the United Kingdom, with around 11,500 fatalities in 2018. It accounts for 7% of all cancer fatalities. More than half of all breast cancer deaths in the United Kingdom (52 per cent) occur in women under the age of 75 each year (Cancer Research UK, 2021).

A mix of variables, according to studies, causes breast cancer. Being a woman and being older are two major variables that affect the risk. Some women get breast cancer despite having no other known risk factors (CDC, 2020). Unfortunately, there is no certain way to prevent breast cancer. There are precautionary steps that may reduce certain risk factors. But the two major risk factors are beyond people's control (American Cancer Society, 2021b).

Breast cancer is still one of the most commonly diagnosed malignancies in the world population. Because it is treatable in its early stages, early detection is the most effective way to reduce mortality, especially given that many women with breast cancer show no symptoms. Therefore, it is vital to get screened for breast cancer frequently from the age of 40. According to the American Cancer Society's breast cancer screening standards, getting either a 2D or 3D mammography is under current screening recommendations (American Cancer Society, 2021a). Mammography screening involves multiple processes, including the identification and analysis of abnormalities, such as masses and calcifications, used to determine the risk of breast cancer. In clinical settings, this analysis is mainly a manual procedure evaluated by a radiologist. Still, regardless of the radiologists' competence in evaluating mammograms, extrinsic variables such as weariness, distractions, and human error must be minimised, as the rate of missed diagnosis by radiologists is as high as 20%.

CAD technologies enable radiologists to make better judgments and this improves diagnosis accuracy. However, because of the low rate of precise performance, building a well-designed CAD system remains a tough task (Ajay and Saurabh, 2018). They typically consist of three steps: lesion detection, lesion segmentation, and lesion classification. The primary problems with these phases are the poor signal-to-noise ratio in lesion imag-

ing and the absence of a consistent lesion location, shape, and appearance. With these difficulties, lesion detection techniques produce a substantial number of false positives while missing a significant proportion of real positives (Oliver et al., 2010). Regardless of the lack of high accuracy or accuracy higher than in human diagnosis, CAD remains an important tool that eases the performance of individual mammogram readers with better precision, so that it may minimise the need for double reading. This will further minimise the costs and time involved in reporting diagnosis during mobile mammography drives.

Deep learning is expected to transform medical CAD and image analysis. Even though machine learning has been utilised in CAD and medical image analysis for over three decades, CAD has not been widely employed in the clinic because of the restricted performance of traditional machine learning techniques (Chan et al., 2020). CNN's recognition capacity in natural image recognition is already thought to be better than human recognition ability. Thus, it is expected to be an excellent discriminator in medical image recognition (Kido et al., 2020). Most studies on deep learning in medical imaging have shown very promising results that often exceeded clinicians' performance. The widespread use of new patients or clinical settings for these deep learning models is still unknown. A large enough training sample containing verified reference truth is one essential requirement for developing a robust machine learning algorithm (Chan et al., 2020). Collecting large labelled datasets for training, which can be time-consuming and expensive when done by an expert such as a doctor in medical applications.

In most data-poor tasks, the models are pre-trained and then fine-tuned to the given dataset problem. This works well when there is a high-level structural similarity to the large pre-trained dataset. They call this transfer learning information through a common set of parameters from one data set to another. However, pre-training has been found to be less useful when the domain of the fine-tuning task is significantly distinct from natural pictures, such as medical images (Raghu et al., 2019).

The accuracy and usefulness of the labels or annotations are important for achieving desired results, and it depends on the task. In 2017, the Digital Mammography DREAM Challenge aimed to develop a model that would help reduce breast cancer screening recall rates. Over 640,000 training mammograms were provided to participants, over 86,000 women. With the number of images, the winning team's false positive rate was higher than that of an experienced breast radiologist at comparable sensitivity even for the top deep learning model. This example shows that the lack of high-quality labelling in deep learning training may decrease its efficacy (Chan et al., 2020).

The Digital Mammogram DREAM Challenge exemplifies an exciting time in medical image processing history. The increasing availability of large annotated medical image datasets and deep learning advances have led to significant achievements in automated medical image understanding. This has allowed the growth in popularity over the last few decades of CAD in medical imaging, increasing its accuracy and allowing for more efficient image assessment and interpretation.

The rising availability of large annotated data is a great development in medical image recognition, but it is not as good as the quantity of labelled natural images and, logically, it may never consider the fact that they are more difficult to annotate and require an expert. Even if supervised learning normally archives better results, experts consider that the future of machine learning is not with supervised learning. As stated by Yann LeCun, "the next AI revolution will not be supervised or purely reinforced. The future is self-supervised learning with massive amounts of data and very large networks".

A vast number of recent successful machine learning applications are in a supervised learning scenario. But getting this labelled data is expensive. Annotated medical data is even more expensive to collect. Given promise of Deep learning-based Computer-Aided Diagnosis (CAD) systems to revolutionize this field by assisting radiologists. The data still seem to be one of the bottleneck in developing such systems, often leads to models that perform poorly and suffer from **database bias** or **domain shift**, the model trained on one dataset fails to generalize to another.

Semi-supervised learning can reduce the need for labelled data by using unlabelled data. It allows the model to learn the high-level data distribution structure and rely only on the labelled data to learn the small specifics of a specific task. This project employs a Meta Pseudo Labels approach with EfficientUnet for segmentation and EfficientNet for classification. With a suitable annotated mammogram dataset with ground truth of tumour mask and their diagnosis.

1.2 Problem Statement

The core problem this research addresses is the development of an automated system for both locating and classifying suspicious regions in mammograms. This is broken down into two primary tasks:

1. Segmentation: Accurately identifying and delineating the boundaries of a potential tumor within a mammogram.

2. Classification: Diagnosing the segmented region as either benign or malignant.

A significant challenge in this domain is the scarcity of large, well-annotated public datasets, which are essential for training high-capacity deep learning models. This project, therefore, also serves as an investigation into the feasibility of training state-of-the-art models a novel semi-supervised learning approach, Meta Pseudo Labels, on a dataset with appropriate ground truth. By using visualisation instruments, as well as the entire model concept process and the implementation and evaluation.

This can enable CAD system developers to understand if a semi-supervised learning approach is suitable for breast cancer diagnosis, in particular Meta Pseudo Labels. And exploring the database bias problem, which is directly related to domain shift, an issue which most machine learning algorithms encounter when deployed in the real world. This may be beneficial for the data science and computational intelligence community.

By seeking to answer the following question:

- Can Meta Pseudo Labels be a suitable approach for CAD in breast cancer diagnosis?
- Can Meta Pseudo Labels be a suitable approach against database bias for CAD in breast cancer diagnosis?

This project is an attempt to answer these questions, detailing the experimental design, the challenges encountered, and the critical lessons learned from the project's outcome.

1.3 Research Objectives

The primary objectives of this thesis are:

1. To design and implement a two-stage deep learning pipeline for breast cancer diagnosis, consisting of a segmentation model followed by a classification model.
2. To investigate the performance of an EfficientNet-B0-U-Net architecture for the tumor segmentation task in Meta Pseudo labels.
3. To evaluate the effectiveness of an EfficientNet-B0 model for both direct image classification and as a feature extractor for the segmentation model.
4. To rigorously analyze the impact of a limited and imbalanced dataset on model training and performance.
5. To identify the key challenges and limitations encountered and propose specific, actionable strategies for future research to overcome them.

1.4 Thesis Structure

This thesis is organized as follows:

Chapter 1 : Provides an overview of the subject's through the problem description and the motivation behind this project, followed by the project's objectives.

Chapter 2 : Is review of the relevant literature the background surrounding of breast cancer detection techniques with and without the deep learning and currently been used and the most used dataset. Also explains their past and possible future.

Chapter 3 : This part details the methodology the datasets, constrains and also shows the results of the procedure.

Chapter 4 : presents the experimental results and provides an in-depth discussion of the findings.

Chapter 5 : concludes the thesis and outlines directions for future work.

Literature Review

2.1 Breast Cancer Diagnosis

Ancient Egyptians described cases of breast cancer as "there is no treatment" for the disease (American Cancer Society, 2002). Mastectomy for breast cancer was first done in anno Domini (AD) 548 when the court physician Aetios of Amida suggested it to Theodora. In the 17th century, the French physician Jean Louis Petit conducted complete mastectomies, which included the removal of the axillary lymph nodes in the armpit (Olson and James, 2002). By the end of the 18th century, doctors were performing them more often (Winchester et al., 2006).

Long before breast cancer screening was part of the diagnosis process, treatment or attempt at the treatment of breast cancer took place. Mammography has its roots in Wilhelm Röntgen's discovery of x-rays in 1895 (Ingram, 2011). One of the first studies was in 1913 by Albert Salomon, a German surgeon, who compared X-rays of the breasts to the actual tissue removed. On the basis of an X-ray, he was able to distinguish between malignant and non-cancerous tumours in the breast. Robert Egan of the University of Texas created the first screening mammography method in the late 1950s (Medich and Martel, 2006). And it was in 1966 that the first large-scale randomised controlled study of mammography screening was conducted. The research, conducted by Philip Strax of New York, looked at the effect of mammograms on mortality and treatment.

Mammography examines the breast with low energy and high-intensity X-ray photons. The radiation is transmitted through the breast, where various tissues absorb varying amounts of radiation. This allows for the differentiation of various features and details associated with the tissues under examination (Yaffe and Mainprize, 2004).

The screening is intended to detect non-homogeneous and subtle breast cancer symptoms. Because of the marginal visual thresholds between the specific abnormality and the complex background tissues, these features are rarely visible. This could be due to the subtlety of the suspicious lesions, or they could be masked by homogeneous tissues of the breast parenchyma at times. It is important to note that mammograms only assist doctors in detecting breast cancer symptoms much earlier when it is treatable with a higher success rate (Tang et al., 2009). They cannot establish that a suspicious spot is cancerous, but they can assist healthcare practitioners in determining if more testing is required. The doctor will check for a variety of anomalies in the breast, including little white spots known as calcifications, bigger abnormal regions known as masses, and other suspicious locations that might be cancerous. If a lump isn't just a cyst, it's cause for concern, and a biopsy may be required to ensure it's not cancer.

Biopsy is the most frequent medical technique performed to diagnose cancer nowadays.

It provides information on the type of cancer, its classification, and the direction of spread, as well as additional information that can aid in treatment. The size of the primary tumour and whether or not it has spread determine how far a tumour has progressed in the body. Among all the diagnostic methods, it stands out. Still, CADs are important for helping experts. It is used as a tool that improves accuracy and reduces the workload.

Researchers began investigating the prospect of developing CAD systems in the late 1950s. Flow charts, statistical pattern matching, theory of probability, or knowledge bases have been utilised in the early CAD systems to drive the decision-making process (Yanase and Triantaphyllou, 2019). The attention focused on the use of data mining methods to utilise more powerful, flexible CAD systems in the late 1980s and the beginning of the 1990s. In 1998, the U.S. Food and Drug Administration (FDA) authorised the first commercial CAD system for mammography (Doi, 2007).

The current CAD system is essentially a complex computer vision-based system that recognises patterns in images. For the detection of suspicious structures in medical images, various types of medical images are used. They follow certain series of steps to prepare and analyse the images. Pre-processing, segmentation, feature extraction, and classification are some of them. The accuracy of this process is usually determined by the conditions of the medical images used to train the system, as well as other factors such as image quality, radiologist's marks, tumour size and location, and so on.

There is a growing interest in the development of deep learning-based mammogram analysis methodologies. There are two publicly available datasets, that do produce the best results in the field but during the project, it was not possible to get access to these datasets, "INBreast-database masses characterization" (Domingues et al., 2012) and "THE DIGITAL DATABASE FOR SCREENING MAMMOGRAPHY" (Heath et al., 2007). In deep learning normally there is 3 deep learning pipeline approach. Because the tumour region is important, these approaches involve detection R-CNN model, segmentation Unet or a combination of both applied on the tumour region then followed by classification. It is possible to have accurate breast mass segmentation results.

The future of the CAD system for breast cancer diagnostics appears to be supported by big companies research and application of deep learning, such as Google that recent study discovered that their system could improve breast cancer detection.

2.2 Efficientnet and U-net

There are a lot of machine learning models that are good at solving linear problems. But the assumption of linearity is very restrictive, since most real world problems are not linear. A simple way to increase the flexibility is to perform feature transformation. By transforming this $f(x;\theta) = Wx + b$ to $f(x;\theta,\theta') = W\phi(x;\theta') + b$. This process can be

repeated recursively with a L function, $f(x;\theta) = f_L(f_{L-1}(...(f1(x))...))$, to create more and more complex functions. This is the key idea behind deep neural networks (DNNs).

In 1967, Alexey Ivakhnenko and Lapa published the first general, working algorithm for supervised, deep, feedforward, and multilayer perceptron (MLP). By 1991, such systems were being used to recognise isolated 2-D handwritten digits. DNNs are a subset of a larger family of models in which are composed of differentiable functions into any type of directed acyclic graph (DAG), mapping input to output. This is because non-differentiable functions make it very difficult to train models with the lack of back-propagation.

For similar problems, like this one or problems related to computer vision, the most used artificial neural network (ANN) is CNN. This is because they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Unlike traditional algorithms, which require hand-engineering of filters (or kernels), the network learns to optimise the filters (or kernels) through automated learning. CNN is comprised of various types of layers, hyper-parameters, as shown in the Table 5, these are some of the building blocks. In most layers of CNN, the data is transformed which may or not result in a change of the shape structure. This process is normally executed with hyper-parameters and uses regulation methods to avoid overfitting.

Layer Type	Hyper-parameters	Shape	Regularization
Convolution	K, filter size	N, number of images	Empirical
Activation	F, number of filters	H, height	1. Dropout
Pooling	S, stride	W, width	Explicit
Fully connected	P, padding	C, number of channels	1. Early stopping

Table 1: *Example of CNN Building blocks*

In 2015, Olaf Ronneberger, Philipp Fischer, and Thomas Brox founded U-Net (Ronneberger et al., 2015). It's an improvement and development of Evan Shelhamer, Jonathan Long, and Trevor Darrell (2014) "Fully convolutional networks for semantic segmentation" (Long et al., 2014). It is a convolutional neural network developed at the University of Freiburg's Computer Science Department for biomedical image segmentation. The main idea is to supplement a usual contracting network with successive layers, where pooling operations are replaced by upsampling operators. The network has a u-shaped architecture as shown in the Figure 14, because it has a contracting path and an expansive path. The contracting path is a typical convolutional network composed of repeated convolutional applications. The spatial information is reduced while the feature information is increased during the contraction. Through a series of up-convolutions and concatenations with high-resolution features from the contracting path, the expansive pathway combines feature and spatial information.

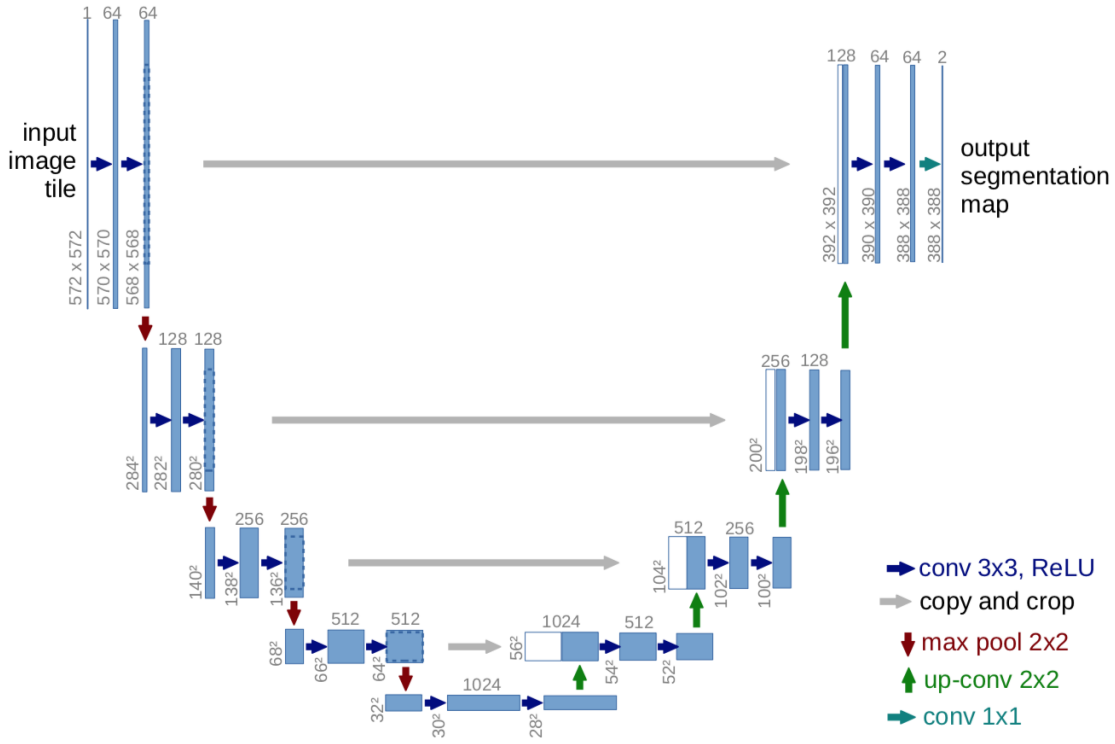


Figure 1: *"U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations". From (Ronneberger et al., 2015)*

Google considered improving the design of its neural architecture search models (Zoph and Le, 2016). The artificial intelligence team at Google created a new type of neural architecture search engine. It is based on the idea that a neural network can be built from the ground up for structure and connectivity. With the goal of ensuring that excellent models can be created from scratch, by training on small proxy tasks and then receiving precise rewards, then reinstating them and repeating the process several times, as shown in the Figure 2 (Tan et al., 2018).

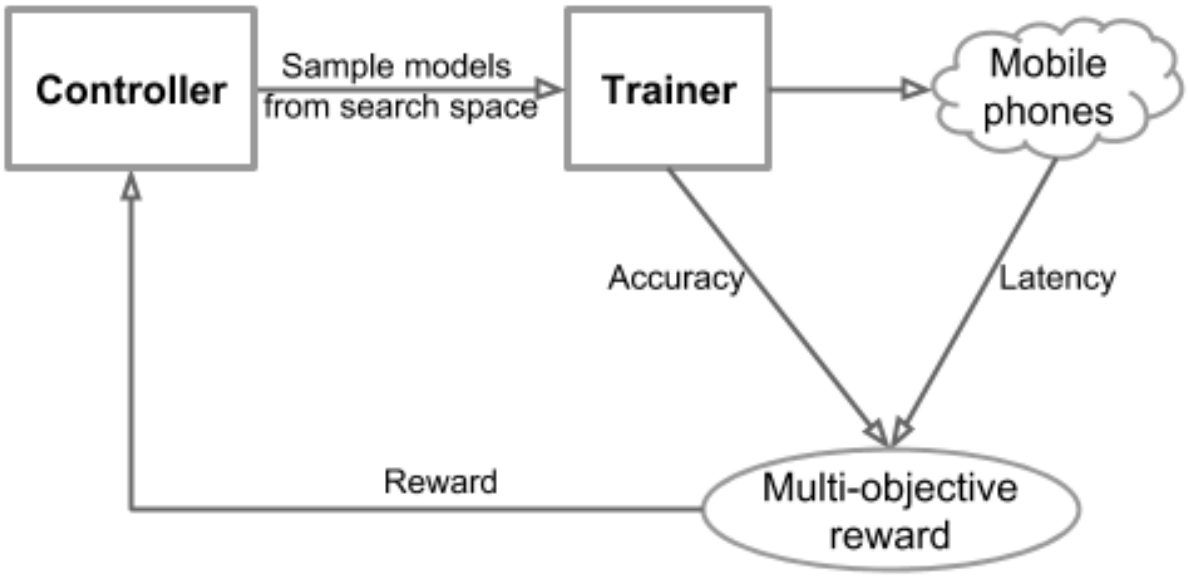


Figure 2: *"Overview of Platform-Aware Neural Architecture Search for Mobile". From (Tan et al., 2018)*

As a result of this process, a new model known as efficient b0 was created. This took about two weeks and about ten thousand central processing unit (CPU)s. This model is extremely efficient when compared to human design. The model outperforms densnet, a subspace of the neural network on which it is based, by a factor of nine. However, they discovered that searching for neural architecture was very expensive when scaling up the model. As a result, they devised a new method of scaling up the model. Unlike the conventional process for model scaling, which is to arbitrarily increase the CNN depth or width, or to use larger input image resolution for training and evaluation. Compound scaling theory combines all of the different scaling strategies into a more accurate model by using a ratio formula. The idea is to balance the upsampling of width, depth and resolution by scaling them with a constant ratio, as shown in the Figure 3. This constant ratio is determined by the parameters of alpha, beta, and gamma (Tan and Le, 2019).

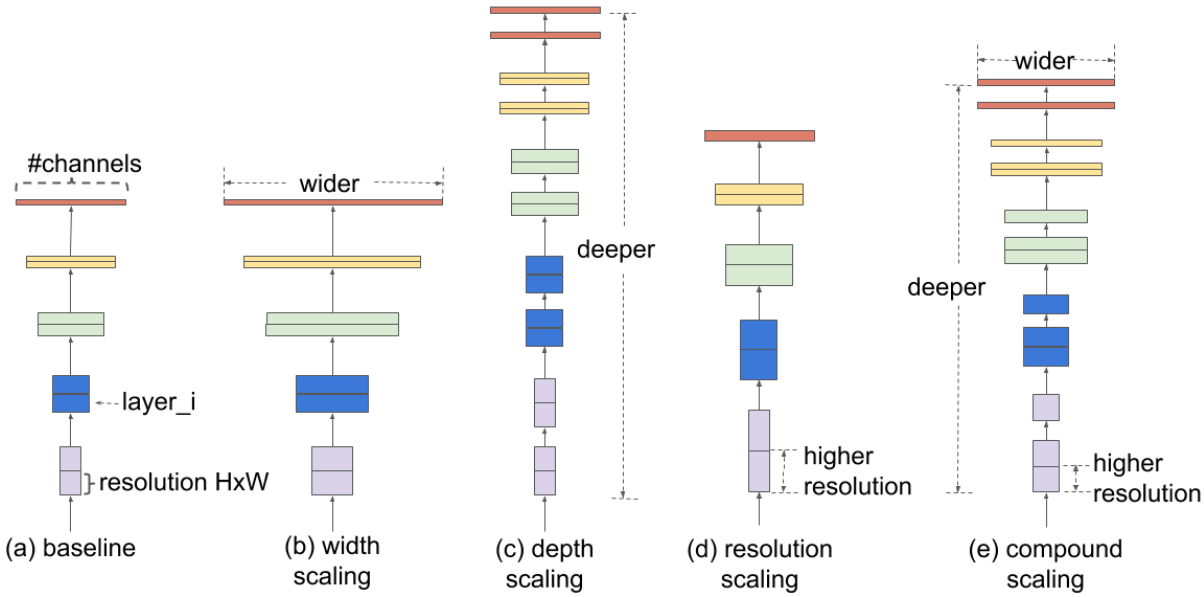


Figure 3: *"Model Scaling. (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is Google's proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio". From (Tan and Le, 2019)*

Its main building block is the mobile inverted bottleneck (cite), to which they also add squeeze-and-excitation optimization(cite). The Mobile inverted bottleneck was originally proposed for the MobileNetV2 CNN architecture. The idea is to have a residual connection from the narrow or bottleneck to the other one using a residual as shown in the Figure 4 (Sandler et al., 2018). The Squeeze-and-Excitation Block receives input from a convolutional block. Using average pooling, each channel is "squeezed" into a single numeric value. A dense layer followed by an activation function Rectified Linear Units (ReLU), $Relu(x) = max(0, x)$, and another dense layer, followed by a sigmoid, $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$. Finally, each convolutional block feature is scaled by the side network; this is the "excitation" as shown in the Figure (Hu et al., 2017).

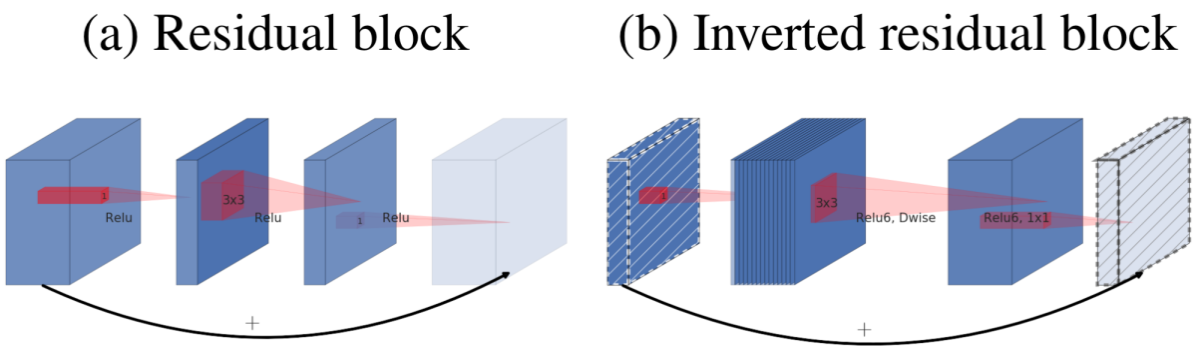


Figure 4: *"The difference between residual block and inverted residual. Diagonally hatched layers do not use non-linearities. We use thickness of each block to indicate its relative number of channels. Note how classical residuals connects the layers with high number of channels, whereas the inverted residuals connect the bottlenecks. Best viewed in color". From (Tan and Le, 2019)*

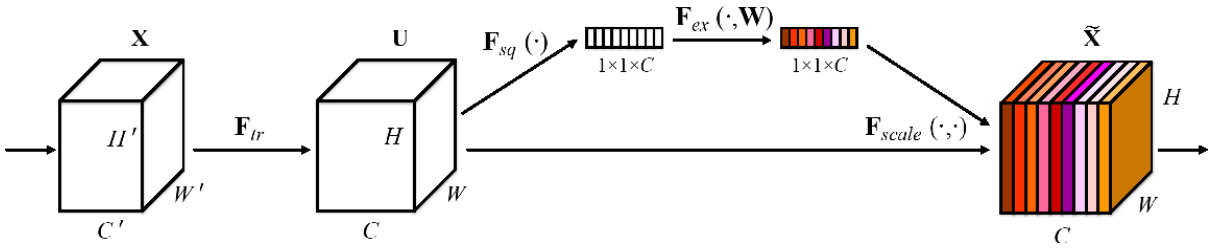


Figure 5: *"A Squeeze-and-Excitation block" From (Hu et al., 2017)*

2.3 Meta Pseudo Labels

Many computer vision tasks, such as image classification, object detection, and semantic segmentation, have used pseudo labels or self-training to improve models. If the pseudo labels are incorrect, the student will learn based on incorrect data. This disadvantage is also known as the confirmation bias problem in pseudo-labelling. This was a problem that also affected Google's previous work, Noisy Student (Xie et al., 2019). This previous work achieved state-of-the-art performance on ImageNet. This is a Semi-Supervised process of transferring knowledge from a teacher model to a student model, as shown in the Figure 6, improving on ideas such as Semi-Supervised Distillation (Hinton et al., 2015). This work is also combined with other works such as UDA (Wang et al., 2021), where data augmentation is applied to unlabeled data to significantly improve semi-supervised learning. This technique uses different existing augmentation methods depending on the

task at hand, including back translation, AutoAugment, and TF-IDF word replacement as shown in Figure 7.

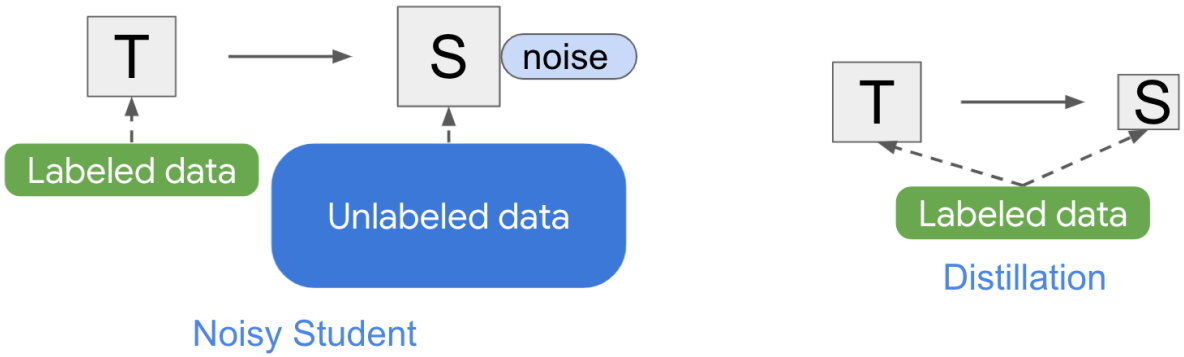


Figure 6: *"illustrations of Noisy Student and knowledge distillation". From (Google AI Blog, 2021)*

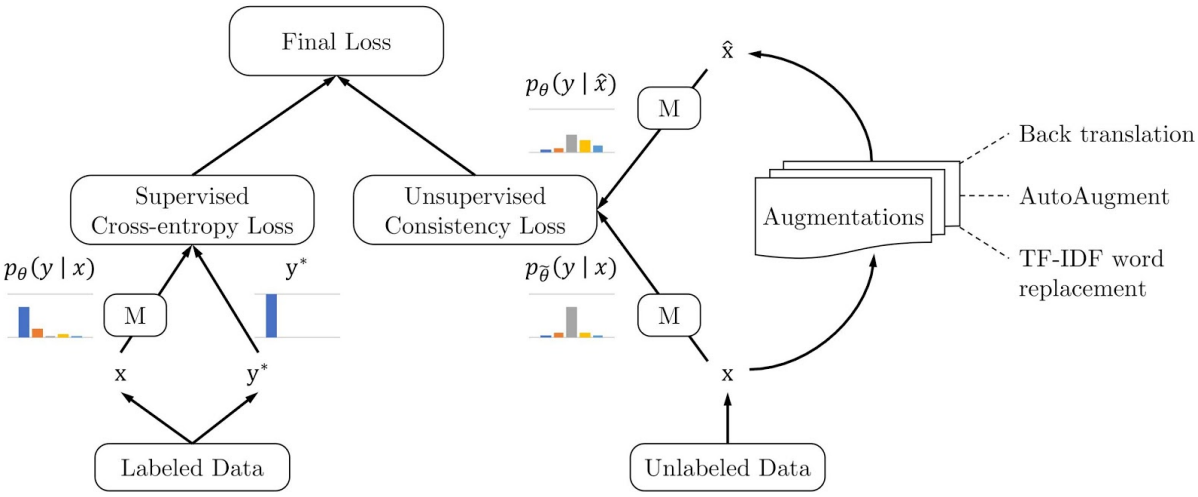


Figure 7: *"Training objective for UDA, where M is a model that predicts a distribution of y given x". From (Wang et al., 2021)*

In Meta Pseudo labels, they design a systematic mechanism for the teacher to correct the bias by observing how the pseudo labels affect the student (Pham et al., 2020). The feedback signal is used as a reward to train the teacher throughout the student's learning, as shown in the Figure 8. The teacher's parameters are optimized according to the performance of the student on labelled data. The hope is that pseudo labels will ultimately achieve a low loss on labelled data. This work also uses ideas from previous work in meta-learning with a one-step gradient update of the student parameters. In the first training step, given a single input example x, the teacher produces the conditional class distribution to train the student. In phase 2, the student learns from the teacher by learning the target distribution for the first time. Then they train the student model using gradient descent in order to make it more efficient.

Figure 8: *"At each training step t, our Meta Pseudo Labels (MPL) update consists of two phases. Updating the Student (top): The teacher network q^{Ψ} assigns the conditional class distribution for a training example x. The student pθ learns from $(x, q\Psi(x))$ by standard supervised learning, updating from $\theta^{(t)}$ to $\theta^{(t+1)}$. Updating the Teacher (bottom): The teacher updates its parameters Ψ based on the resulting student's cross-entropy loss on validation data (xval, yval)". From (MPL, 2020)*

Methodology

3.1 Pipeline Design

The project was designed as a two-stage AI pipeline:

1. Stage 1: Segmentation with Meta Pseudo Labels. An EfficientNet-B0-U-Net architecture was chosen for both the teacher and student models. The goal was to train the student to segment tumors using a small labeled dataset and a larger unlabeled dataset via the Meta Pseudo Labels training loop.

2. Stage 2: Classification. The output of the trained segmentation model (the tumor mask) was intended to be used to crop the tumor region, which would then be fed into an EfficientNet-B0 classifier to determine if the tumor was benign or malignant.

3.1.1 Meta Pseudo Labels Training

The core of the implementation was the MPL training loop for the segmentation task:

- Models: A teacher-model and an identical student-model (EfficientNet-B0-U-Net).
- Data: A labeled-dataset (images with ground-truth masks) and an unlabeled-dataset (images without masks).
- Optimizers: Separate Adam optimizers for the teacher and student.
- Loss Function: A Dice loss function to measure segmentation accuracy.

The intended training step was as follows:

1. Train the student-model on a batch from the labeled-dataset.
2. Use the teacher-model to predict pseudo-label masks for a batch from the unlabeled-dataset.
3. Train the student-model on the batch of images with their corresponding pseudo-labels.
4. Evaluate the updated student-model's performance by calculating its Dice loss on the labeled batch. This loss serves as the meta-loss.
5. Use the meta-loss to calculate gradients and update the teacher-model's weights.

3.2 Dataset

To detect cases of breast cancer in mammograms and assess their effectiveness, the deep learning model planned to be developed throughout this project would require real-world data to understand the underlying patterns. But because medical data is sensitive, it is important to take into account ethical problems. When using sensitive medical data, such as mammograms, the main ethical issue is whether the data can be linked back to the original patient. As a result, this project employed completely anonymous, open-source, and public datasets. The type of data used in this project was ethically approved and authorised by Coventry University.

The mini-MIAS dataset for more tests regarding dataset bias. This dataset includes both abnormal and normal instances, yielding three classifications: malignant, benign, and malignant. The mini-DDSM dataset was chosen for the training stage, not because it has better samples than the mini-Mias, but because it has counter masks. And this is important for the classification of the tumour (C.D. Lekamlage and Cheddad, 2020) and (et al, 1994).

3.2.1 Constraints

These images from the mini-DDSM dataset are from the 1990s. They are far from ideal. But in an ideal digital X-ray imaging system, the mammogram has only one source of the noise. This noise is caused by the detector's random absorption. Other noise sources exist that are linked to the detector constraints. These noises are technically classified and referred to as radiographic noise or 'Mottle.' They may occur as a result of the granularity and structure of the screen (Yaffe, 2010).

The difficulties in detecting these anomalies stem from the fact that they vary in size, morphology, and brightness, and are sometimes visually similar to or camouflaged with healthy tissues. There are also a variety of normal mammogram appearances as women's tissue structure changes with age, due to hormonal influences, or there may be temporal changes in densities in the breasts. The clarity with which lesion margins are visualised is related to the correct detection of malignancy as well as the severity of the case. Edge enhancement for feature detection is required in such cases because the clarity with which the lesion margin is visualised correlates with the severity of the diagnosis.

3.2.2 Data Extraction

This dataset lacks all the information needed for the proposed task, segmentation and classification. First, for segmentation, there is a need to use masks because semantic

segmentation relies on selecting pixels that are related to the object. And given the fact that contour masks are not often used, it's important to generate masks from these contour masks. Second, because in the project there is a need to classify the region of interest, in this case, the tumour region.

To generate this data, there is a need to crop the region of interest related to the tumour. And then, depending on the task, certain data points will be selected and others won't.

Algorithm 1: Pseudocode for masks creation

```
finalMask = imageMinimumIntensity
Image ← image
Masks ← getMasks(Image)
for mask ∈ Masks do
    mask = resize(mask, Image.height, Image.width)
    thresh = threshold(mask, otsu.threshold)
    Contours = findContours(thresh)
    for contour ∈ Contour do
        drawContoursOn(mask, contour.coordinates)
    end
    finalMask = maximum(finalMask, mask)
end
```

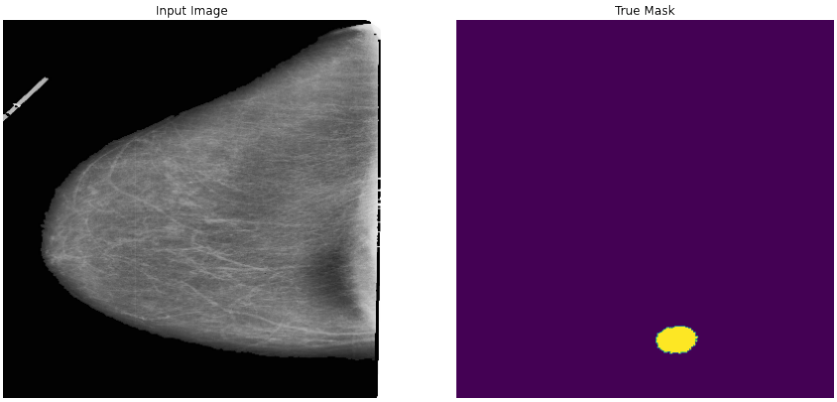


Figure 9: *Enter Caption*

3.3 Pre-Processing

Pre-processing data is a critical step in the data mining process. Data collection methods are frequently uncontrolled, resulting in out-of-range values. Knowledge discovery during the training phase gets more challenging when there is a lot of irrelevant and redundant data, as well as noisy and untrustworthy data. When the interpretation of the results is critical, this aspect should be carefully considered.

In this project, the main focus of the pre-processing phase is on noise reduction and feature enhancement. Because the main problems with these images are the poor signal-to-noise ratio in lesion imaging and the absence of a consistent lesion location, shape, and appearance (Oliver et al., 2010). Image feature enhancement of mammograms at post-acquisition stages forms an effective tool for CAD techniques. The basic anatomy of the breast consists of various objects, apart from anomalies that appear on mammograms. Dense tissues in the breast region display brighter intensities in mammograms, while regions with fat or skin appear comparatively darker. This leads to poor contrast as well as increased levels of noise. The basic goal of mammography enhancement, in theory, is to improve the contrast between regions of interest and the background, as well as to sharpen the margins of the lesion zone. This means that the detection of masses has become easier. In a nutshell, contrast and edge enhancement can be applied to mammograms as a part of computer-aided analysis before a final decision can be made by the radiologist.

There are several ways to reduce the noise in the mammogram image. One of the most effective ways to do so is with a type of breast segmentation with tag removal. Sometimes, a pectoral muscle removal is applied to the mammogram image to further segment the region of interest. The problem is with the loss of data with these approaches, even if they are very effective and widely used. This is one of the concerns found during this project. Other ways of reducing the noise level from the mammogram image are to apply simple filters. These are less effective and less used, but the loss of data is minimised.

To assess the improvement from the previous image. An image quality assessment (IQA) is normally used. In medical imaging, this plays an important role in the design and manufacturing processes of image acquisition and processing systems. In medical diagnosis, any visual artefact in the reconstruction may hinder diagnostic conclusions and lead to severe, unpredictable consequences. A study on dedicated medical image quality assessors is strongly required and can be utilized as a guide to evaluate and improve the performance of medical imaging and processing algorithms. Medical image evaluation is a multifaceted process that can be approached from various angles. A variety of methods for assessing objective image quality have been proposed over the years, including subjective quality assessment and objective quality assessment.

3.3.1 Noise removal

For simple filter noise reduction, the Median filter seems to perform better by reducing the noise and preserving edges in noisy mammogram images, resulting in improved image clarity. This is compared to other filters, such as the Average filter and the Gaussian filter (Pati and Panda, 2020). But this filter was then compared to the mask segmentation removal method in a different paper and it did show that tag removal performed better than the Median filter (Ashenafi et al., 2021).

Despite performing better than the Median filter, as shown in the Algorithm 2. It is also important because different tags are attached to mammograms, such as the type of view, the date. Most of the time, these tags are written in the mammogram's background region and maybe readily erased by utilising the connected-component labelling (CCL) approach.

Algorithm 2: "Pseudocode for Tag Removal." From (Pati and Panda, 2020)

1 Read grayscale mammogram
2 Apply Wiener Filter To remove Noise
3 $iWeiner = Weinerer(grayimage)$
4 Apply Otsu's thresholding $mask = im2bw(iWeiner, graythresh(iWeiner))$
5 Fill holes i the Binary Image
6 $mask = imfill(mask,' holes')$
7 Select The Largest binary region as a mask $mask = bwpropfill(mask,' area', 1)$
8 Remove Tags $masked = (iwiener(mask) = 0)$
9 Tags Removed Mammogram

Based on the Algorithm 2 and 3. In this project, a new algorithm was created after finding that adaptive thresholding, a method of calculating the threshold value for smaller regions, worked well for selecting or segmenting tags. Resulting in different threshold values for various regions. In this project, the Otsu's thresholding in the Algorithm2 is used after applying the median filter from the Algorithm 3, Then the largest areas are selected and areas not related to the mask are then turned to zero, as shown in the algorithm.

Algorithm 3: "Remove label." From (Ashenafi et al., 2021)

1 Input: Image from MIA's database
2 Output: Filtered image without label
3 Step 1: Apply 3*3 median filters.
4 Step2: convert the image to a binary form with an appropriate threshold.
5 Step3: find the area for all components in the image were Arranged from the largest to the smallest.
6 Step4: Create Mask 1, which represents the two largest Areas in the image breast and background and Exclude small areas including the label.
7 Step 5: multiply mask1 with the filtered image to obtain the Pre-processing image.
8 Step 6: end

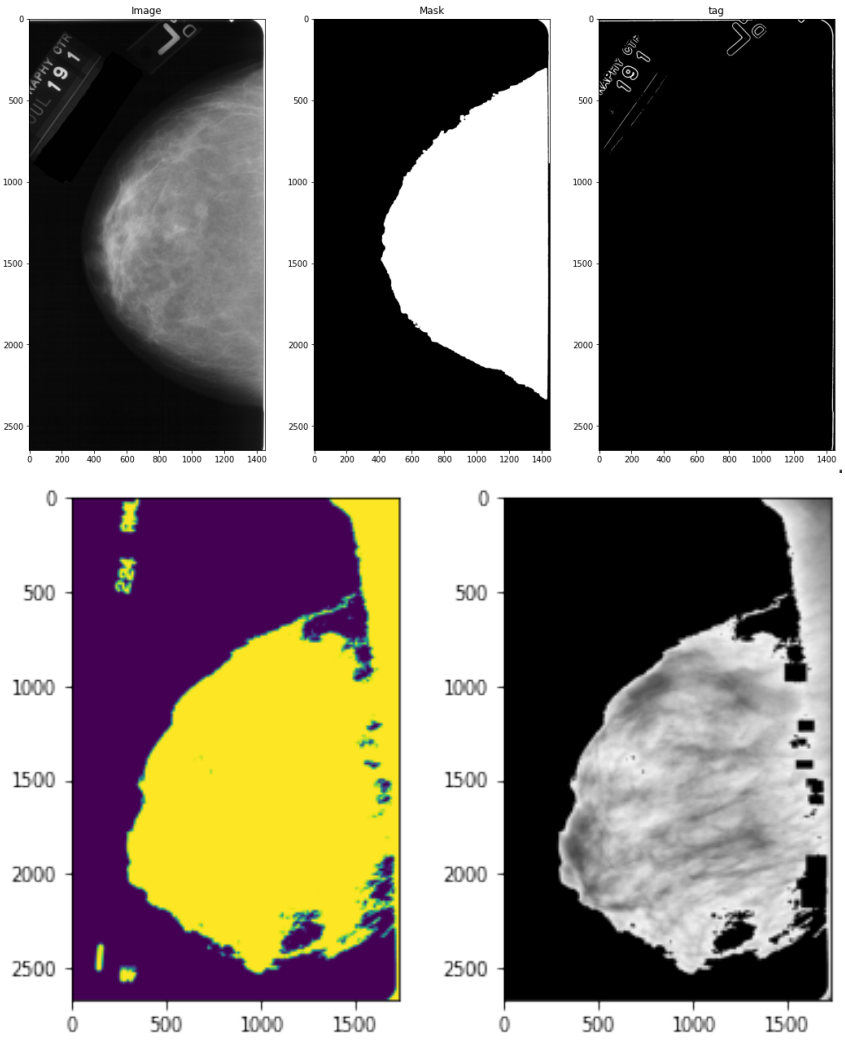


Figure 10: *Tag and label removal*

The pectoral muscle is the muscle that connects the front of the human chest with the bones of the upper arm and shoulder. Pectoral segmentation is important as the pectoral region and the breast region may have a similar intensity or texture appearance. The process to remove the pectoral muscle is shown in the Algorithm 4. This algorithm also shows the combination of previous algorithm two approaches from Algorithm 2 and 3.

Algorithm 4: Noise removal

```
img =medianFilter(image)
tagMask =adaptiveThreshold(img,invertedBinaryThresh)
breastSidePixels = halfOfImageWithMostPixels ≠ 0
thresholdPixels =amountAboveMedian(breastSidePixels)
breastMask =threshold(img, otsuThresh)
2ndbiggestArea = findAreas(breastMask)
for area in findAreas(breastMask) do
    if area < 2ndbiggestArea then
        breastMask.whereIsEqual(area) = 0
    end
end
if amountOfPixels(breastMask) ≥ thresholdPixels
 then
    Image.where(breastMask == 0) = 0
end
```

- **Pectoral Muscle removal** (Gsunit)

```
if Not right-oriented then
    Flip image
end
canny edge detection
Obtain Hough transform lines
Shortlist lines
if More than 1 line is short then
    Choose one with the least area removed Darken
    pixels of area segmented by selected line
end
Choose one with the least area removed Darken pixels
of area segmented by selected line
```

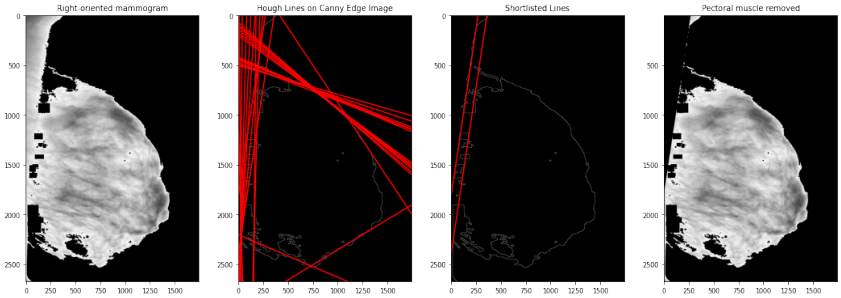


Figure 11: *Pectoral Muscle Removal*

3.3.2 Image Enhancement

The primary goal of image enhancement is to change, improve the visual characteristics of a digital image so that it is more appropriate for a specific task from the standpoint of both human perception and computer vision. Sharpening is the process of increasing the contrast between bright and dark regions to properly highlight the image's features. It also improves contrast near the edges, with the overall goal of implementing the high emphasis attributes as applicable in the frequency domain.

For image enhancement, the method chosen took time, subjective, and objective image quality assessment into consideration. In subjective image quality assessment, human experimenters are employed to rate the image quality using various methods. The development of the subjective quality assessment can be dated back to a long time ago and is still in continuous research by several studies. This is also important just for the simple fact that it can make the image clear for the radiologist to diagnose.

In this case, a participating radiologist was shown enhanced mammographic images for clinical validation. The radiologist graded the enhanced images in comparison to the original images (without enhancement) based on visual perception, with grade values ranging from 1 to 10, with 1 indicating poor visual quality and 10 indicating high visual quality. The Table 4 shows the average grades obtained for each enhancement method. Three enhancement methods—contrast stretching, morphological enhancement and contrast stretching, and unsharp masking and contrast stretching (UMCA)—provide clinically acceptable images in which the abnormalities are better highlighted, according to the radiologist's grading (Kaur et al., 2019).

Enhancement Method	**Grade**
Alpha trimmed mean filter	3
Contrast stretching	8
Histogram equalization	1
CLAHE	6
RMSHE	4
Contra-harmonic mean filter	5
Mean filter	5
Median filter	5
Hybrid median filter	5
Morphological enhancement	6
Morphological enhancement and contrast stretching	9
Unsharp masking	6
Unsharp masking and contrast stretching	9
Wavelet-based subband filtering	5

Table 2: *"Radiologist's Grading of Enhanced Images." From (Kaur et al., 2019)*

Normally, these approaches are classified into several categories, such as contrast stretching techniques, region-based, feature-based and non-linear enhancement and so on. As it's clear to see, image enhancement that involves contrast stretching are very good from a subjective and objective point of view. This project doesn't use anything that involves contrast stretching, it's a very slow process. Because of this, the contrast limited adaptive histogram equalization (CLAHE) was used instead. This is a variant of adaptive histogram equalization in which the contrast amplification is limited to reduce this problem of noise amplification(Hummel, 1977).

Enhancement Method	OCA (%)	ICAA (%)	ICAN (%)
Raw images with no enhancement	84.3	74.5	94.1
Alpha trimmed mean	79.4	70.6	88.2
Contrast stretching	88.2	86.3	90.2
Histogram equalization	74.5	74.5	74.5
CLAHE	88.2	76.5	100
RMSHE	84.3	72.5	96
Contra harmonic mean	83.4	74.5	92.2
Mean filter	82.4	72.5	92.2
Median filter	80.4	68.7	92.2
Hybrid median filter	81.4	72.5	90.2
Morphological enhancement	74.5	68.7	80.4
MECA	92.2	90.2	94.1
Unsharp masking	84.3	72.5	96.1
UMCA	85.3	80.4	90.2
Wavelet-based subband filtering	84.3	74.5	94.1

Table 3: *"Comparison of Classification Performance of the Enhancement Methods* OCA, overall classification accuracy; ICAA, individual abnormal class accuracy; ICAN, individual normal class accuracy*". From (Kaur et al., 2019)*

3.3.3 Regularization

Because the distribution is not uniform, the mini-MIAS dataset is heavily skewed. To avoid training a biassed CNN model, this imbalance must be considered. Oversampling the dataset would reduce the number of samples from which the model could learn by dropping samples from the majority class. Adding class weights, on the other hand, would turn the loss into a weighted average, giving under-represented classes more weight. However, because mini-DDSM is used for training, the imbalanced data is not an issue.

When training a model for generalisation, data augmentation is important to reduce overfitting and indirectly regularise the model. It is an oversampling of the data to possible samples that are less present in the training sample. This is also significant because more data can be generated to train the model during this process.

In model training, deep learning regularisation strategies such as data balancing and augmentation are commonly used. Data balancing and breast image monitoring to avoid bias during the training process (Dhungel et al., 2017). In the project's case, the mammogram images are flipped randomly left, right, top, down and zoomed as well. The

mammograms are also normalized by the maximum pixel intensity of 255 and then normalised by the image normalised minus the mean of the training sample. All this is divided by the standard deviation. This is a technique for altering the range of pixel intensity values. Normalization is also known as histogram stretching or contrast stretching. The goal is to achieve dynamic range consistency for a set of signals.

3.4 Training

The mini-DDSM dataset is divided the data into training, and test sets. In Would nice to to a normal cross-validation, there was nor resources for this. And it was establish the mini-MIAS dataset is left for test as well.

Normalisation is explained on previous subsection, Regularization. It is also important to mention that the input images are resized to input size of the model, 384. Because the labels for each mammogram are in categorical string format, the status of the mammogram is encoded into a numerical format. A similar thing is done with the segmentation image for dice loss, which is derived from the Srensen–Dice coefficient, a 1940s statistic used to assess the similarity of two samples. In 2016, it was introduced to the computer vision community for medical image segmentation used in papers like (Li et al., 2019). This because the per-pixel loss is calculated discretely in cross entropy loss, regardless of whether the adjacent pixels are boundaries or not.

It is already known and explained that for this project Efficientnet and EfficientUnet. Efficient B0 with around 4 Million parameters and EfficientUnet with around 14 Million parameters. For Overall Classification of 3 classes, benign, cancer and normal. Code implementation ://drive.google.com/drive/folders/19FoL6YdFd5S1c3aELJlUfZcvzLsU4yct?usp=sharing,https://drive.google.com/drive/folders/1NRSkEEx29DNeOVry0jtIBKrek6aGF2IX?usp=sharing

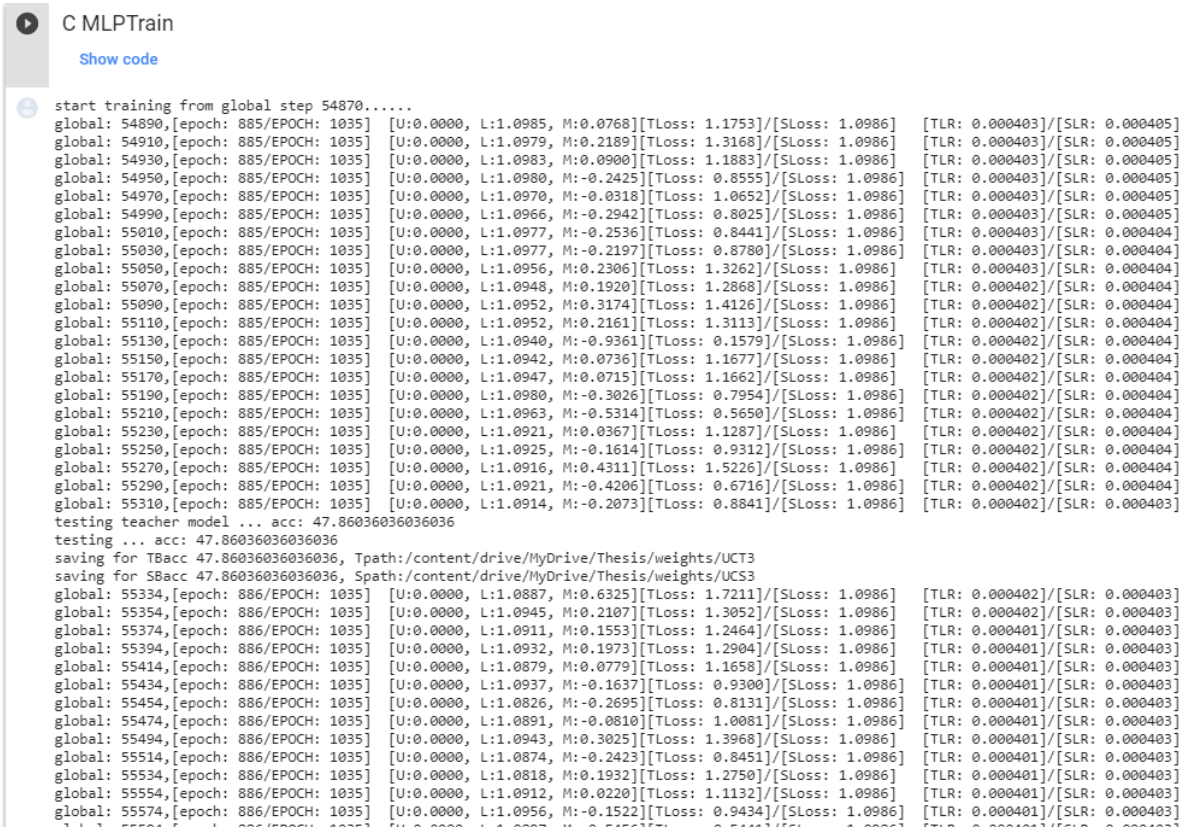


Figure 12: *3 (Normal, benign, cancer) Class classification performance after at least 5 epochs (47%) softmax cross entropy*

Region of tumour classification Meta Pseudo Labels Train

Show code

```
global: 59575,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6871, M:-0.2153][TLoss: 0.4718]/[SLoss: 0.6931]  [TLR: 0.000387]/[SLR: 0.000389]
global: 59595,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6984, M:0.1934][TLoss: 0.8917]/[SLoss: 0.6931]   [TLR: 0.000387]/[SLR: 0.000389]
global: 59615,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6940, M:-0.1890][TLoss: 0.5050]/[SLoss: 0.6931]  [TLR: 0.000387]/[SLR: 0.000389]
global: 59635,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6881, M:0.0284][TLoss: 0.7165]/[SLoss: 0.6931]   [TLR: 0.000387]/[SLR: 0.000389]
global: 59655,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6974, M:0.1064][TLoss: 0.8037]/[SLoss: 0.6931]   [TLR: 0.000387]/[SLR: 0.000388]
global: 59675,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6873, M:0.0094][TLoss: 0.6967]/[SLoss: 0.6931]   [TLR: 0.000387]/[SLR: 0.000388]
global: 59695,[epoch: 898/EPOCH: 1035]  [U:0.0000, L:0.6932, M:-0.1135][TLoss: 0.5797]/[SLoss: 0.6931]  [TLR: 0.000387]/[SLR: 0.000388]
global: 59720,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6906, M:-0.0861][TLoss: 0.6045]/[SLoss: 0.6931]  [TLR: 0.000387]/[SLR: 0.000388]
global: 59740,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6932, M:-0.0379][TLoss: 0.6553]/[SLoss: 0.6931]  [TLR: 0.000387]/[SLR: 0.000388]
global: 59760,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6852, M:0.1443][TLoss: 0.8295]/[SLoss: 0.6931]   [TLR: 0.000387]/[SLR: 0.000388]
global: 59780,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6959, M:-0.1208][TLoss: 0.5751]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000388]
global: 59800,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6860, M:0.2734][TLoss: 0.9594]/[SLoss: 0.6931]   [TLR: 0.000386]/[SLR: 0.000388]
global: 59820,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6847, M:-0.2276][TLoss: 0.4570]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000388]
global: 59840,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.7000, M:-0.0571][TLoss: 0.6429]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000388]
global: 59860,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6969, M:-0.0186][TLoss: 0.6783]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000388]
global: 59880,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6923, M:-0.0826][TLoss: 0.6098]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000388]
global: 59900,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6878, M:-0.1436][TLoss: 0.5441]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000388]
global: 59920,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.7013, M:-0.2672][TLoss: 0.4341]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000387]
global: 59940,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6897, M:0.0452][TLoss: 0.7349]/[SLoss: 0.6931]   [TLR: 0.000386]/[SLR: 0.000387]
global: 59960,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6896, M:0.0056][TLoss: 0.6952]/[SLoss: 0.6931]   [TLR: 0.000386]/[SLR: 0.000387]
global: 59980,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6914, M:-0.0764][TLoss: 0.6151]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000387]
global: 60000,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6985, M:0.0040][TLoss: 0.7025]/[SLoss: 0.6931]   [TLR: 0.000386]/[SLR: 0.000387]
global: 60020,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6897, M:-0.0599][TLoss: 0.6298]/[SLoss: 0.6931]  [TLR: 0.000386]/[SLR: 0.000387]
global: 60040,[epoch: 899/EPOCH: 1035]  [U:0.0000, L:0.6975, M:0.0264][TLoss: 0.7239]/[SLoss: 0.6931]   [TLR: 0.000386]/[SLR: 0.000387]
global: 60065,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6931, M:-0.0113][TLoss: 0.6818]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000387]
global: 60085,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6939, M:-0.0276][TLoss: 0.6663]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000387]
global: 60105,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6923, M:0.1617][TLoss: 0.8540]/[SLoss: 0.6931]   [TLR: 0.000385]/[SLR: 0.000387]
global: 60125,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6947, M:-0.0542][TLoss: 0.6405]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000387]
global: 60145,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6891, M:0.0005][TLoss: 0.6896]/[SLoss: 0.6931]   [TLR: 0.000385]/[SLR: 0.000387]
global: 60165,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6923, M:-0.1929][TLoss: 0.4995]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000387]
global: 60185,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6956, M:-0.0861][TLoss: 0.6095]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000387]
global: 60205,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6954, M:0.0821][TLoss: 0.7775]/[SLoss: 0.6931]   [TLR: 0.000385]/[SLR: 0.000386]
global: 60225,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6931, M:-0.1267][TLoss: 0.5664]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000386]
global: 60245,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6938, M:-0.5176][TLoss: 0.1763]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000386]
global: 60265,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6931, M:0.0797][TLoss: 0.7728]/[SLoss: 0.6931]   [TLR: 0.000385]/[SLR: 0.000386]
global: 60285,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6894, M:0.2393][TLoss: 0.9286]/[SLoss: 0.6931]   [TLR: 0.000385]/[SLR: 0.000386]
global: 60305,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6884, M:-0.2267][TLoss: 0.4616]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000386]
global: 60325,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6931, M:-0.0944][TLoss: 0.5988]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000386]
global: 60345,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6907, M:-0.0795][TLoss: 0.6112]/[SLoss: 0.6931]  [TLR: 0.000385]/[SLR: 0.000386]
global: 60365,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6948, M:0.0479][TLoss: 0.7426]/[SLoss: 0.6931]   [TLR: 0.000384]/[SLR: 0.000386]
global: 60385,[epoch: 900/EPOCH: 1035]  [U:0.0000, L:0.6891, M:-0.1162][TLoss: 0.5729]/[SLoss: 0.6931]  [TLR: 0.000384]/[SLR: 0.000386]
testing teacher model ... acc: 52.38784370477568
testing ... acc: 52.38784370477568
global: 60410,[epoch: 901/EPOCH: 1035]  [U:0.0000, L:0.6854, M:-0.0593][TLoss: 0.6261]/[SLoss: 0.6931]  [TLR: 0.000384]/[SLR: 0.000386]
```

Figure 13: *2 (benign, cancer) Class classification performance after at least 5 epochs (52%) softmax cross entropy*

Segmentation Meta Pseudo Labels Train

Show code

```
global: 55130,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4901, M:-10500.7627][TLoss: 587089.9375]/[SLoss: 90960.1172]       [TLR: 0.000475]/[SLR: 0.000476]
global: 55150,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4935, M:10916.1084][TLoss: 608506.8750]/[SLoss: 90960.1172] [TLR: 0.000475]/[SLR: 0.000476]
global: 55170,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4964, M:19000.0176][TLoss: 616590.7500]/[SLoss: 90960.1172] [TLR: 0.000475]/[SLR: 0.000476]
global: 55190,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4950, M:10129.2480][TLoss: 607719.9375]/[SLoss: 90960.1172] [TLR: 0.000475]/[SLR: 0.000476]
global: 55210,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4957, M:-17296.9609][TLoss: 580293.8125]/[SLoss: 90960.1172]       [TLR: 0.000475]/[SLR: 0.000476]
global: 55230,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4925, M:39493.8633][TLoss: 637084.6250]/[SLoss: 90960.1172] [TLR: 0.000475]/[SLR: 0.000476]
global: 55250,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4951, M:-28522.8242][TLoss: 569067.8750]/[SLoss: 90960.1172]       [TLR: 0.000475]/[SLR: 0.000476]
global: 55270,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4916, M:15379.2598][TLoss: 612970.0625]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000476]
global: 55290,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4960, M:6244.0840][TLoss: 603834.8750]/[SLoss: 90960.1172]  [TLR: 0.000474]/[SLR: 0.000476]
global: 55310,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4961, M:-35645.6328][TLoss: 561945.1250]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000476]
global: 55330,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4950, M:-359.2812][TLoss: 597231.5000]/[SLoss: 90960.1172]  [TLR: 0.000474]/[SLR: 0.000476]
global: 55350,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4959, M:4352.6577][TLoss: 601943.3750]/[SLoss: 90960.1172]  [TLR: 0.000474]/[SLR: 0.000476]
global: 55370,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4957, M:-413.5147][TLoss: 597177.3125]/[SLoss: 90960.1172]  [TLR: 0.000474]/[SLR: 0.000476]
global: 55390,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4975, M:-22732.1191][TLoss: 574858.6875]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000476]
global: 55410,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4954, M:-10878.7207][TLoss: 586712.0000]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000476]
global: 55430,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4898, M:14279.4873][TLoss: 611870.1875]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000476]
global: 55450,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4942, M:-19173.3164][TLoss: 578417.5625]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55470,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4923, M:-16028.0938][TLoss: 581562.6250]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55490,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4935, M:-4184.1694][TLoss: 593406.6250]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
global: 55510,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4926, M:-14694.5918][TLoss: 582896.1250]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55530,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4972, M:-27478.3789][TLoss: 570112.3750]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55550,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4959, M:23503.3770][TLoss: 621094.1250]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
global: 55570,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4933, M:-1033.1158][TLoss: 596557.6250]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
global: 55590,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4952, M:17070.3242][TLoss: 614661.0625]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
global: 55610,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4956, M:-4949.8057][TLoss: 592640.9375]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
global: 55630,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4892, M:-13003.8789][TLoss: 584586.8750]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55650,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4944, M:-19414.0742][TLoss: 578176.6875]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55670,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4958, M:-41553.3828][TLoss: 556037.3125]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55690,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4952, M:13817.1533][TLoss: 611407.8750]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
global: 55710,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4933, M:-10202.7998][TLoss: 587387.8750]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55730,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4962, M:-26500.3398][TLoss: 571090.3750]/[SLoss: 90960.1172]       [TLR: 0.000474]/[SLR: 0.000475]
global: 55750,[epoch: 885/EPOCH: 1035]  [U:597590.1250, L:0.4957, M:46187.1641][TLoss: 643777.8750]/[SLoss: 90960.1172] [TLR: 0.000474]/[SLR: 0.000475]
testing teacher model ... acc: 1.0882598859769805
testing ... acc: 1.0882598859769805
saving for TBacc 1.0882598859769805, Tpath:/content/drive/MyDrive/Thesis/weights/b4ST2
saving for SBacc 1.0882598859769805, Spath:/content/drive/MyDrive/Thesis/weights/b4SS2
```

Figure 14: *2 (normal, tumour) segmentation performance after at least 1 epoch (1%) dice loss*

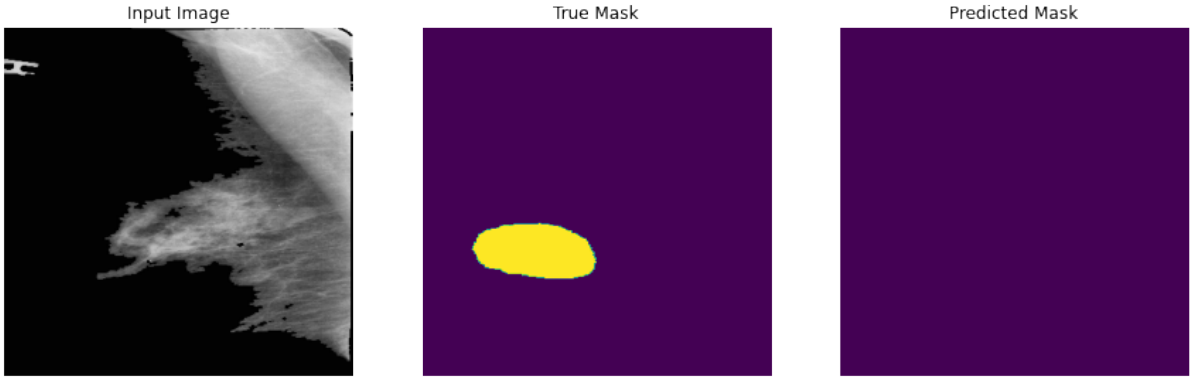


Figure 15: *The model predicting empty masks*

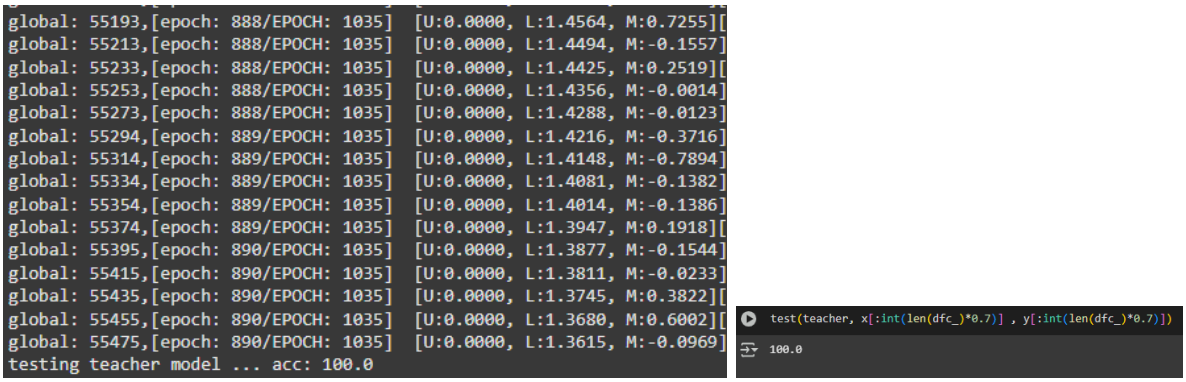


Figure 16: *Sanity test using .keras/datasets/oxford-102-flowers/ 100% while training, while testing as well, maybe too good to be true*

Project Results and Management

4.1 Research Discussion

The primary objective of implementing the Meta Pseudo Labels pipeline could not be achieved due to a foundational failure at the first stage of the process: the student model's inability to learn from the supervised data.

4.1.1 Experiment 1: Tumor Segmentation

Before initiating the Meta Pseudo Labels loop, the student model (EfficientNet-B0-U-Net) was trained exclusively on the small, labeled dataset to establish a baseline. The model completely failed to converge. After the first epoch, the Dice loss was approximately 1.0, and the Dice coefficient was near 0, indicating no meaningful overlap between the predicted masks and the ground-truth masks. The model was effectively producing random noise or empty masks, when it came to the segmentation section. Trying to introduce or create synthetic label data with ground-truth by rotating them, unfortunately this resulted to a Data explosion to our google's drive, having images saved on the first layer of the driver did make it unusable.

This failure could be attributed to several factors:

- Data scarcity, the number of unique tumor examples might have been insufficient for the model to learn the complex patterns and shapes of tumors.
- The pre-processing choices steps might have not the right ones, even if the were justified given current researches in the field.
- The masks generated from the contours might have been inaccurate, further complicating the learning task.
- The model EfficientNet-B0-U-Net, might have been too complex for such a small dataset, leading to an immediate failure to generalize from the initial training batches.
- The lack of computational power, forcing the use of very few examples on each training batches might affected the learning process of the model.

4.1.2 Experiment 2: Direct Classification

The classification model also failed to converge on the mammography data.

- Three-Class Problem (Normal, Benign, Malignant): The model achieved an accuracy of only 47%, which is slightly better than random chance (33%) but indicates a failure to learn distinguishing features.
- Two-Class Problem (Benign vs. Malignant): When simplifying the problem, the accuracy was 52%, effectively random guessing.

In stark contrast, the exact same EfficientNet-B0 architecture achieved 100% accuracy on the Oxford Flowers dataset. the use of this dataset was a random choice sanity purposes.

This discrepancy is a critical finding. The model's success on a standard dataset confirms that the architecture and implementation were sound to certain degree. The failure on the mammography data, therefore, points directly to issues with the complexity of the problem itself:

- The features that distinguish between benign and malignant tumors, or between normal tissue and a tumor, might be incredibly subtle and complex, requiring a large amount of data to learn. The Oxford Flowers dataset has more distinct, high-level features differentiating classes.
- The dataset being imbalanced, with far more normal and benign cases than malignant ones. This might have introduce biases to the model making it difficult to learn the features of the minority class.
- Mammograms images may exhibit high variability within the same class (e.g., tumors have different shapes, sizes, and textures), making the learning task much harder than for the more uniform classes in the Flowers dataset.

The results from both experiments demonstrate that the primary limitation of this project was the dataset and the complexity of Mammograms images. State-of-the-art deep learning models are data-hungry and computationally demanding, and their performance is fundamentally capped by the quality and quantity of the data they are trained on. This research serves as a practical case study in the failure of complex models when applied to insufficient data in the medical domain.

4.1.3 Research Questions

- Can Meta Pseudo Labels be a suitable approach for CAD in breast cancer diagnosis? This study was unable to determine the suitability of Meta Pseudo Labels. The technique's effectiveness is predicated on a student model that can achieve at least a minimal level of competence through initial supervised training. When the data is so scarce or complex that even this first step fails, the Meta Pseudo Labels framework cannot be successfully applied.

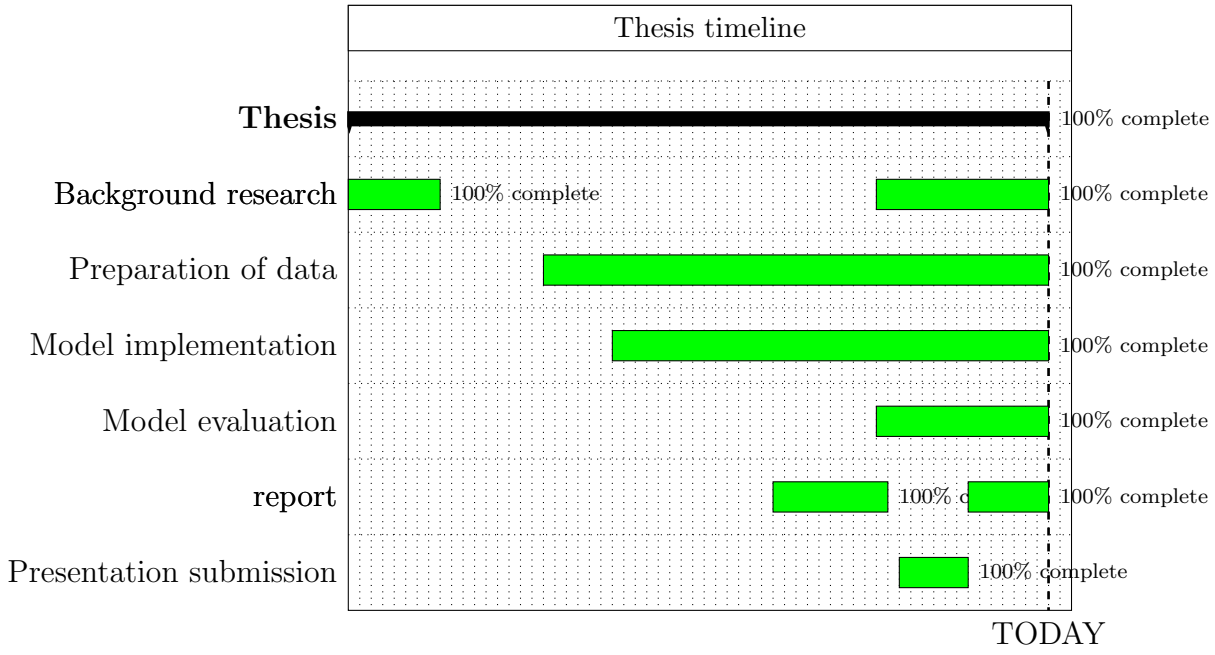
- Can Meta Pseudo Labels be a suitable approach against database bias? Similarly, this could not be evaluated. Meta Pseudo Labels is designed to improve generalization and thus combat database bias by learning from a wider, unlabeled data distribution. However, this process requires the model to first learn the fundamental task, which was not achieved.

4.2 Project Schedule

It was hard to keep up with the schedule proposed. From the day this project was proposed 14 of June 2021. To be working for the next 9 weeks to finish the project. The project was broken down into sub-tasks, that were proposed to be followed to make it easier and to guarantee that the research would go smoothly and that its aim would be reached. But over the next five weeks from the proposed schedule was ignored in a way.

It would be nice to lie and say every went as planned but that would be far from the truth. The feedback or guidance that one may have expected where never matched.

This is all to say that there is nothing to say there was no project management done about the project management. And this is also because mistakes were made during the research with unaccepted complications. And this is the nature o research of doing that challenge one's ability in a topic. A realistic approximation of the project timeline would similar to the chart bellow 1.



4.3 Risk Management

In this project, the risk involved the data explosion, unmountable google drive caused by augmented data saved on the first layer of the drive. Google colab pro, RAM crash, GPU with not enough memory. All these errors did affect the schedule but not at the level that would stop the project. Most times a restart of the notebook was required.

Risk	**Mitigation**	**Solutions**	**Impact**
Data explosion	Augmented images saved on first Google drive's layer	To delete each image or change to different account	High
Memory crash	Google colab notebook crash	Using subscription plan	Medium
Data loss	Google colab notebook crashing does not save the notebook and other information	Constant file backup	Medium

Table 4: *Risks and Corresponding Solution*

4.4 Quality Management

Since the schedule that was not followed, not weekly meetings and feedback to keep the project on track. There was no strategy to employed employed. The one strategy employed was to keep working to able to submit the work before deadline and to follow the proposed research. If these can be considered as quality management.

4.5 Social, Legal, Ethical and Professional Considerations

When utilising sensitive medical data, such as mammograms, there are ethical concerns. However, a comprehensive ethics proposal was filed and accepted early in the research. The data utilised in this study were collected from internet sources that made them freely available to the public. All of the data are devoid of any sensitive aspects that might harm someone, and the legal ownership of the data has been acknowledged.

Conclusion and Future Work

5.1 Conclusion

This thesis set out to develop a deep learning pipeline for breast cancer diagnosis. Through two distinct experiments—tumor segmentation and direct classification—it was demonstrated that the chosen state-of-the-art models were unable to learn from the limited and complex mammography dataset provided. The successful application of the same classification model to a benchmark dataset confirmed that the failure might have been data-driven, and maybe not so much model-driven, although there's no way of having certainties about it.

The key takeaway is a practical illustration of the critical need for large, high-quality, and balanced datasets in medical imaging AI, and that even possibly different architectures, approaches to the problem and a lot of computational power, given that it took 7 hours to run one epoch for the segmentation section while using Google Colab.

5.2 Future work

For future work, it would nice with better equipment to explorer further this idea, or ideas such as:

- Exploring models like R-CNN detection model.
- Attention U-net for segmentation
- Using more Advanced Data Augmentation beyond basic augmentations (rotation, flips) and implement techniques like Generative Adversarial Networks (GANs) to synthesize realistic mammogram images.
- Using models pre-trained on large-scale medical imaging datasets. This would provide the model with a much more relevant feature-learning head start.

.

Student Reflections

The objective was not achieved since the model did not perform well in the training dataset. It was almost impossible to answer the questions that were proposed. It is disappointing, to say the least, with the outcome. But still, one may think that it's still important to explore new ideas, even if sometimes the results are far from ideal. For those reasons, and for the reasons that let this project go down this path, someone may pick up where this project stopped.

This project was a good way to learn more and explore further concepts in the field of data science and machine learning. Even if the project schedule was not followed, it does show its importance. It seems like this is a problem in research overall. There is a wish that this was a better reflection, but it is not, at least at the moment of writing. And the effect of most things takes time. Hopefully, this project will have a positive impact.

Bibliography

(23, March, 2020). Meta pseudo labels deepai. https://deepai.org/publication/meta-pseudo-labels.

Ajay, K. and Saurabh, J. (2018). Why cad failed in mammography. *Journal of the American College of Radiology*.

American Cancer Society (2021b). U.s. breast cancer statistics. https://www.breastcancer.org/symptoms/understand_bc/statistics. Online; accessed 19 July 2021.

American Cancer Society (22, April, 2021a). American cancer society recommendations for the early detection of breast cancer. https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html. Online; accessed 19 July 2021.

American Cancer Society (25, March, 2002). The history of cancer.

Ashenafi, D. Y., Debelee, T., and Rahimeto, S. (2021). Automatic pectoral muscle removal in mammograms. *Evolving Systems*, 12.

Cancer Research UK (2021). Breast cancer statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/. Online; accessed 19 July 2021.

C.D. Lekamlage, F. Afzal, E. W. and Cheddad, A. (09, 10, 2020). Mini-ddsm: Mammography-based automatic age estimation. *3rd International Conference on Digital Medicine and Image Processing (DMIP 2020)*.

CDC (14, September, 2020). Breast Cancer what are the risk factors. https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm. Online; accessed 19 July 2021.

Chan, H.-P., Samala, R. K., Hadjiiski, L. M., and Zhou, C. (07, February, 2020). Deep learning in medical image analysis. *Advances in Experimental Medicine and Biology 1213*.

Dhungel, N., Carneiro, G., and Bradley, A. P. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical Image Analysis*, 37:114–128.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4):198–211. Computer-aided Diagnosis (CAD) and Image-guided Decision Support.

Domingues, I., Sales, E., and Pereira, W. (2012). Inbreast-database masses characterization.

et al, J. S. (1994). The mammographic image analysis society digital mammogram database exerpta medica. international congress series 1069. pages 375–378.

Google AI Blog (14, july, 2021). From vision to language: Semi-supervised learning in action...at scale. https://ai.googleblog.com/2021/07/from-vision-to-language-semi-supervised.html. Online; accessed 8 August 2021.

Gsunit. Gsunit/pectoral-muscle-removal-from-mammograms: Algorithm to segment pectoral muscles in breast mammograms. https://github.com/gsunit/Pectoral-Muscle-Removal-From-Mammograms.

Heath, M. D., Bowyer, K., Kopans, D., and Moore, R. H. (2007). The digital database for screening mammography.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network.

Hu, J., Shen, L., and Sun, G. (2017). Squeeze-and-excitation networks. *CoRR*, abs/1709.01507.

Hummel, R. (1977). Image enhancement by histogram transformation. *Computer Graphics and Image Processing*, 6(2):184–195.

Ingram, A. (22, March, 2011). *Breast Cancer Pioneer - Was the First Person to Use X-rays to Study Breast Cancer*. Science Heroes.

Kaur, H., Virmani, J., Kriti, and Thakur, S. (2019). Chapter 10 - a genetic algorithm-based metaheuristic approach to customize a computer-aided classification system for enhanced screen film mammograms. In Dey, N., Ashour, A. S., Fong, S. J., and Borra, S., editors, *U-Healthcare Monitoring Systems*, Advances in Ubiquitous Sensing Applications for Healthcare, pages 217–259. Academic Press.

Kido, S., Hirano, Y., and Mabu, S. (07, February, 2020). Deep learning for pulmonary image analysis: Classification, detection, and segmentation. *Advances in Experimental Medicine and Biology 1213*.

Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., and Li, J. (2019). Dice loss for data-imbalanced NLP tasks. *CoRR*, abs/1911.02855.

Long, J., Shelhamer, E., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038.

Medich, D. and Martel, C. (2006). *Medical Health Physics*. Medical Physics Publishing.

NHS (28, October, 2019). Breast cancer in women. https://www.nhs.uk/conditions/breast-cancer/. Online; accessed 19 July 2021.

Oliver, A., Freixenet, J., Martí, J., Pérez, E., Pont, J., Denton, E. R., and Zwiggelaar, R. (2010). A review of automatic mass detection and segmentation in mammographic images. In *Medical image analysis*, volume 10, page 87–110. ScienceDirect.

Olson and James, S. (2002). *Bathsheba's breast: women, cancer & history*. The Johns Hopkins University Press.

Pati, D. P. and Panda, S. (2020). Feature extraction and enhancement of breast cancer mammogram noisy image using image processing. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pages 1–5.

Pham, H., Xie, Q., Dai, Z., and Le, Q. V. (2020). Meta pseudo labels. *CoRR*, abs/2003.10580.

Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.

Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., and Chen, L. (2018). Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381.

Tan, M., Chen, B., Pang, R., Vasudevan, V., and Le, Q. V. (2018). Mnasnet: Platform-aware neural architecture search for mobile. *CoRR*, abs/1807.11626.

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946.

Tang, J., Rangayyan, R. M., Xu, J., Naqa, I. E., and Yang, Y. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances. *IEEE Transactions on Information Technology in Biomedicine*, 13(2):236–251.

Wang, H., Tian, J., Li, S., Zhao, H., Tian, Q., Wu, F., and Li, X. (2021). Unsupervised domain adaptation for image classification via structure-conditioned adversarial learning. *CoRR*, abs/2103.02808.

Winchester, David, J., Winchester, D. P., Hudis, C. A., and Norton, L. (2006). *Breast Cancer*. TPMPH-USA.

Xie, Q., Hovy, E. H., Luong, M., and Le, Q. V. (2019). Self-training with noisy student improves imagenet classification. *CoRR*, abs/1911.04252.

Yaffe (2010). Detectors for digital mammography. *Diagnostic Imaging*, page pp. 13–31.

Yaffe, M. J. and Mainprize, J. G. (2004). Detectors for digital mammography. *Technology in Cancer Research & Treatment*, 3(4):309–324. PMID: 15270582.

Yanase, J. and Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, 138:112821.

Zoph, B. and Le, Q. V. (2016). Neural architecture search with reinforcement learning. *CoRR*, abs/1611.01578.

Appendix

A.1 Meeting Records

Presentation of the project topic and project discussion on Friday, November June, 2021, from 10:30 a.m. to 11:00 a.m.

A.2 Ethics application

☑	I submitted my ethics application and my application has been approved. I include my ethics certificate in the appendix as evidence.
☐	I submitted my ethics application and my application is currently under review.
☐	I have not submitted my ethics application.

Table 5: *Ethics application process indication.*

Figure 17: *Illustration for the certificate of ethical approval*