

6-2022

DEEP LEARNING BASED SEGMENTATION AND CLASSIFICATION FOR IMPROVED BREAST CANCER DETECTION

Bitra Khosrow Asadi Khosrow Asadi

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/all_theses

 Part of the [Engineering Commons](#)



MASTER THESIS NO. 2022:60

College of Engineering

Department of Electrical and Communication Engineering

**DEEP LEARNING BASED SEGMENTATION AND
CLASSIFICATION FOR IMPROVED BREAST CANCER
DETECTION**

Bita Khosrow Asadi

CLASSIFICATION MODEL

```
[ ] from sklearn.model_selection import train_test_split
from tensorflow import keras
from tensorflow.keras import layers
from keras.layers import Input, Conv2D
from keras.models import Model
from keras.utils.np_utils import to_categorical

#from keras.models import Sequential
#from keras.optimizers import Adam
```

June 2022

United Arab Emirates University

College of Engineering

Department of Electrical and Communication Engineering

**DEEP LEARNING BASED SEGMENTATION AND
CLASSIFICATION FOR IMPROVED BREAST CANCER
DETECTION**

Bitra Khosrow Asadi

This thesis is submitted in partial fulfilment of the requirements for the degree of Master
of Science in Electrical Engineering

June 2022

**United Arab Emirates University Master Thesis
2022: 60**

Cover: Benign tumor generated after programming experiment.

(Photo: By Bitā Khosrow Asadi)

© 2022 Bitā Khosrow Asadi, Al Ain, UAE

All Rights Reserved

Print: University Print Service, UAEU 2022

Declaration of Original Work

I, Bitra Khosrow Asadi, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled “*Adaptive Feature Selection and Classification to Improve Breast Cancer Detection*”, hereby, solemnly declare that this is the original research work done by me under the supervision of Dr. Qurban A. Memon, in the College of Engineering at UAEU. This work has not previously formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature: Bitra

Date: 21-5-2022

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

- 1) Advisor (Committee Chair): Dr. Qurban Memon

Title: Associate Professor

Department of Electrical Engineering

College of Engineering

Signature 

Date 22/09/2022

- 2) Member: Dr. Rachid Erroussi

Title: Associate Professor

Department of Electrical Engineering

College of Engineering

Signature 

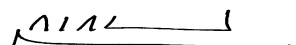
Date 29/09/2022

- 3) Member (External Examiner): Dr. Uvais Qidwai

Title: Associate Professor

Department of Electrical Engineering and Computer Engineering

Institution: Qatar University

Signature 

Date 30/09/2022

This Master Thesis is accepted by:

Acting Dean of the College of Engineering: Professor Mohamed Al-Marzouqi

Signature Mohamed AlMarzouqi

Date 20/10/2022

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature Ali Hassan

Date 20/10/2022

Abstract

Breast Cancer is a leading killer of women globally. It is a serious health concern caused by calcifications or abnormal tissue growth in breast. Doing a screening and identifying the nature of the tumor as benign or malignant is important to facilitate early intervention, which drastically decreases mortality rate. Usually, it uses ultrasound images, since they are easily accessible to most people and have no drawbacks as such, unlike in the other most famous screening technique of mammograms where in some cases you may not get a clear scan. In this thesis, the approach to this problem is to build a stacked model which makes prediction on the basis of the shape, pattern, and spread of the tumor. To achieve this, typical steps are pre-processing of images followed by segmentation of the image and classification. For pre-processing, the proposed approach in this thesis uses histogram equalization that helps in improving contrast of the image, making the tumor stand out from its surroundings, and makes it easier for the segmentation step. Through segmentation, the approach uses UNet architecture with a ResNet backbone. The UNet architecture is made specifically for the bio-medical imaging. The aim of segmentation is to separate the tumor from the ultrasound image so that the classification model can make its predictions from this mask. The metric result of F1-score for the segmentation model turned out to be 97.30%. For classification, CNN base model is used for feature extraction from provided masks. These are then fed into a network and the predictions is done. The base CNN model used is ResNet50 and the neural network used for the output layer is a simple 8-layer network with ReLU activation in the hidden layers and softmax in the final decision-making layer. The ResNet weights are initialized from training on ImageNet. The ResNet50 returns 2048 features from each mask. These are then fed into the network for decision making. The hidden layers of the neural network have 1024, 512, 256, 128, 64, 32, 10 neurons respectively. A classification accuracy achieved for the proposed model was 98.61% with F1 score of 98.41%. The detailed experimental results are presented along with comparative data.

Keywords: UNet, ResNet, F1-score, CNN, Softmax, ImageNet, ResNet50.

Title and Abstract (in Arabic)

اختيار الميزات التكيفية وتصنيفها لتحسين اكتشاف سرطان الثدي

الملخص

يعتبر سرطان الثدي من الأسباب الرئيسية لوفاة النساء على مستوى العالم. انها مشكلة صحية خطيرة ناتجة عن التكتلات او نمو الانسجة غير الطبيعي في الثدي. يعد إجراء الفحص وتحديد طبيعة الورم على أنه ورم حميد أو خبيث أمرًا مهمًا لتسهيل التدخل المبك، مما يقلل بشكا كبير من معدل الوفيات. يمكن أن يساعد نموذج التعلم الآلي في التشخيص المبكر للمرض لمساعدة المرضى على تحديد ما إذا كانوا بحاجة إلى تدخل طبي أم لا. عادةً ما تستخدم صور الموجات فوق الصوتية، حيث يسهل الوصول إليها لمعظم الناس وليس لها عيوب على هذا النحو، على عكس تقنية الفحص الأكثر شهرة لتصوير الثدي بالأشعة السينية حيث قد لا تحصل في بعض الحالات على مسح واضح.

في هذه الأطروحة، يتمثل نهج هذه المشكلة في بناء نموذج يقوم بالتنبؤ على أساس شكل ونمط وانتشار الورم. لتحقيق ذلك، فإن الخطوات النموذجية هي المعالجة المسبقة للصور متبوعة بتجزئة الصورة والتصنيف. للمعالجة المسبقة، استخدم النهج المقترح في هذه الأطروحة معادلة الرسم البياني التي تساعد في تحسين تباين الصورة، مما يجعل الورم بارزًا عن محيطه، ويسهل خطوة التجزئة. من خلال التجزئة، يستخدم النهج بنية UNet مع العمود الفقري ResNet. تم تصميم بنية UNet خصيصًا للتصوير الطبي الحيوي. الهدف من التجزئة هو فصل الورم عن صورة الموجات فوق الصوتية حتى يتمكن نموذج التصنيف من عمل تنبؤاته من هذا القناع. كانت النتيجة المترية لدرجة F1 لنموذج التجزئة 97.30%. بالنسبة للتصنيف، يتم استخدام نموذج قاعدة CNN لاستخراج الميزات من الأفتعة المتوفرة. ثم يتم إدخالها في شبكة ويتم تنفيذ التوقعات. نموذج CNN الأساسي المستخدم هو ResNet50 والشبكة العصبية المستخدمة لطبقة المخرجات عبارة عن شبكة بسيطة من 8 طبقات مع تنشيط ReLU في الطبقات المخفية و softmax في طبقة اتخاذ القرار النهائية. تتم تهيئة أوزان ResNet من التدريب على ImageNet. يُرجع ResNet50 2048 ميزة من كل قناع. ثم يتم إدخالها في الشبكة لاتخاذ القرار. تحتوي الطبقات المخفية للشبكة العصبية على 1024، 512، 256، 128، 64، 32، 10 خلايا عصبية على التوالي. كانت دقة التصنيف المحققة للنموذج المقترح 98.61% مع درجة F1 98.41%. يتم عرض النتائج التجريبية التفصيلية مع البيانات المقارنة.

مفاهيم البحث الرئيسية: المعالجة المسبقة، الطبي الحيوي، الشبكة العصبية.

Acknowledgements

My thanks go to Dr. Qurban Memon who introduced me to this field of computer vision with applications in medical area and whose endless ideas and encouragement led to this and most other studies in which I have been involved.

I would like to thank my committee for their guidance, support, and assistance throughout my preparation of this thesis. I would like to thank the chair and all members of the Department of Electrical Engineering at the United Arab Emirates University for assisting me all over my studies and research.

Dedication

To my beloved parents and family who helped me along the way.

Table of Contents

Title.....	i
Declaration of Original Work.....	iii
Approval of the Master Thesis	iv
Abstract.....	vi
Title and Abstract (in Arabic).....	vii
Acknowledgements.....	viii
Dedication.....	ix
Table of Contents.....	x
List of Tables	xi
List of Figures.....	xii
List of Abbreviations	xiii
Chapter 1: Introduction.....	1
1.1 Literature Review	3
Chapter 2: Material and Methods	7
2.1 Dataset	7
2.2 Pre-processing.....	8
2.3 Image Features.....	9
2.4 Feature Selection	10
2.5 Training.....	11
2.6 Classification	12
2.7 Evaluation Metrics.....	14
2.8 Proposed Approach.....	16
Chapter 3: Experimental Results	19
3.1 Pre-processing.....	19
3.2 Experimental Platform.....	21
3.3 Segmentation Results	22
3.4 Classification Results	25
Chapter 4: Discussion and Conclusions	33
References.....	35

List of Tables

Table 1: Comparing ResNet50, ResNet101, ResNet152	26
Table 2: Comparing VGG16 and VGG19.....	29
Table 3: Comparing MobileNet and MobileNet_V2	30

List of Figures

Figure 1: Model block diagram	17
Figure 2: Ultrasound images.....	21
Figure 3: Segmentation results for sample images.....	22
Figure 4: Training and validation loss.....	23
Figure 5: Tracking accuracy during training and validation	23
Figure 6: Tracking precision - training and validation.....	24
Figure 7: Tracking recall - training and validation.....	24
Figure 8: Tracking F1 - score during training and validation	25
Figure 9: Training and validation loss function.....	26
Figure 10: Tracking accuracy during training and validation	27
Figure 11: Tracking Precision - training and validation.....	27
Figure 12: Tracking recall - training and validation.....	28
Figure 13: Tracking F1 - score during training and validation	28

List of Abbreviations

ANN	Artificial Neural Network
AP	Average Precision
AUC	Area Under the Curve
BCDR	Breast Cancer Digital Repository
CAD	Computer-Aided Design
CNN	Convolution Neural Network
DBF	Decision Boundary Features
DM	Digital Mammography
DL	Deep Learning
ERFNet	Efficient Residual Factorized Network
FN	False Negative
FP	False Positive
GA	Genetic Algorithm
GLCM	Gray-Level Co-occurrence Matrix
IoU	Intersection of Union
IRMA	Image Retrieval in Medical Application
mAP	Mean Average Precision
MLO	Mediolateral Oblique
PCA	Principle Component Analysis
REF	Recursive Feature Selection
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SVM	Support Vector Machine
TN	True Positive

TP	True Positive
US	Ultrasound
WBC	Wisconsin Breast Cancer
WDBC	Wisconsin Diagnostic Breast Cancer
WPT	Wavelet Packet Transform
YOLO	You Only Look One

Chapter 1: Introduction

Based on research conducted at a cancer research institute (Islam, Iqbal, Haque, & Hasan, 2017), breast cancer is common and the most serious disease affecting women around the world, with 2 million new cases diagnosed in 2018. When compared to other types of cancers, breast cancer is the fifth leading cause of death in women. Breast cells affected by cancer grow cancer tissues abnormally, thereby increasing the affected cell rate gradually. Breast lesions that can be detected by mammography fall into two categories: malignant and benign lesions. Benign lesions are a heterogeneous group that includes inflammatory lesions and developmental abnormalities. Diagnosis of these lesions can be made using mammography and, therefore, surgery is not needed because most of these lesions are not linked to an increased risk of later breast cancer. Breast cancer is divided into two types based on where the cancer begins. They could be breast lesions that are either ductal or lobular in shape. Mammography can show malignant lesions as microcalcifications, infiltrative masses or irregularly shaped masses. A mass may be considered as an abnormal breast tissue area with shape and edges that differ it from the rest of the breast tissue on a mammogram. A mass may be seen either calcified or non-calcified. Lumps may be many things, including cysts that are non-cancerous or a liquid filled capsule, and solid tumors that are non-cancerous, but they may also be a cancer symptom.

The malignant calcifications frequently found in clusters, are small, vary in size and shape, are angular, and have irregular shapes (Valvano et al., 2019). Breast cancer is that type of cancer, which is made up of cells that are splitting apart. Some breast cancers are cancerous and can spread to other parts of the body, while other breast cancers are benign and will not spread. Despite substantial research by medical professionals and academics, there are no viable solutions for giving a suitable approach to a much-anticipated therapy and stringent proof of breast cancer prevention. Furthermore, certain significant malignant cells linked with cancer appear to be aggressive, and present bigger risk to patients' lives since they have the greatest risk of infecting other organs of the body. This type of cancer must thus be diagnosed and treated as soon as possible. Breast cancer detection, however, must be automated and intelligent enough due to the factors such as: a) the unreliability of

competence due to involvement of humans in detection, b) the degree of mistake in diagnosis, c) consumes time, d) overworked radiologists and e) inaccuracy in detection and prognosis. Furthermore, because manually detecting breast cancer can take months, an intelligent system to recognize cancer is essential because the stage of cancer infection in the localized stage may proceed to the critical stage when survival chances are unachievable.

The five-year survival rate of breast cancer in its initial phase is around 80-90% in countries with advanced medical technology, but it falls to 24% for diagnosis of cancer at its initial stage (Krizhevsky, Sutskever, & Hinton, 2012). Breast cancer has been diagnosed using a variety of invasive-based techniques. For testing, breast tissues are collected using biopsy technique (Ramadan, 2020), with extremely accurate results. A biopsy of that issue, on the other hand, is painful for the patient. The mammogram (Zou et al., 2019), which is used to diagnose breast cancer, is another breast cancer diagnosis technique. This can also be used to reduce the number of unnecessary biopsies (Dhungel, Carneiro, & Bradley, 2016). A 2-dimensional projected image of the tissue is created using this technique. The mammogram technique, on the other hand, is ineffective in detecting benign cancer. It is a very complex test (Gardezi, Elazab, Lei, & Wang, 2019), which can provide very good results for three-dimensional images with dynamic functionality. It is another technique for breast cancer diagnosis. These invasive-based breast cancer diagnosis techniques are difficult to carry out and do not accurately and effectively diagnose the disease. Furthermore, the results from these techniques take longer time to produce results (Dhungel, Carneiro, & Bradley, 2015). A noninvasive-based techniques, like machine learning, are considered effective and more reliable to resolve these complexities in breast cancer diagnosis. The non-invasive use of machine learning has attracted many to investigate its potential strength in many applications (Memon, 2013, 2019; Memon & Laghari, 2006; Valappil & Memon, 2021) including healthcare. Some have reported success in classifying cancer tissues as benign or malignant. The research on these non-invasive techniques for cancer diagnosis in breast tissues is discussed in the next section.

1.1 Literature Review

Digital mammography carries few shortcomings, like a low sensitivity found especially with dense breasts. The techniques (Ultrasound, Mammography, MRI), however, must be interpreted by a radiologist manually. Many artificial intelligence algorithms are thus becoming more popular because they perform better than traditional methods in image recognition tasks (Jadoon, Zhang, Haq, Butt, & Jadoon, 2017). Various methods of breast image classification, such as computer-aided diagnostic/detection systems based on machine learning or modern deep learning systems, have been used to facilitate physicians in interpreting medical images (Noguchi, Nishio, Yakami, Nakagomi, & Togashi, 2020). The study based on CAD system (Ramadan & El-Banna, 2020) found that it is possible to obtain 92% accuracy in classifying mammography readings. In general, the DL-CAD system focuses on convolutional neural networks. It is widely used intelligent image analysis model with accurate detection of the cancer (Urbanowicz, Meeker, La Cava, Olson, & Moore, 2018). The generation of technologies used in DL-CAD system are better equipped in solving problems, which are difficult to be handled by traditional methods. These problems encompass learning from complex data, recognizing images, diagnosing medical conditions, and enhancing images (Haq et al., 2018). The image analysis techniques used include preprocessing, selecting a Region of Interest (ROI) known as segmentation, extraction of features, followed by classification. In the recent past, key researchers in this arena have widely pioneered automated and intelligent breast cancer detection systems including machine learning. This invention is primarily concerned with calculating the maximum number of parameters that might be positively upgraded throughout the diagnostic procedure. Furthermore, an intelligent breast cancer detection that is also automated is chosen over manual breast cancer detection because intelligent systems have the capacity to imitate the human brain's unique behavior. In Artificial Neural Network (ANN) based cancer detection, this mimicry of the human brain is done by embedding an extra capacity for approximating and resolving non-linear and difficult issues, which aids in enabling the understanding of systems' mathematical representations caused by the process. Because the former's performance is dependent on parameter optimization, it can be easily judged that the prediction due to ANN-based cancer diagnosis is noteworthy and above standard statistical detection approaches. Furthermore,

the rate of ANN-based cancer diagnosis performance is determined by key parameters, such as: a) selection of features, b) algorithms used for learning, c) number of hidden layers, d) nodes used in each hidden layer, and e) weights of the factors initialized for optimization. The first and most important parameter to study in the creation of ANN-based cancer detection systems is feature selection. The importance of the subset of features in the design of ANN-based detection techniques has been demonstrated by a large body of research. It is also clear from studies that the feature subset and design elements used in the development of the ANN inspired detection carry a reciprocal influence. As a result, improving feature subset and corresponding design parameters used in the process of ANN-based cancer diagnosis becomes necessary. Because of these capabilities in artificial neural networks, it may be utilized to analyze medical pictures (Pérez, Guevara, Silva, Ramos, & Loureiro, 2014). Several unique breast cancer detection strategies have been presented by certain potential researchers for parallelized optimization of feature subset and corresponding design parameters. However, such approaches are not capable of meeting challenges, which develop as a result of the dynamics in feature subset during complexity in design processes and when requirements are coupled (Pérez, Guevara, & Silva, 2013). As a result, the primary goal in hybridizing the diagnostic process employing different algorithms in the ANN design becomes critical. Furthermore, the hybridizing of ANN design processes employing approaches such as brute force, destructive or constructive network, and trial-error schemes has been shown to be inefficient as each one is incapable of simultaneously managing large quantities of design parameters (Ibrahim, Yousri, Ismail, & El-Makky, 2014). Deep learning is a powerful tool used in imaging systems especially used in medical field to detect tumors and cancers (Haq et al., 2019). The Convolution Neural Network (CNN) classifier sees artifacts and pectoral muscle in mammograms as distractions, so they must be removed. Before feeding the mammograms to the CNN as input images, image cropping is done manually to isolate areas or interest regions. This isolation process has been automated by many researchers. The authors of (Zhang, Zou, Zhou, & He, 2018) employed genetic algorithms to automatically figure out interest regions using fitness value of area under the curve defined by receiver operating characteristic.

The ability for global search optimization, optimization algorithms such as Genetic

Algorithms (GA) have shown to be very suited for the optimization of design parameters. GA has also been shown to be more effective because of its ability to maintain the accuracy of Support Vector Machine (SVM) technique under optimum feature subset and parameter generation for implementation (Shao et al., 2011). However, GA-ANN-based cancer diagnosis has drawback of being tracked towards its local optimal spots while parameters are being optimized.

The authors developed a deep learning-based technique for extracting characteristics from histopathology pictures of the breast, with a specific focus on mitosis detection. The breast mitosis was recognized using the suggested model, which retrieved features from CNN and sent them to the support vector machine for training. AlexNet was used to build CNN in order to differentiate benign mitosis from malignant mitosis based on histological pictures (Mahmood, Arsalan, Owais, Lee, & Park, 2020). For the identification of mitosis from breast histology slides, a deep cascade network was proposed. Then, for mitosis classification, a CaffeNet model (Dabeer, Khan, & Islam, 2019) was fine-tuned and pre-trained using ImageNet pictures. The outputs were provided in the form of various scores or probabilities. Using annotation (Dabeer et al., 2019), a Stacked Sparse Autoencoder (SSAE)-based approach was used to classify nuclei from breast histopathology images. The greedy technique was used to optimize SSAE, in which only one hidden layer at a time was trained, and the previous layer's output became the input to the next hidden layer. In addition to histopathological image-based breast cancer detection, screening mammography images were employed in the research for breast cancer detection. In case, the training data for a CNN model turned out to be inadequate, the authors (Shin et al., 2016) used a transfer learning approach to train it. It was feasible to detect the mass from the available mammograms using this deep learning model. In a separate research (Dhungel, Carneiro, & Bradley, 2017), Dhungel et al. suggested a mass detection approach based on a cascade of random forest classifiers and deep learning.

The authors (Martynenko & Bück, 2018) developed a breast cancer detection system based on Wisconsin Breast Cancer datasets and using a genetic algorithm for selection of features and random forest for classification. On selected characteristics chosen by the GA method, the Random Forest (RF) achieved above 95 percent classification accuracy. For the classification of benign and malignant breast cancers, Zheng et al. (Zheng, Yoon, & Lam,

2014) employed a K-means approach towards feature extraction and integrated it with support vector machine. The suggested method produced good classification accuracy while consuming little processing resources. Ramadevi (Ramadevi, Rani, & Lavanya, 2015) utilized combination of principal component analysis with several classifiers involving various cancer datasets and obtained relatively high accuracy. In another work, the author suggested a strategy for detecting breast cancer using a memetic Pareto artificial neural network in (Abbass, 2002). The experimental findings showed that the suggested approach has a high level of classification accuracy and a short processing time. Liu et al. (Liu, Wang, & Zhang, 2009) introduced a decision tree-based breast cancer prediction algorithm that used the under-sampling technique to balance the training data. The results of the experiments reveal that the suggested approach has a high level of accuracy. The author (Onan, 2015) devised a clever method for detecting breast cancer. For instance, selection, the author employed fuzzy-rough, and feature selection by consistency. He employed the fuzzy-rough nearest neighbor method to identify breast cancer. The last and not the least work (Mohan, Bhattacharya, Kaluri, Feng, & Tariq, 2020) developed a breast cancer prediction method based on particle swarm optimization combined with non-parametric kernel density estimation. The rest of the thesis is structured as follows. In the next section, materials and methods are proposed, where dataset selected, preprocessing done on images, the proposed model and performance metric done on results are presented. The chapter three presents results, and the chapter four concludes the findings of this thesis.

Chapter 2: Materials and Methods

2.1 Dataset

Not only should an automated identification technique be able to match a sickness to one of the known issues, but it should also be able to reject diseases that were not part of the training set. Benchmark datasets are accessible, however obtaining a live dataset is quite challenging. This is because gathering a real-life event and then imagining it in a lab or in the field takes a lot of time and work. When existing databases are examined, it becomes evident that they were developed by a small group of people and contain specimen from a limited geographic region. As stated in previous section, the occurrence of large variability in photographs renders these datasets unsuitable for widespread usage. As a result, sufficient training data that incorporates all of the traits and changes in breast cancer is required for realistic issue categorization. Furthermore, studies should include more realistic pictures of various circumstances that have yet to be included in subsequent study, such as numerous and overlapped scenarios. Images should be shot in a natural setting as well as with complicated backgrounds, in various lighting situations, and at various times of the day to cover all of the bases. The examples of breast mammography pictures include, for example, the Mini-MIAS dataset (Mohan et al., 2020) that has 208 normal photos, 62 benign images, and 52 malignant cancer images. Because most DL-CAD systems demand a great quantity of data, the number of public medical pictures is growing. Thus, deep learning algorithms are applied to classify digitized mammograms, such as obtained from DDSM (Digital Database for Screening Mammography), which contains 2620 patients and includes mediolateral oblique and craniocaudal images (Araújo et al., 2017). The other examples include Image Retrieval in Medical Application (IRMA), which has 736 biopsies that have lesion in 344 patients and includes mediolateral oblique and craniocaudal images (Aswathy & Jagannath, 2017), INbreast with 419 cases for detection and diagnosis, including mediolateral oblique and craniocaudal images of 115 patients (Li et al., 2018), Breast Cancer Digital Repository (BCDR) with 322 digitized mediolateral oblique images of 161 patients for classification (Yousefi, Ting, Mirhassani, & Hosseini, 2013).

The developers at the University of Wisconsin generated the dataset "Wisconsin

Diagnostic Breast Cancer (WDBC)," that is available at the UCI machine learning repository (Fan, Upadhye, & Worster, 2006). The repository contains 569 people, 32 characteristics, and 30 real-value features. There are 212 malignant and 357 benign patients in the class. It consists of a 569 32 feature matrix. The dataset was separated into two parts: 70% for classifier training and 30% for classifier validation. The information and description of 569 occurrences are automatically generated using 32 dataset attributes and certain statistical metrics. In a dataset, there are 357 benign patients and 212 malignant ones.

2.2 Pre-processing

Database features are well recognized to have a major impact on the performance of a design scheme or a specific approach for processing. It also may devise a system, which produces perplexing or even erroneous outcomes. Because raw photos contain noise, preprocessing is the initial stage in the detection process. For example, removing undesirable picture information, often known as image noises, might improve the quality of an image to be utilized further. If this issue is not addressed appropriately, the categorization may contain several errors. In addition to errors, poor contrast between skin lesions and surrounding healthy skin, uneven boundaries, and skin artifacts such as hairs, skin lines, and black frames necessitate this preprocessing. The filters such as mean, Gaussian, median, or adaptive median, and wiener may all be used to remove speckle, Gaussian, Poisson, and salt and pepper noise (Delen, Walker, & Kadam, 2005).

A picture with hairs in it, as well as the lesion, may induce misdiagnosis. Gaussian blur and histogram equalization may also be used for clearing general noise from images. Histogram equalization helps in improving the contrast of the image by spreading out frequency intensity values. Pre-processing operations such as contrast adjustment, vignetting effect removal, color correction, picture smoothing, hair removal, normalization, and localization are designed to eliminate or alter image noises. More precision is achieved by combining the proper pre-processing processes.

2.3 Image Features

In picture categorization, the processing phase is very significant. The separation of regions of interest (masses, lesions, microcalcifications) from the picture backdrop is known as segmentation. The duties of specifying the region of interest, such as lesions, or initial boundary are completed using knowledge of radiologists in contemporary Computer-Aided Design (CAD) systems. Because several characteristics are employed to discriminate between malignant and benign tumors, the accuracy of segmentation does affect outcomes of CAD systems (contour, texture, shape of lesions). As a result, the characteristics may be retrieved actively if tumor segmentation is done with considerable precision (Khan, Choi, Shin, & Kim, 2008). This is why academics are turning to deep learning (DL) approaches, particularly CNN, to solve segmentation problems. Furthermore, DL-CAD systems are self-contained and able to model breast Ultrasound (US) and Digital Mammography (DM) information utilizing constraints without the need for human intervention. In order to train CNNs for DM and US instead of regions of interest, two methodologies have been used: (1) one with higher resolution (T.-N. Wang, Cheng, & Chiu, 2013) (2) at patch level (Chao, Yu, Cheng, & Kuo, 2014) pictures. The You Only Look One (YOLO) (Afshar, Ahmadi, Roudbari, & Sadoughi, 2015), Decoder Network (SegNet), Encode Network (UNet) (Sun, Wang, & Li, 2018), Generative Adversarial Network (GAN) (N. Ibrahim, Kudus, Daud, & Bakar, 2008), and Efficient Residual Factorized Network (ERFNet) (Sohail, Jiadong, Uba, & Irshad, 2019) are examples of modern network topologies that have been utilized to construct segmented areas. Another study used morphology of features in a picture. Because these procedures rely solely on the relative ordering of pixel values, they are best suited for binary image processing (Dai, Cheng, Bai, & Li, 2017), although they may also be used with greyscale pictures.

The features in the ultrasound image include the abnormalities in form of a cancerous tumor or benign cyst, lesions, tissues, and other masses. Since the dataset for breast cancer may provide us with corresponding masks of the ultrasound images, it can be deployed during training to learn to identify a malignant tumor from a benign abnormality. The segmentation may also be applied on ultrasound images to separate out the abnormalities from the image creating a mask that only has the abnormality highlighted in it. This mask

is later fed into the prediction model. Thus, the prediction model itself learns to distinguish between malignant tumors and benign cysts possibly by their different sizes and shapes. The dimension of feature vectors is significant throughout the classification process since it affects the classifier's performance. Morphological, model based, descriptor, and texture characteristics are the four feature categories seen in breast ultrasound pictures.

2.4 Feature Selection

After segmentation, the extraction of features and their selection are the subsequent steps to remove redundant and irrelevant data that is being processed. The form and margin of lesions, masses, and calcifications are used to create features in the ROI. Texture and morphologic characteristics (Kohli & Arora, 2018), descriptors, and model-based features (Aličković & Subasi, 2017) are some of the aspects that assist distinguish between benign and malignant tumors. The gray-level value and morphological characteristics are used to compute the majority of texture features from the full picture or Region of Interest (ROIs). Searching algorithms, gain ratio, random forest, chi-square test and recursive feature removal are some examples of conventional feature selection approaches (Rodriguez, Kuncheva, & Alonso, 2006). The wavelet packet transform, principal component analysis (Oyewola, Hakimi, Adeboye, & Shehu, 2016), grey-level co-occurrence matrix (Meesad & Yen, 2003), Gaussian derivative kernels (Birkett, Arandjelović, & Humphris, 2017), and decision boundary features (Furundzic, Djordjevic, & Bekic, 1998) are some of the other traditional feature extraction approaches. However, the size of the vectors affects both the performance and computational time (Cover & Hart, 1967) in classification schemes such as ANN or support vector machine. By reducing duplicate features, feature selection approaches minimize the feature space size, which boosts accuracy and reduces computation time (Fu et al., 2017). DL models, in particular, generate a collection of picture features from data (Fu et al., 2017), which have the benefit of extracting features and performing classifications immediately. For DL CAD systems, providing effective feature extraction and selection is critical; for example, many authors have provided CNNs that are capable of extracting features (Wang, Chu, & Xie, 2007).

Another feature selection process may be thought of as an approach that picks a feature subset from a larger feature set. The data has a huge amount of space hence subspace

feature selection is crucial for specificity of data. There are two advantages to feature selection. First, it enhances the classifier's accuracy, and second, it reduces the machine learning algorithm's calculation time owing to feature selection (Sarkar & Leong, 2000). The Recursive Feature Selection (REF) is another algorithm for feature selection, which fits a model by removing irrelevant ones until the desired quantity of features is obtained. The remaining qualities are the ones that have the most impact on the target class. The feature elimination approach that works recursively computes performance measures such as specificity, sensitivity, accuracy and F1-score by training the SVM model on the training dataset (Medjahed, Saadi, & Benyettou, 2013). In case of doing predictions using the ultrasound image, first, the approach may isolate the tumor from the image by creating the mask, which is done by the segmentation model. The mask is then passed as input to the classification model for identifying the tumor as benign or malignant. Consequently, any new mask passed to the classification model is classified as benign or malignant according to training. As the result, possible features that the model finds are when the model compares the size and shape of the tumor. Therefore, using pixels of imaging techniques is a feature selection of the image. It selects the pixels that make up the tumor and color them white while all other pixels are colored black.

2.5 Training

The dataset of mammograms used for training the model to predict tumor came from the Kaggle website. Kaggle is a renowned community of machine learning enthusiasts and data scientists, hosting competitions regularly, and providing datasets and projects to work on. It is a subsidiary of Google and is one of the most well-known communities in the data science field. So, this website is typically preferred for getting datasets. The dataset of Mammograms has 780 images including benign and malignant ultrasound images of women between the ages of 25-75 years. To increase the size of the dataset, data augmentation may be used to achieve better results. This may be done using a library known as Albumentations, the functions used for the same are VerticalFlip, HorizontalFlip, GridDistortion, Transpose, and RandomRotate90. The ImageNet dataset is another one to pre-train and for the weight initialization of the model. ImageNet is a large dataset of 14 million images across 200 different classes. The size of each image on average was 470*390 but cropped to 256*256. So, pre-training the model on that dataset

is the best way to have weights initialized for an image segmentation and image classification model. From this dataset, 2500 images were used for training, and 280 for image testing, so roughly a ratio of 90% to 10% was selected for training and testing.

2.6 Classification

Following the extraction and selection of characteristics, the ROI is classified into malignant and benign classifications using a classifier. Linear, ANN, Bayesian neural networks, decision trees, SVM, template matching, and CNNs are some of the most often used classifiers (Dalle, Leow, Racoceanu, Tutac, & Putti, 2008). Deep convolutional neural networks are hierarchical architectures trained on large datasets recently demonstrated impressive object identification and detection results, implying that they might enhance lesion diagnosis in both US and DM approaches. Lesion (Dundar et al., 2011), microcalcification, and mass categorization in DM and US images using CNN algorithms are of attention to a number of researchers. When it comes to deep learning and image processing, CNNs are the most extensively utilized neural networks. Convolution, pooling, and full-connection layers are the three types of layers in the CNN structure, which are layered in numerous layers. The number of layers, the learning rate, the activation function, the pooling layer for feature map extraction, the loss function, and the classification specific fully connected layers all influence the structure of a CNN. Furthermore, dropout and batch normalization are two strategies for increasing the performance of a CNN. Dropout is a regularization approach for preventing overfitting in CNN models. A normalization layer speeds up CNN training, and thus minimizes network initialization sensitivity.

CNN's performance has been amazing in recent years as it has grown deeper, with well-known networks with layers ranging from 7 to 1000. For breast cancer study, several of these cutting-edge architectures may be employed for transfer learning to range of applications, which are pre-trained on ImageNet.

Researchers from Oxford University published multiple versions of the very deep convolutional network Visual Geometry Group (VGG) (Arevalo, Cruz-Roa, Arias, Romero, & González, 2015). VGG16 is one of the finest networks and widely recognized for simplicity. This network's design is deep, mostly contains alternate convolution and

dropout layers. This network was the first to integrate many tiny 3x3 filters in a sequence in each convolutional layer to mimic the impact of bigger receptive fields. Although the network's design is basic, it is costly in terms of computation and memory since the expanding kernels result in longer calculation times and a larger model size.

The ResNet50 is one of the models presented by the Microsoft research team in residual learning for image recognition (Shin et al., 2016). This basic yet beautiful concept takes a traditional deep convolutional neural network and adds shortcut connections that skip few convolutional layers. The convolutional layers output is added to the input of the remaining blocks created by the shortcut connections. The ResNet50 model, for example, is made up of fifty layers of similar blocks connected by shortcuts. These connections keep calculations to the minimum while providing a wealth of combination features.

There are 16 residual modules in the ResNet50 model, which contains one (1) convolutional layer, a normalization layer, and two (2) pooling layers in between. There are two types of such modules, one with four convolutional layers and the other with three, and each convolutional layer is followed by normalization. The residual block with four (4) convolutional layers is utilized first, followed by two or more such blocks with three convolutional layers, and so forth.

The Google research team, led by Christian Szegedy, was primarily concerned with decreasing the computational weight of CNNs while preserving their performance. They proposed a novel module known as "The Inception Module," which is essentially four parallel routes of 1x1, 3x3, and 5x5 convolution filters. The model's execution time is faster than VGG or ResNet.

For Classification, shape and size of the tumour are selected. The segmentation model is used to separate out only the tumour from the rest of the ultrasound image. The resulted mask has the tumour in white (pixel value 255) and the rest of the mask is black (pixel value 0). To do classification, ResNet50 is used as the base model with attached output neural network with convolutional layers. A kernel size of 3 is used, which means that in each layer a 3*3 matrix runs over the image. This matrix carries the weights for that particular layer. During training, the target dataset for the model is employed whether the mask is learning benign or malignant. Since the model is pre-trained on the ImageNet

dataset, which is a huge dataset of about 1.2 million images organized into 1000 classes, it already knows aspects of image classification. When trained model is used on the dataset by initializing it with the pre-trained weights, it asks the model to adapt its image classification techniques to suit our job. Since in most cases the shapes of benign and malignant tumours are very different, the model learns to extract features that give higher weight to this aspect. Now the base model returns 2048 features which we feed into the output neural network which is a simple 8-layer network. The output model takes in these chosen features and gives the verdict. It uses ReLU activation in the hidden layers of this network, and the number of neurons from top to bottom is 1024,512,256,128,64,32,10. It uses softmax in the activation of the output layer which gives the probability of the mask belonging to a benign or malignant class, where the one with a higher probability is chosen as the result.

2.7 Evaluation Metrics

As mentioned earlier, segmentation of the ultrasound images is done using those masks to separate out the abnormalities. These masks are then passed into the classification model for making the decision. To use ResNet50, the weights are initialized by training on ImageNet as our base model. The output of the model consists of layers with 1024,512,256,128,64,10,2 neurons respectively. Binary cross entropy may be used as the loss function. Binary cross entropy compares the probabilities predicted by the model to the actual class output which can be either malignant (1) or benign (0). It then calculates the score that penalizes the probabilities based on how far the prediction is from the expected value. Mathematically, it is expressed as:

$$\log loss = \frac{1}{N} \sum_{i=1}^N -(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)) \quad (1)$$

Here y_i represents the class to which the i^{th} image belongs to, 1 for malignant, 0 for benign. P_i refers to the probability of the image being of a cancerous tumour while $1-P_i$ refers to its probability of being benign. The Adam optimizer (of gradient descent) is used with an initial learning rate (say 0.0001), which is modified if the model hits a plateau during training by the Keras library function. The method is efficient when working with large problems involving a lot of data or parameters. It requires less memory and is efficient. Broadly speaking it is a combination of two gradient descent methodologies, ‘gradient

descent with momentum' and 'Root Mean Square Propagation'. The former algorithm makes gradient descent converge faster by taking 'exponentially weighted average' of the gradients. The later is an adaptive learning algorithm working on 'an adaptive moving average'. Adam basically combines the strengths of these two and uses that to optimize gradient descent.

The authors (Medjahed, Saadi, & Benyettou, 2013) provided measures based on the spatial intersection of ground-truth and system-generated results and then produced multiple performance metrics, which were then averaged for all sampled frames. The examples of various detector performance keys precision, recall, and mean average precision (mAP). mAP is also a metric for calculating the accuracy of machine learning algorithms. The True Positive is the number of good (safe and uncluttered) detection found by the algorithm. The amount of non-good detection mistakenly identified as excellent detection by the algorithm, and the number of good detections overlooked by the system, are known as false positives and false negatives, respectively.

$$mAP = \sum_{q=1}^Q \frac{AveP(q)}{Q} \quad (2)$$

The metrics list includes accuracy, sensitivity, specificity, F1-score, recall, Mathew Correlation-coefficient (MCC), AUC-score, and ROC curve, all calculated from the confusion matrix, which contains True negative, True positive, false positive, and false negative values. The evaluation matrices are described below:

Accuracy: This shows the overall performance of the model and can be calculated by the formula given below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (3)$$

Specificity: This is ratio of the recently classified healthy people to the total number of healthy people. The formula for calculating specificity is given as follows:

$$Specificity = \frac{TN}{TN+FP} * 100 \quad (4)$$

Sensitivity (Recall): This is the ratio of recently classified heart patients to the total patients having heart disease. It can be calculated as:

$$Sensitivity = \frac{TP}{TP+FN} * 100 \quad (5)$$

Precision: Precision is the ratio of the actual positive score and the positive score predicted by the classification model, and can be calculated by the following formula:

$$Precision = \frac{TP}{TP+FP} * 100 \quad (6)$$

Softmax: This is a function which returns the probability of the image belonging to one of the two classes.

$$f_i(x) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (7)$$

The higher probability is displayed as the result.

F1-Score: F1 is the weighted measure of both recall precision and sensitivity. Its value ranges between 0 and 1. The closer the value to one, the better the performance of the classification model. It can be calculated as:

$$F1 = \frac{2*(Precision*Recall)}{recision+Recall} \quad (8)$$

MCC: It is a correlation coefficient between the actual and predicted results. MCC gives resulting values between - 1 and + 1, and can be calculated as given below:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (9)$$

Finally, we examine Area Under the Curve (AUC) that describes the ROC of a classifier. The performance of the classification algorithms is directly linked with AUC. For example, the larger the value of AUC, the greater will be the performance of the classification algorithm. The Receiver Operating Characteristic (ROC) curve consists of the True Positive (TP) rate as the y-axis and False Positive (FP) rate as the x-axis with the area under the ROC curve being calculated to show the performance of the classifier.

2.8 Proposed Approach

The approach adopted in this work can be shown graphically, as illustrated in Figure 1. The dataset, first, undergoes preprocessing to enable further processing steps. Next, it is

segmented before classification. Finally, the results are evaluated using well known metrics. Below, each step is discussed in detail.

Dataset: The dataset used in this research is ImageNet--a dataset that has 100,000+ images across 200 different classes. The size of each image as an average was 470*390 but cropped to 256*256. From this dataset, 2500 images were used for training, and 280 for images testing, so roughly a ratio of 90% to 10% for training and testing.

Preprocessing: For clearing the noise from the image, Gaussian blur and histogram equalization were used. Gaussian blur is best suited for clearing general noise from images. Histogram equalization helps in improving the contrast of the image by spreading out frequency intensity values. The size of Gaussian blur being used at the moment is (19,19). It can be modified depending on the performance of the model, so as to achieve maximum accuracy.

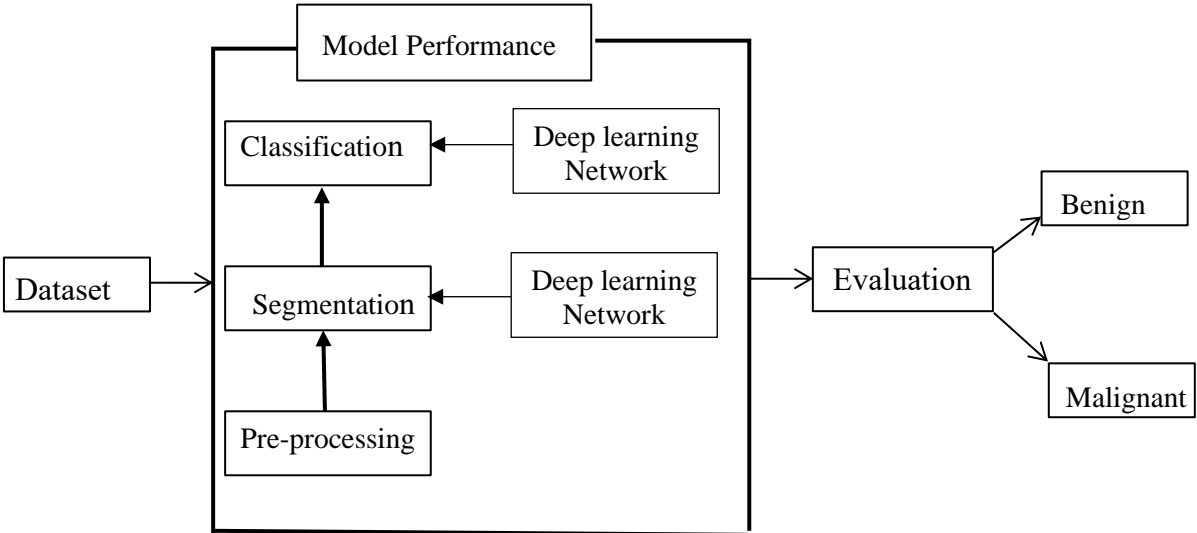


Figure 1: Model block diagram

Two models are explored in this research. The first model is Resnet101, which is a 101-layer convolutional neural network (CNN) that may be used as a state-of-the-art classification model. It is different from conventional neural networks in such a way that it retains residuals from every layer and applies them in next connected layers. It builds network by stacking residual blocks on top of one another and can stretch to many layers

per network by efficiently learning all the parameters from early activations. The Resnet101 model used in the segmentation model was imported from the Keras library and was pre-trained on the ImageNet dataset.

U-Net is a U-shaped (encoder-decoder) network architecture and is an expanded version of convolutional neural network and was developed for biomedical applications with fast and precise segmentation in situations, where the target is not only to classify whether there is an infection or not but also to identify the area of infection. In UNet architecture, the encoder decreases the spatial dimensions in each layer with increase in channels, whereas the decoder increases the spatial dimensions with decrease in channels. This model was imported from segmentation model library and was pre-trained on ImageNet dataset. Image segmentation gives us a mask that has only the tumor isolated for our study. All the unnecessary background information like lesions and masses are removed from the image.

Chapter 3: Experimental Results

In this section, the experimental results are presented, which are obtained using the proposed approach and a number of well-known classifiers implemented on breast cancer dataset. For the diagnosis of cancer, the ultrasound images are used, since that is what doctors use in diagnosing the disease. Another way could have been by going with numerical features instead of ultrasound images. But if a model could do the decision-making straight from the images, then it is much more efficient than having to first determine features like smoothness, compactness, concavity, texture, etc., before it is finally fed into a model. So, it was decided to use the image itself for doing the classification. The approach to this problem is to have two models, one which will separate out the tumour from the rest of the ultrasound images and a classification model which will give the verdict. Ultrasound images have a lot of other things besides tumours, lesions, masses, and harmless cysts. But what was needed to diagnose cancer is to determine whether the anomaly in the ultrasound is a malignant tumour or not. So, image segmentation was employed to separate it out from the rest of the image and study it. The base of the classification is using a mask of the anomaly since malignant tumours and benign cysts are in general very different in size and shape. The net thus deployed is a learning architecture specifically built for biomedical image segmentation. So, it was decided to go with that for mask generation or segmentation process. For doing this differentiation, it was decided to go with Convolutional Neural Networks. There are proven best performers in the field of image-based classification. The performance-based use various CNNs to determine which one gives the best predictions for the proposed model. The models selected were VGG16, VGG19, ResNet101, ResNet50, ResNet152, and MobileNet. The best performance received was from ResNet50, so it was decided to stick with it.

3.1 Pre-processing

First, data augmentation is done to increase the size of the dataset since our dataset only had 760 ultrasound images which were not enough for the job. To do data augmentation, a library known as Albumentations was imported. The process uses functions of this library to do data augmentations. The brief description of which functions are used for

which purpose, and what augmentation was done, is described below:

VerticalFlip: Reversing all rows and columns of the image vertically.

HorizaontalFlip: Reversing all rows and columns of the image horizontally.

RandomRotate90: Give the image a 90-degree rotation clockwise or anti-clockwise.

GridDistortion: If we put a uniform grid on an image then Grid Distortion will basically alter the structure of that grid which creates changes in the image structure.

One or more of these four processes are randomly applied to the images and new images are created, it is applied simultaneously to the images and masks so that it does not so happen that masks and images undergo different transformations. It increased the dataset size to 10000 images by augmentation. The size of each image as an average was 470*390 but cropped to 256*256. From this dataset, 2500 images were used for training, and 280 images for testing, so roughly a ratio of 90% to 10% was set for training and testing. Next, histogram equalization was applied to the images to improve the contrast of the image so that the tumours or cysts are more prominent. Histogram equalization helps to improve the contrast of the image. Moreover, Gaussian blur, Newtonian blur, and other blurring algorithms were also tried, but these require specifying kernel size and all images need a unique kernel size for the best result. Since the results from blurring were inconsistent across the dataset, it was decided to drop these blurring and proceed with histogram equalization. After this, the images that are basically 2D NumPy arrays are stacked into one array to be fed into training. Before feeding into the model, all models come with their specific pre-processing technique provided in respective library. So, for every model tried, it had to run through specific pre-processing module that came with the model. For example, for image segmentation, pre-processing module of ResNet101 provided by the segmentation models library was used. For the classification model, pre-processing module of ResNet50 was deployed from the Tensorflow library. For illustration purposes, five sample images (two benign and three malignant cases) are shown in following Figure 2.

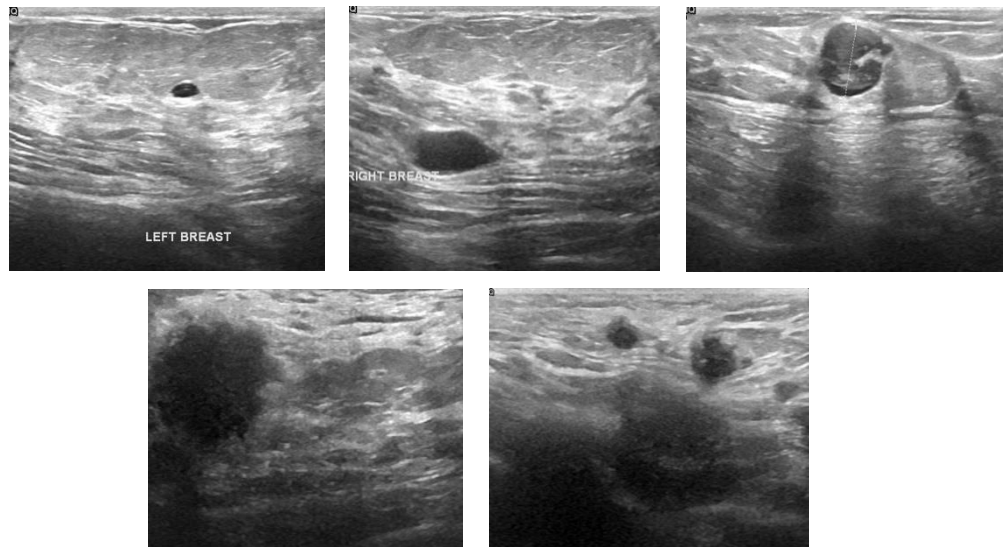


Figure 2: Ultrasound images (From top left: benign, benign, malignant, malignant, malignant)

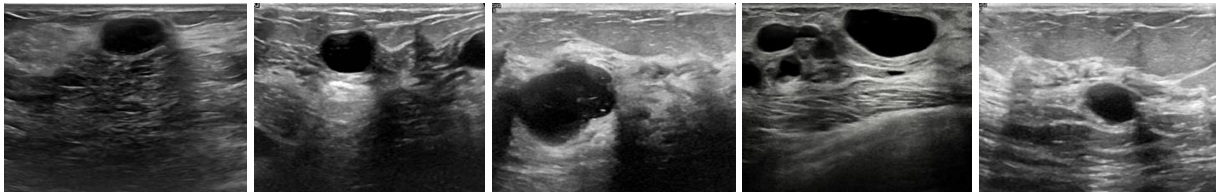
3.2 Experimental Platform

The model was run on an HP Pavilion Laptop 15-eg2xx. It has 16 GB DDR4-3200 SDRAM, 512 GB PCIe NVMe M.2 SSD, Intel Core i5-1155G7(4 cores, 8 threads) with an integrated iRIS xe graphics card. OS is Windows 11. The model was coded and trained as a Google Colab Pro notebook. Colab hosts the notebook and runs it on a cloud. It was provided with 24 GB RAM, 2x vCPU, and P100 GPU. The models were built on Keras framework. The libraries used in implementation are segmentation_models and TensorFlow. The UNet architecture and its ResNet101 backbone were taken from the segmentation_models library. The pre-processing module for ResNet 101 too was taken from segmentation_models. The model imported from the library was already pre-trained on the ImageNet dataset with the corresponding weights initialized accordingly. Tensorflow library was used to import the metrics: MeanIOU, Precision, Recall, and Accuracy. MeanIOU and Accuracy were used only in the segmentation model and classification model, respectively. The other metrics were used in both. ResNet50 model with weights already initialized according to pre-training on ImageNet was also taken from Tensorflow. The corresponding pre-processing module was taken from TensorFlow too.

3.3 Segmentation Results

For a sample of ten (10) images, the segmentation results for 5 benign and 5 malignant cases are shown in following Figure 3.

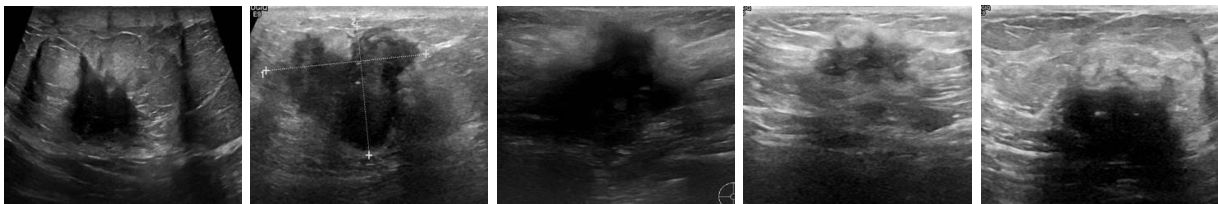
Benign cases: 1 through 5



Segmentation results 1 through 5



Malignant cases: 1 through 5



Segmentation results 1 through 5



Figure 3: Segmentation results for sample images

The masks of the images are basically the tumors separated from the rest of the image. The tumor is shown by the white area and black is the background. The images received for training are manually classified by doctors to be benign or malignant. The proposed model learns from that and further classifies the tumors as benign or malignant. Mostly it

is the size and shape of the tumor, which decides. It turned out that the smaller masks represent benign, and the bigger masks represent malignant.

During segmentation training and testing, different metric measurements were also computed, and the results are shown in following Figures 4 and 5.

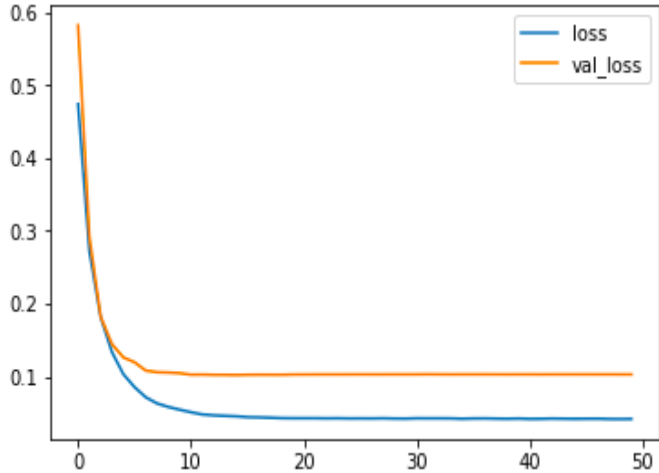


Figure 4: Training and validation loss function

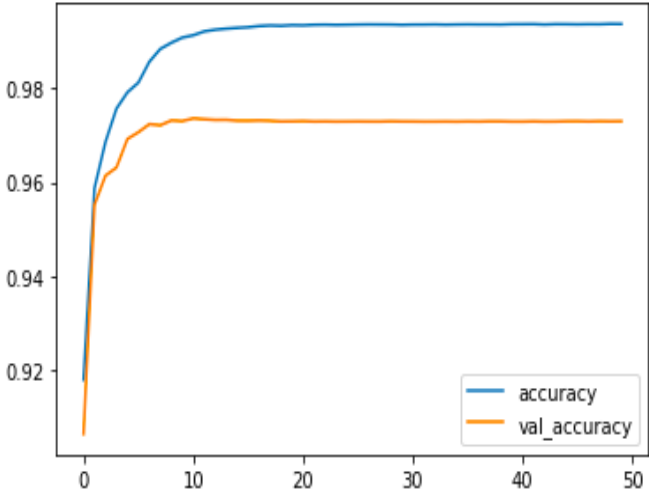


Figure 5: Tracking accuracy during training and validation

The Figures (4, 5) relate to segmentation and show the value of a function versus the number of epochs. The Figure 4 tracks the value of the loss function that is monotonically

decreasing before being flattened as the model trains. The blue line varies at each epoch and tracks the training loss while orange line tracks validation loss. The Figure 5 displays tracking the accuracy as monotonically increasing once the model trains. The variation is obvious for validation at each epoch, where the blue line tracks accuracy at that epoch during training, while orange line tracks the same but during validation.

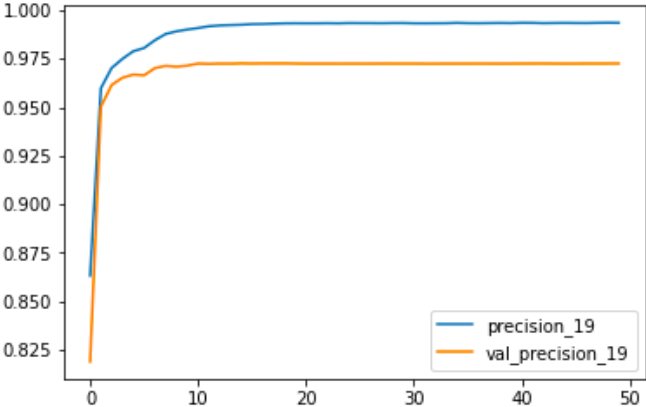


Figure 6: Tracking precision - training and validation

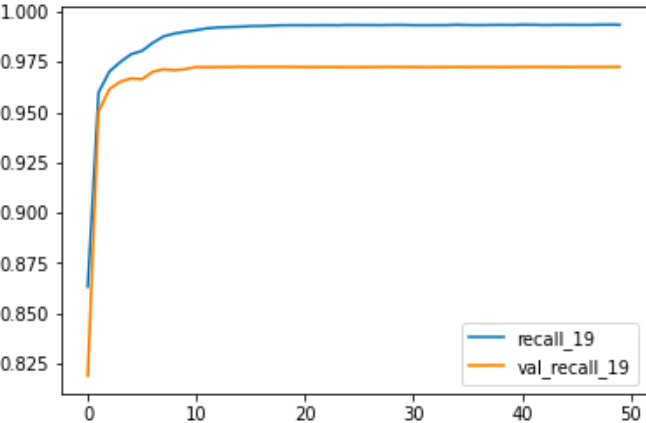


Figure 7: Tracking recall - training and validation

The Figures (6, 7) relate to segmentation and show the value of a function versus the number of epochs. The Figure 6 displays tracking the precision as monotonically increasing as the model trains. The variation is obvious for validation at each epoch. The blue line tracks accuracy at that epoch during training, while orange line tracks the same

but during validation. The Figure 7 displays recall model and shows tracking of the recall function as monotonically increasing once the model trains. The variation is obvious for validation at each epoch. The blue line tracks recall at that epoch during training, while orange line tracks the same but during validation.

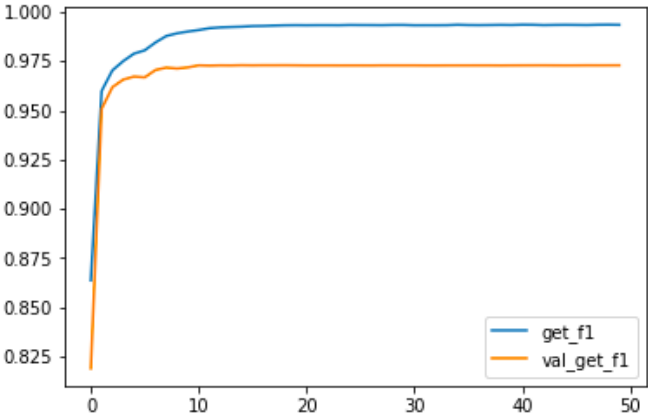


Figure 8: Tracking F1 - score during training and validation

The Figure 8 displays the F1 score of the model versus the number of epochs. The graph tracks the F1 score as a monotonically increasing function as the model trains. The line of validation shows variation at each epoch. The blue line tracks F1 score of the model at each epoch during training data, while orange line tracks the same but during validation.

3.4 Classification Results

Three different ResNets50, 101, and 152 were used differing in a number of layers, all share the same architecture. ResNet was developed because it was noticed that contrary to what theory says, after a certain number of layers, increasing the size of the neural network was leading to worse training results. This was happening due to various reasons like vanishing or exploding weights or the layers just being unnecessary. Hence Residual Networks brought the idea of skip connections. That is where one can skip layers in the middle and straight away go to a deeper layer. This allowed deep models to ignore layers that were not improving performance. The evaluation results on these three ResNet models using the metrics mentioned earlier are shown in Table 1.

Table 1: Comparing ResNet50, ResNet101, ResNet152

	ResNet 50	ResNet 101	ResNet 152
Trainable params	26,332,506	45,350,746	61,017,434
Training Accuracy	0.9861	0.9785	0.9725
Validation Accuracy	0.9677	0.9631	0.9534
Training Precision	0.9861	0.9785	0.9725
Validation Precision	0.9677	0.9631	0.9534
Training Recall	0.9861	0.9785	0.9725
Validation Recall	0.9677	0.9631	0.9534
Training F1 score	0.9841	0.9766	0.9706
Validation F1 score	0.9678	0.9632	0.9534

During classification training and testing, different metric measurements were also computed, and the results are shown in following figures.

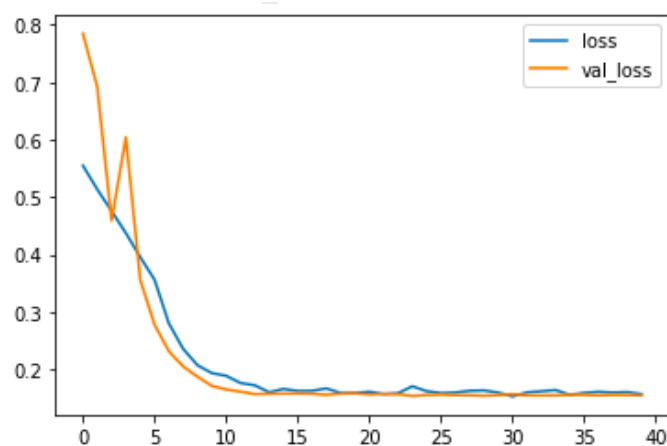


Figure 9: Training and validation loss function

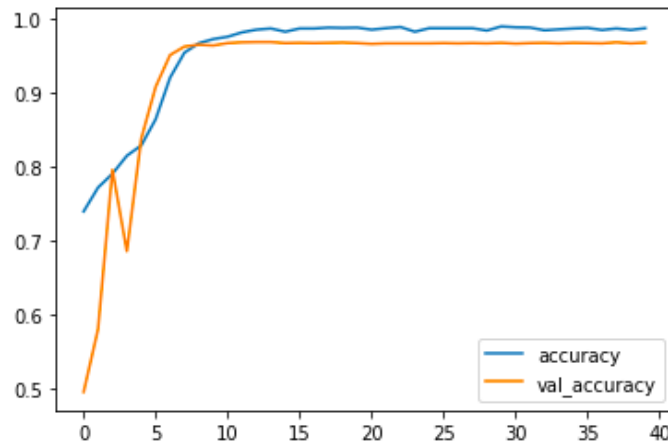


Figure 10: Tracking accuracy during training and validation

The Figures (9, 10) relate to classification and show the value of a function versus the number of epochs. The Figure 9 tracks the value of the loss function that is monotonically decreasing but with more variations comparing to segmentation result. The blue line varies at each epoch and tracks the training loss while orange line tracks validation loss. The Figure 10 displays tracking the accuracy as monotonically increasing once the model trains. The variation is obvious for validation at each epoch, where the blue line tracks accuracy at that epoch during training, while orange line tracks the same but during validation.

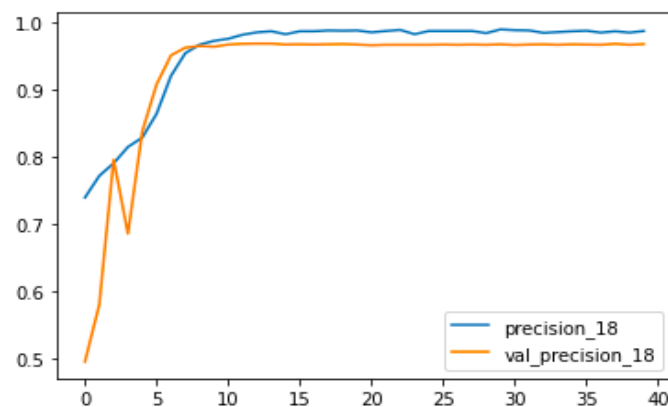


Figure 11: Tracking precision - training and validation

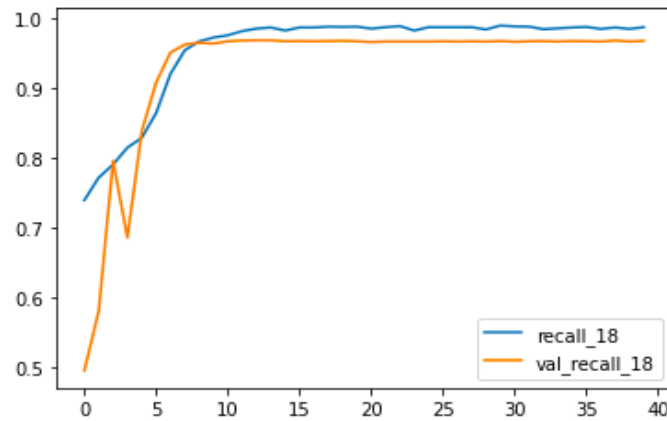


Figure 12: Tracking recall - training and validation

The Figure 11 displays tracking the precision related to classification as monotonically increasing as the model trains. The blue line tracks accuracy at that epoch during training, while orange line tracks the same but during validation. Similarly, the Figure 12 displays recall model of classification, and shows tracking of the recall function as monotonically increasing once the model trains. The blue line tracks recall at that epoch during training, while orange line tracks the same but during validation.

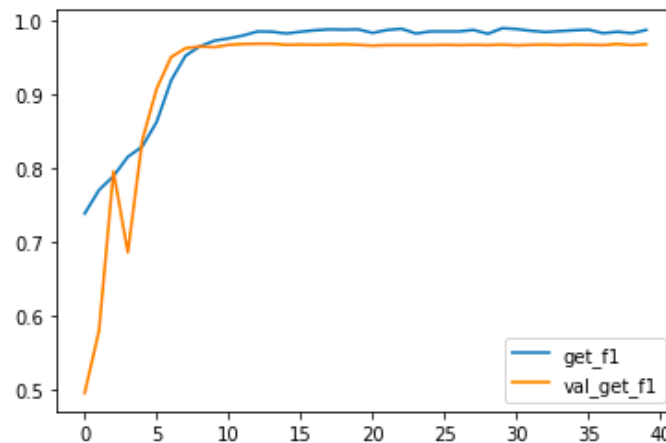


Figure 13: Tracking F1 - score during training and validation

The Figure 13 displays the F1 score of the model during classification versus the number of epochs. The graph tracks the F1 score as a monotonically increasing function as the model trains. The line of validation shows variation at each epoch. The blue line tracks F1 score of the model at each epoch during training data, while orange line tracks the same

but during validation. Typically, the models tend to perform lower during validation than during training, as the model becomes familiar with data during training. The curves are smoother most of the time meaning that either they are monotonically increasing or decreasing. Additionally, the gap between the two (training and validation) curves in all of metrics is shorter, which means the model is performing well. It can be seen from the graphs that all metrics graphs are consistent for both segmentation and classification models.

The next group of networks used was from VGG Net, the full name being Visual Geometry Group. It is a standard deep CNN with multiple layers. This model was developed for improving performance in object recognition tasks. Developed as a deep neural network, the VGG Net surpassed baselines on many tasks and datasets beyond ImageNet. It is still one of the most popular image recognition architectures. It replaced the large kernel-sized filters used in then best AlexNet with 3*3 kernel-sized filters in each layer one after the other which made it significantly better. The addition of ReLU units after each convolution also started with this model, which made training much easier and efficient compared to AlexNet. The evaluation results on these two VGG models using the metrics mentioned earlier are shown in Table 2.

Table 2: Comparing VGG 16 and VGG 19

	VGG 16	VGG 19
Trainable parameters	15,939,738	21,249,434
Training Accuracy	0.8124	0.9116
Validation Accuracy	0.8089	0.9157
Training Precision	0.8124	0.9116
Validation Precision	0.8089	0.9157
Training Recall	0.8124	0.9116
Validation Recall	0.8089	0.9157
Training F1 Score	0.8124	0.9099
Validation F1 Score	0.8084	0.9155

MobileNet architecture has 13 blocks. Inside each block, there is a depth-wise convolution for all three channels followed by a pointwise convolution which increases the number of channels. After all the convolutions, we do average pooling followed by the fully connected layer and then softmax in the final layer. MobileNet_v2 architecture has 17 blocks. In this case, all blocks have a pointwise convolution, known as expansion, followed by the depth-wise convolution for all three channels and another pointwise convolution known as projection. Each block also has a skip connection which skips all layers in the block, so if the block does not help in improving the model, then the model can disable it with this skip connection. The evaluation results on these two MobileNet models using the metrics mentioned earlier are shown in Table 3.

Table 3: Comparing MobileNet and MobileNet_V2

	MobileNet	MobileNet_V2
Trainable params	4,978,222	4,235,354
Training Accuracy	0.9789	0.9741
Validation Accuracy	0.9638	0.7522
Training Precision	0.9789	0.9741
Validation Precision	0.9638	0.7522
Training Recall	0.9789	0.9741
Validation Recall	0.9638	0.7522
Training F1 Score	0.9790	0.9741
Validation F1 Score	0.9639	0.7519

The model which worked best for the classification of the masks in our case was ResNet50, with the best metric values among the eight models we used. MobileNet can be considered the most efficient since it needed only 4.9 million parameters compared to 26.3 million

needed by ResNet50 and still the metrics it generated were very close to that of ResNet50. The worst performing model was the VGG model, with VGG 16 being the worst to be specific. VGG models are simple deep convolutional networks. These were the first models which introduced smaller sized 3*3 kernels and a ReLU unit after each convolution. This was the first architecture to make a big jump on AlexNet. However, the problem with VGG architecture, which was later solved by the Residual Networks, was that if there were layers that are not helping improve the model then the model tries to nullify them by transforming them into identity mappings. In a deep network like VGG, it was very difficult for the model to be able to perform this operation and it either needed a huge or impossible number of epochs to achieve that. Residual Networks with their skip connections solved this problem as the model could now just choose the path of identity mapping and ignore the layer instead of trying to convert that layer into an identity mapping. This may be the reason VGG performed poorly on our data compared to the Residual Networks because they had a lot of redundant layers that they could not get rid of. MobileNet performed exceptionally well, a possible reason for this can be that the number of parameters it provided was optimal for classification on our model. It did not have any redundant layers dragging the model performance down during training, hence delivering such high metrics, while also being the most efficient. MobileNet_V2 is based upon MobileNet architecture with a skip connection across each module of the model. This allows for removing the modules which are not helping. MobileNet_V2 performed just as well as MobileNet in training as it should be since they have a similar number of total parameters. However, MobileNet_V2's extremely poor performance in validation can be due to overfitting by the model. MobileNet_V2 has 17 modules compared to 13 of MobileNet and each module has an extra pointwise convolution layer, so the model could have performed poorly on validation due to overfitting. Thus, ResNet50 from the Residual Networks group was chosen as classifier due to its ability to use skip connections for bypassing poorly performing layers. Although MobileNet was doing only a marginally inferior job despite having less than a fifth of the number of parameters of ResNet50 and being the most efficient amongst all, we decided to go with higher metrics for a little loss of inefficiency. There are, however, some wrong classifications. With high accuracy

achieved, there are errors, mostly due to special cases in tumor shapes, or an error in segmentation. However, the model differentiates tumors varying in size perfectly.

Chapter 4: Discussion and Conclusions

Breast cancer generally develops in the milk ducts and glands. They initially grow as epithelial tumors in these places before forming a lump which causes cancer. Many time these lumps are benign, but they can also be premalignant and may develop into a malignant tumor in the future. Whether the detected abnormal masses are malignant or benign lumps, they can be stated with respect to their shape (lobular, oval, round and irregular) or their margin (indistinct, obscured, and spiculated) characteristics. The masses with lines radiating from their margins are particular kind of masses with a high probability of malignancy. In most cases, spiked and irregularly shaped tumors are indicators of cancer. Benign calcifications are typically large and coarse with smooth and round contours. Malignant calcifications appear to be clustered, small, varying in shape and size, angular, and irregularly shaped.

Breast cancer diagnosis targets to identify these abnormalities and diagnose them so that they can be treated before symptoms start to develop. Mammograms, MRI scans, US scans, and CT scans are all popular methods used for cancer screening. These help us visualize the features of the tumor and do the diagnosis. Mammograms and Ultrasounds are particularly famous when it comes to detecting tumors in their pre-invasive state. In this thesis work, ultrasound images were used because mammograms have some drawbacks, in some cases the mammograms of the patients are not clear, and there is very low contrast between the abnormalities and surrounding lesions. For preprocessing, histogram equalization was used to increase contrast. Next, segmentation of the dataset was done to separate the tumor from its surroundings based on shape, spread, and patterns of the tumor to help decide whether it is a malignant or a benign one. The model used in segmentation was UNet architecture with a ResNet backbone. This architecture is pretrained on the ImageNet dataset for initializing the weights. With these as the target for segmentation, the model was trained with F1-score of 0.9671 F1. After that, these masks were used in base model of ResNet50 to extract features from the mask. Then these were fed into a self-written neural network to do the prediction. The masks are the input data, and their category is the target. The ResNet50 base model extracts 2048 features from the mask and then carried onto output network for making the decision whether the tumor is

malignant or benign. Each layer in the output network uses the ReLU activation function. In the final layer, softmax function is used to generate probabilities of whether the mask is of a benign or a malignant tumor. The classification accuracy thus achieved was 98.61% with an F1 score of 0.9861.

The proposed model may be improved to bring improvement in image pre-processing and the segmentation part. The simple histogram equalization was used to improve contrast of the image but more advanced tools like active contour techniques can do a much better job. The more advanced methods such as advanced statistical techniques such as wavelet methods and intensity-based methods may also be investigated. This helps a lot in segmentation and can push up accuracy of segmentation to 99%. The classification model may also be improved by applying more feature selection techniques like principal component analysis, recursive feature selection, etc. To further optimize the decision making of the network, support vector machine will potentially do a better job than softmax at the end layer, as it has a non-linear overlapping dataset. The SVM reduces the generalization error during testing of data and is accurate and efficient in computations due to reduced parameters. This helps in more accurate predictions.

Deploying a machine learning model in real-time is simple and easy to do. It may be integrated into a website or software at its backend as a part of online social network, similar to works in (Abdulhameed & Memon, 2022; Memon & Mustafa, 2015) where patient data is set for real time monitoring. Ultrasound images are to be fed into the front, and the model does all the processing on the image before giving out the result. The intermediate processing of the images can also be displayed for medical practitioners to study and understand.

References

- Abbass, H. A. (2002). An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial intelligence in Medicine*, 25(3), 265-281.
- Abdulhameed, A., & Memon, Q. (2022). Support Vector Machine Based Design and Simulation of Air Traffic Management for Prioritized Landing of Large Number of UAVs. *European Journal of Artificial Intelligence and Machine Learning*, 1(2), 17-21.
- Afshar, H. L., Ahmadi, M., Roudbari, M., & Sadoughi, F. (2015). Prediction of breast cancer survival through knowledge discovery in databases. *Global journal of health science*, 7(4), 392.
- Aličković, E., & Subasi, A. (2017). Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Computing and applications*, 28(4), 753-763.
- Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., . . . Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *Plos one*, 12(6), e0177544.
- Arevalo, J., Cruz-Roa, A., Arias, V., Romero, E., & González, F. A. (2015). An unsupervised feature learning framework for basal cell carcinoma image analysis. *Artificial intelligence in Medicine*, 64(2), 131-145.
- Aswathy, M., & Jagannath, M. (2017). Detection of breast cancer on digital histopathology images: Present status and future possibilities. *Informatics in Medicine Unlocked*, 8, 74-79.
- Birkett, C., Arandjelović, O., & Humphris, G. (2017). *Towards objective and reproducible study of patient-doctor interaction: automatic text analysis based VR-CoDES annotation of consultation transcripts*. Paper presented at the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2638-2641. IEEE.
- Chao, C. M., Yu, Y. W., Cheng, B.-W., & Kuo, Y. L. (2014). Construction the model on the breast cancer survival analysis use support vector machine, logistic regression and decision tree. *Journal of medical systems*, 38(10), 1-7.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Dabeer, S., Khan, M. M., & Islam, S. (2019). Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked*, 16, 100231.

- Dai, X., Cheng, H., Bai, Z., & Li, J. (2017). Breast cancer cell line classification and its relevance with breast tumor subtyping. *Journal of Cancer*, 8(16), 3131.
- Dalle, J. R., Leow, W. K., Racoceanu, D., Tutac, A. E., & Putti, T. C. (2008). *Automatic breast cancer grading of histopathological images*. Paper presented at the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 3052-3055. IEEE.
- Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in Medicine*, 34(2), 113-127.
- Dhungel, N., Carneiro, G., & Bradley, A. P. (2015). *Tree re-weighted belief propagation using deep learning potentials for mass segmentation from mammograms*. Paper presented at the 2015 IEEE 12th international symposium on biomedical imaging (ISBI). 760-763. IEEE.
- Dhungel, N., Carneiro, G., & Bradley, A. P. (2016). The automated learning of deep features for breast mass classification from mammograms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 106-114. Springer, Cham.
- Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). *Fully automated classification of mammograms using deep residual neural networks*. Paper presented at the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI). 310-314. IEEE.
- Dundar, M. M., Badve, S., Bilgin, G., Raykar, V., Jain, R., Sertel, O., & Gurcan, M. N. (2011). Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, 58(7), 1977-1984.
- Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, 8(1), 19-20.
- Fu, Q., Luo, Y., Liu, J., Bi, J., Qiu, S., Cao, Y., & Ding, X. (2017). *Improving learning algorithm performance for spiking neural networks*. Paper presented at the 2017 IEEE 17th International Conference on Communication Technology (ICCT). 1916-1919. IEEE.
- Furundzic, D., Djordjevic, M., & Bekic, A. J. (1998). Neural networks approach to early breast cancer detection. *Journal of systems architecture*, 44(8), 617-633.

- Gardezi, S. J. S., Elazab, A., Lei, B., & Wang, T. (2019). Breast cancer detection and diagnosis using mammographic data: Systematic review. *Journal of medical Internet research*, 21(7), e14464.
- Haq, A. U., Li, J., Memon, M. H., Khan, J., Din, S. U., Ahad, I., . . . Lai, Z. (2018). *Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease*. Paper presented at the 2018 15th International computer conference on wavelet active media technology and information processing (ICCWAMTIP). 101-106. IEEE.
- Haq, A. U., Li, J. P., Memon, M. H., Malik, A., Ahmad, T., Ali, A., . . . Shahid, M. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access*, 7, 37718-37734.
- Ibrahim, N., Kudus, A., Daud, I., & Bakar, M. A. (2008). Decision tree for competing risks survival probability in breast cancer study. *Int J Biol Med Sci*, 3(1), 25-29.
- Ibrahim, R., Yousri, N. A., Ismail, M. A., & El-Makky, N. M. (2014). *Multi-level gene/MiRNA feature selection using deep belief nets and active learning*. Paper presented at the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 3957-3960. IEEE.
- Islam, M. M., Iqbal, H., Haque, M. R., & Hasan, M. K. (2017). *Prediction of breast cancer using support vector machine and K-Nearest neighbors*. Paper presented at the 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). 226-229. IEEE.
- Jadoon, M. M., Zhang, Q., Haq, I. U., Butt, S., & Jadoon, A. (2017). Three-class mammogram classification based on descriptive CNN features. *BioMed Research International*, 2017.
- Khan, M. U., Choi, J. P., Shin, H., & Kim, M. (2008). *Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare*. Paper presented at the 2008 30th annual international conference of the IEEE engineering in medicine and biology society. 5148-5151. IEEE.
- Kohli, P. S., & Arora, S. (2018). *Application of machine learning in disease prediction*. Paper presented at the 2018 4th International conference on computing communication and automation (ICCCA). 1-4. IEEE.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.

- Li, H., Weng, J., Shi, Y., Gu, W., Mao, Y., Wang, Y., . . . Zhang, J. (2018). An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. *Scientific reports*, 8(1), 1-12.
- Liu, Y. Q., Wang, C., & Zhang, L. (2009). Decision tree based predictive models for breast cancer survivability on imbalanced data. In 2009 3rd international conference on bioinformatics and biomedical engineering. 1-4.
- Mahmood, T., Arsalan, M., Owais, M., Lee, M. B., & Park, K. R. (2020). Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *Journal of clinical medicine*, 9(3), 749.
- Martynenko, A., & Bück, A. (2018). *Intelligent control in drying* (Vol. 3): CRC Press Boca Raton, FL, USA.
- Medjahed, S. A., Saadi, T. A., & Benyettou, A. (2013). Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. *International Journal of Computer Applications*, 62(1).
- Meesad, P., & Yen, G. G. (2003). Combined numerical and linguistic knowledge representation and its application to medical diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 33(2), 206-222.
- Memon, Q. (2013). Smarter Healthcare Collaborative Network. *Building Next-Generation Converged Networks: Theory*, 451-476
- Memon, Q. (2019). On Assisted Living of Paralyzed Persons through Real-Time Eye Features Tracking and Classification using Support Vector Machines: Array. *Medical Technologies Journal*, 3(1), 316-333.
- Memon, Q. A., & Laghari, M. S. (2006). Building relationship network for machine analysis from wear debris measurements. *International Journal of computational intelligence*, 3(2).
- Memon, Q. A., & Mustafa, A. F. (2015). Exploring mobile health in a private online social network. *Int. J. Electron. Heal.*, 8(1), 51-75.
- Mohan, S., Bhattacharya, S., Kaluri, R., Feng, G., & Tariq, U. (2020). Multi-modal prediction of breast cancer using particle swarm optimization with non-dominating sorting. *International Journal of Distributed Sensor Networks*, 16(11).
- Noguchi, S., Nishio, M., Yakami, M., Nakagomi, K., & Togashi, K. (2020). Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Computers in biology and medicine*, 121, 103767.

- Onan, A. (2015). A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer. *Expert Systems with Applications*, 42(20), 6844-6852.
- Oyewola, D., Hakimi, D., Adeboye, K., & Shehu, M. D. (2016). Using five machine learning for breast cancer biopsy predictions based on mammographic diagnosis. *International Journal of Engineering Technologies IJET*, 2(4), 142-145.
- Pérez, N., Guevara, M. A., & Silva, A. (2013). *Improving breast cancer classification with mammography, supported on an appropriate variable selection analysis*. Paper presented at the Medical Imaging 2013: Computer-Aided Diagnosis. Vol. 8670, 520-533. SPIE.
- Pérez, N., Guevara, M. A., Silva, A., Ramos, I., & Loureiro, J. (2014). *Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection*. Paper presented at the 2014 Federated Conference on Computer Science and Information Systems. 209-217. IEEE.
- Ramadan, S. Z. (2020). Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *Journal of healthcare engineering*, 2020.
- Ramadan, S. Z., & El-Banna, M. (2020). Breast cancer diagnosis in digital mammography images using automatic detection for the region of interest. *Current Medical Imaging*, 16(7), 902-912.
- Ramadevi, G. N., Rani, K. U., & Lavanya, D. (2015). Importance of feature extraction for classification of breast cancer datasets—a study. *International Journal of Scientific and Innovative Mathematical Research*, 3(2), 763-368.
- Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10), 1619-1630.
- Sarkar, M., & Leong, T. Y. (2000). Application of K-nearest neighbors algorithm on breast cancer diagnosis problem. In Proceedings of the AMIA Symposium. 759. American Medical Informatics Association.
- Shao, Y. Z., Liu, L. Z., Bie, M. J., Li, C. c., Wu, Y. p., Xie, X. m., & Li, L. (2011). Characterizing the clustered microcalcifications on mammograms to predict the pathological classification and grading: a mathematical modeling approach. *Journal of digital imaging*, 24(5), 764-771.
- Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., . . . Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN

- architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), 1285-1298.
- Sohail, M. N., Jiadong, R., Uba, M. M., & Irshad, M. (2019). A comprehensive looks at data mining techniques contributing to medical data growth: a survey of researcher reviews. *Recent developments in intelligent computing, communication and devices*, 21-26.
- Sun, D., Wang, M., & Li, A. (2018). A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(3), 841-850.
- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 85, 189-203.
- Valappil, N. K., & Memon, Q. A. (2021). CNN-SVM based vehicle detection for UAV platform. *International Journal of Hybrid Intelligent Systems* (Preprint), 17, 1-12.
- Valvano, G., Santini, G., Martini, N., Ripoli, A., Iacconi, C., Chiappino, D., & Della Latta, D. (2019). Convolutional neural networks for the segmentation of microcalcification in mammography imaging. *Journal of healthcare engineering*, 2019.
- Wang, L., Chu, F., & Xie, W. (2007). Accurate cancer classification using expressions of very few genes. *IEEE/ACM transactions on computational biology and bioinformatics*, 4(1), 40-53.
- Wang, T. N., Cheng, C. H., & Chiu, H. W. (2013, July). Predicting post-treatment survivability of patients with breast cancer using Artificial Neural Network methods. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 1290-1293. IEEE.
- Yousefi, B., Ting, H. N., Mirhassani, S. M., & Hosseini, M. (2013, November). Development of computer-aided detection of breast lesion using gabor-wavelet BASED features in mammographic images. In 2013 IEEE International Conference on Control System, Computing and Engineering. 127-131. IEEE.
- Zhang, D., Zou, L., Zhou, X., & He, F. (2018). Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. *Ieee Access*, 6, 28936-28944.
- Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications*, 41(4), 1476-1482.

Zou, L., Yu, S., Meng, T., Zhang, Z., Liang, X., & Xie, Y. (2019). A technical review of convolutional neural network-based mammographic breast cancer diagnosis. *Computational and mathematical methods in medicine*, 2019.

The logo for the United Arab Emirates University (UAEU) is displayed in a red rectangular box. It consists of the letters 'UAEU' in a white, bold, sans-serif font.

جامعة الإمارات العربية المتحدة
United Arab Emirates University



UAE UNIVERSITY MASTER THESIS NO. 2022:60

Breast Cancer is a leading killer of women globally. It is a serious health concern caused by calcifications or abnormal tissue growth in the breast. Doing a screening and identifying the nature of the tumor as benign or malignant is important to facilitate early intervention, which drastically decreases the mortality rate. Usually, it uses ultrasound images, since they are easily accessible to most people and have no drawbacks as such, unlike the other most famous screening technique of mammograms where in some cases you may not get a clear scan.

Bitu Asadi received her Master of Science in Electrical Engineering from the Department of Electrical and Communication Engineering, College of Engineering at UAE University, UAE. She received her BSc in Electronics Engineering from Higher Colleges of Technology, Dubai, UAE.

www.uaeu.ac.ae