AFRICAN CENTRE OF EXCELLENCE
IN DATA SCIENCE

Classification of Invasive Ductal Carcinoma Breast Cancer Using Deep Neural Network

By

Sunge Makawa

Registration Number: 220001715

A Dissertation submitted in partial fulfilment of the requirements for the Degree of Master of Data Science in Biostatistics

University Of Rwanda, College Of Business and Economics

The African Centre of Excellence in Data Science (ACE-DS)

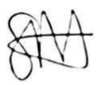
Supervisors: Prof. Weiwei Pan and Dr. Melanie Fernandez

September 2022

Declaration

I declare that this dissertation entitled **“Classification of Invasive Ductal Carcinoma Breast Cancer Using Deep Learning”** is the result of my own work and has not been submitted for any other degree at the University of Rwanda or any other institution.

Names: Sunge Makawa Signature:



Date:

Approval sheet

This dissertation entitled **"Classification of Invasive Ductal Carcinoma Breast Cancer Using Deep Learning"** written and submitted by Sunge Makawa in partial fulfilment of the requirements for the degree of Master of Science in Data Science majoring in Biostatistics is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 14% which is less than 20% accepted by the African Centre of Excellence in Data Science (ACE-DS).

Melanie F. Pradier

Supervisor

Supervisor

Head of Training

Dedication

I would like to dedicate this work to all breast cancer patients and survivors across the world. Sending you love and light all the way. Hoping more health improvements and advancements will be realised to help more patients graduate to becoming breast cancer survivors.

Acknowledgement

I would like to thank God for helping me throughout my studies. This thesis has been a long and tiresome process which would have not been possible without the help and support of my supervisors and mentors: Professor Weiwei Pan, Dr. Melanie Fernandez and Professor Pavlos Protopapas.

I would like to extend my gratitude to my family for being a source of my emotional and physical energy. And to my ACEDS cohort II classmates, thank you for the support.

Last but not least I would like to sincerely thank Inter-University Council for East Africa (IUCEA) for the scholarship and financial support rendered to me during my entire master's program at ACE-DS.

Abstract

Invasive Ductal Carcinoma (IDC) is a type of breast cancer that is common and it requires correct and early treatment. With technology advancement, automated computer tools have been developed for analyzing, diagnosing and predicting IDC.

In this work, a study has been conducted to compare the performance of different methods of classifying images using Deep Neural Network. Two main approaches were used. Building a neural network model from scratch and using transfer learning. The methods had different input values. However, both approaches used Convolution Neural Network architecture. 3-way cross validation was used to split the datasets. The evaluation methods for analyzing the results were learning curves, confusion matrix, and classification report. The results showed that ResNet50 model had a better performance and had a lot of images that were correctly classified compared to the other methods with a total accuracy of 87%. This was in comparison of total accuracy of 84% and 85% achieved by a model built from scratch and Mobilenet respectively. The study went further to check the significance of color in IDC breast cancer image classification by comparing the performance of models with colored images and images on a gray scale. According to overall accuracy, precision and recall, CNN model from scratch and ResNet50 trained on colored images is performing better compared to CNN model from scratch and ResNet50 trained on grayscale images.

Keywords: classification, breast cancer, deep neural network, transfer learning

TABLE OF CONTENTS

Glossary

Adam - Optimizer used to update the weight of a network when training iteratively

ANN-Artificial Neural Networks

BC-Breast Cancer

CNN-Convolutional Neural Networks

DL-Deep Learning

IDC-Invasive Ductal Carcinoma

List of figures

List of tables

Chapter 0: Introduction

0.1 Background and Rational

Breast cancer (BC) is one of the most common cancer that affects women and the second most common among all types of cancer. Over 2 million new cases aroused in 2018(World Cancer Research Fund, 2018). In most low- and middle-income country (LMIC) breast cancer is also a problem among women.

Breast cancer occurs when there is growth of abnormal cells in a woman's. The abnormal cells may start growing at a specific area of the breast and be spread to other parts. It is women who are susceptible to breast cancer compared to men because of the anatomy of the human body. There are different reasons and risk factors for breast cancer. Age, family history, breast density, obesity, and alcohol intake are some of the reasons for breast cancer (Skandalakis, 2009).

Women's breasts are developed by lobules, ducts, nipples, and fatty tissues. Milk is made in lobules and passed to nipple by ducts. Usually epithelial tumors develop inside lobules as well as ducts and there after form cancer inside the breast (Skandalakis, 2009). Among all types of breast cancers, Invasive Ductal Carcinoma (IDC) is the foremost common cause of death in women (DeSantis, Ma, Bryan and Jemal 2013). This kind of breast cancer begins in the duct cells and spreads to the tissues near the duct cells.

There are a number of strategies that have been utilized by researchers to explore and investigate breast images from diverse points of view depending on the status of the disease, the quality of the images and the demand of the disease. Among the many kinds tasks, for classifying the breast image, machine learning (ML) and the Artificial Intelligence (AI) are mostly utilized.

Deep learning is a king of machine learning that allows computers to learn by understanding the world in terms of a hierarchy of concepts and through experience (Goodfellow, Bengio, & Courville, 2016). The hierarchy of concepts enables the computer to learn complex ideas and concepts that cannot be learnt by human through building them out of simpler ones; a graph of these hierarchies would be many layers that are deep. Much attention has been given to Deep learning as of late because of its computational ability to run much larger models, its ability to handle growing dataset sizes, and increasing accuracy, application complexity and real world impact.

The Artificial Neural Networks(ANNs) which is deep learning contains neurons. Neurons are used to store and accept data before transferring to the next layer. Using multiple layers, the system that is built is complex. As such, it is possible for the system to reclaim and retrieve information without human interference (Jiang, et al., 2017). A good example of ANN is convolutional neural network(CNN) (Quang & Xie, 2016).

Convolutional Neural Network commonly known as CNN, is used in classifying images including medical images. One or more labels can be assigned to images using image classification technique of the CNN. The ability of CNN to classify images has been further extended to healthcare for example breast cancer.

Image classification can be used to classify images with breast cancer and images without cancer. This is necessary especially in low to middle income countries like where cancer specialist and resources are limited. Image classification reduces the need of specialist consults and reduce medical error.

The overall objective of this study predict the presence of breast cancer which is vital in early prediction using mammogram images. The rationale for this research is that many women are diagnosed with advanced stage disease, therefore there is need to identify techniques to increase timely detection of breast cancer which would result in improved cancer outcomes.

In most developing countries, there is shortage of cancer specialist who can identify if a mammogram image is cancerous or not. Therefore, it is nearly possible for the few specialists available to look at the images manually. However, deep learning through image classification can help oncology departments to identify cancer patients in time. For this reason, there is need to understand deep learning techniques that can help to predict cancer effectively and efficiently in order to enhance early detection interventions.

The main objective of this research study is build and analyze a Deep Learning Neural Network models that have been used to predict IDC breast cancer using images. Using these models can be helpful in early detection of cancer and early treatment.

The CNN model that will be built will be able to automatically flag women that have breast cancer using the mammograms. The emphasis is to point out, notify, and discuss a breast cancer predictive

Artificial Intelligent models. Moreover, discuss the benefits and challenges in cancer detection models.

This study uses two main approach of deep neural networks training. Firstly, it has used models trained from scratch to classify IDC images. Secondly, it has used transfer learning to classify IDC. Thereafter, evaluate and compare which method worked best for the breast cancer images. A recommendation is made on the most effective and efficient deep learning method that can be used in classifying IDC images.

In addition to analyzing the different approaches to classification of IDC images using deep learning, this study will also go further to understand why the models built were misclassifying images. It will be able to understand the reason the classifiers making mistakes.

The study will also understand the significance of color in image classification by training the models in three color channel: Red Green and Blue (RGB) and on one color channel that is the grayscale. Thereafter, it will compare the performance of the models built on images in 3 channel color and one color channel.

In this thesis, transfer learning using ResNet50 has performed better compared to training a deep learning from scratch and using MobileNet. After converting the train, validation and test set to grayscale, the performance on the validation and test set was reduced for both the CNN models trained from scratch and using ResNet50.

Convolutional Neural Network is a machine learning technique for classifying breast cancer images and this will help in analyzing the model's performance hence recommending medical sector in this.

0.2 Thesis Overview

In the first chapter of this thesis background of the study has been explained. This includes the theoretic background of the methods that have been used. Chapter two gives a review of the related works that have been done in IDC classification using deep learning. The studies have been summarized and the results have been outlined. In Chapter three data exploration has been done, where data description, data transformation or any data preprocessing method that was used in this thesis has been explained. Chapter four explains the methods that have been used. Chapter five gives the results achieved from the methods and experiments in this study. Each deep learning method that has been used in the study is explained here and the results have been shown. Chapter six, seven and eight give the discussion of the results, conclusion and future works respectively.

Chapter 1: Background

This chapter discusses the necessary theories and methods that have been used in this study to have an understanding in the subject matter of image classification using deep learning in health care sector.

1.1 Theoretic Background

Machine Learning

Machine learning which is a core concept that has been used in this thesis is a method of analyzing data by automating analytical model building. Machine learning is a subdivision of artificial intelligence built on the idea that systems can learn from data, identify trends or patterns of the data and make decisions using minimal human interference (Bell, 2014).

Although human beings undertake various task routinely, there has not been a sufficient elaboration in extracting a program that is well defined. For instance, there are activities like speech recognition, driving and image understanding. The art of machine learning is "learning from experience" achieve quite satisfactory results in these tasks if there are sufficiently many training examples (Shalev-Shwartz, Shai, & Ben-David, 2015).

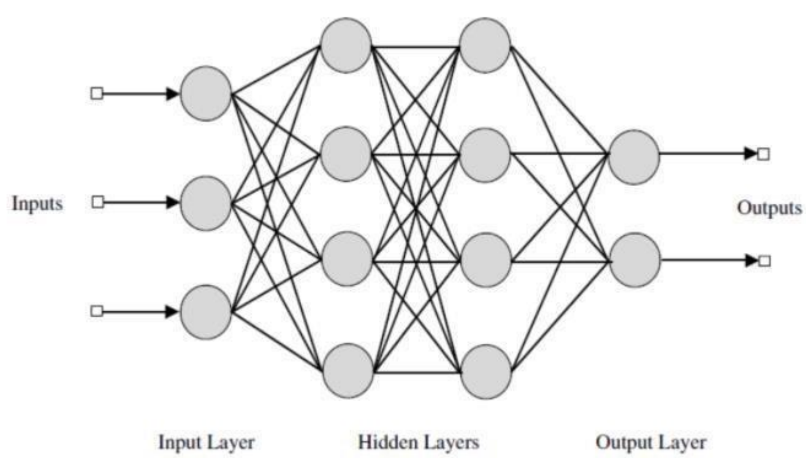
The area of machine learning has branched into several subfields that deal with different types of learning tasks. In machine learning there are different systems commonly known as algorithms that are used. The algorithms are grouped into two main types and that is supervised learning and unsupervised learning. On one hand, supervised learning group is referred to as a method utilizing a set of training data. Training data set is data is used to learn or know the features of the dataset. For each case, the dataset has an input commonly known as features in machine learning and output object which is referred to as labels (Bell, 2014). Training dataset is very crucial and highly depended in supervised learning. Consequently, training data set is required to be correct for the algorithm to make sense and understands the pattern and trends of the data. On the other hand, unsupervised learning is an algorithm that finds hidden patterns in a massive amount of data where it gives room for the computer algorithm to perform and see what the outcome patterns are going to be. Hence, the models is unsupervised learning is not provided correct answers (Bell, 2014). Machine learning also consists of dependent and independent variables which are also referred to as predictor or control input. The independent variables hold the labels that control the experiment.

1.1.2 Deep Learning

Deep Learning which is a subdivision of Machine Learning is was designed to have an architecture of a human brain and have similar functions. In the same way a human brain has neural networks that join billions of neurons deep learning has the same design. The idea of deep learning is for the machines to have a perception. The architecture comprises of artificial neural networks that connect a number of mathematical units called neurons (Lakshmanamoorthy, 2021).

Although the architectural design and functionality of a brain and deep learning is similar to each other, deep learning performs better than the human brain and has the capability of modelling complex problems. Deep Learning has achieved a lot in the field of machine learning as such different great domains of deep learning have emerged generalizing deep learning processes from data pre-processing to model deployment.

Artificial Neural Network (ANN) is a concept that is used in deep learning. The algorithm in ANN generates layers that enter input values from one layer to the next. Figure 1 shows an outcome result of the architecture of ANN.



Source: Virtual Labs, 2005

Figure 1. 1: Architecture of a multilayer feedforward neural network

During data training in deep learning when the information is being processed, there is no human interference with the layers. There is no need for humans to manually handle the training because the system algorithms are trained with data and learning procedures. This method has the ability to handle data with higher dimensionality. Translations of more advanced areas, classification and analysis are some of the promising result that the system of deep learning has showed (LeCun, Bengio, & Hinton, 2015).

1.1.3 Convolutional Neural Networks

Convolutional Neural Networks(CNN) are deep artificial neural networks. Artificial Neural Networks (ANN) are the predecessor to CNNs and first appeared after the development of the perceptron in 1958 (Rosenblatt, 2018). The fundamental element of CNNs is back propagation which minimizes an error function by manipulating the network weights using gradient descent.

Convolutional Neural Networks mainly centers on text, image data and times-series. CNN has different levels of necessity, one based on spatial distances. In spatial distances, it works in gridstructures, which are data with dimensional images and spatial dependencies in the local region, that is related to the color values of each pixel in an image (Aggarwal, 2018). With 3D structured input enables it to capture color. With CNN, it shows a different level of translation and interpretations, which could process an augmented image which is an image that is upside down or shifted in different directions. This is not usual with other grid-structure data (Aggarwal, 2018).

CNN is composed of at least one convolutional layer but can have as many more layers as possible and is regarded as an easy neural network to train. The convolutional layer is followed by a fully connected layer(s) in a standard multi-layer network. An image process through the convolutional layer extract feature from an input that goes through different kernels. An input is down sampled by the pooling layer by reducing the dimensionality of the data making sure the essential features or information is not changed or lost. The previous layers to the next layer neurons is tied to the output by the fully connected layer. There are variables in CNN called hyperparameters which determines the structure of the network (Convolutional neural network, 2021).

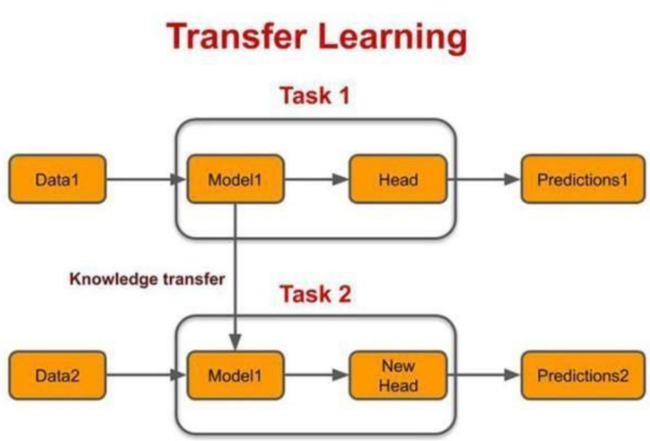
1.1.4 Transfer Learning

Building a model from scratch can take long and can be tiresome. However, in machine learning this can be avoided by using a technique called transfer learning. Transfer learning is technique in machine learning where a model that was developed and used for a particular task is reused as the starting point for a model on another task which is related to original task in order to improve the performance of the model in terms of generalization and other aspects of a model performance.

For instance, one can reuse the knowledge of a classifier that was trained to predict whether an image contains cars and aircrafts (Donges, 2020).

There are several benefits of transfer learning. The main advantages being saving training time, running away from using a lot of data, and better performance of the neural network models. In most cases the huge amount of data that is required in training neural network from scratch is not always available especially in low income areas such as Africa. Consequently, transfer learning come becomes useful in such a circumstance. Since the model that is being used is already a pretrained one, another machine learning model can be built with relatively very little data when training.

The most interesting thing is that there are several ways that transfer learning can be incorporated. For example, only the top layer of a neural network can be used. This is commonly known as freezing the fully connected layer which is at the top of the network in the original model. Freezing reduces training time as the backward passes go down in number. Below is an architecture showing transfer learning where the output layer is frozen.



Source: topbots.com; An Ultimate Guide to Transfer Learning Figure 1. 2: Transfer Learning architecture showing a frozen output layer

MobileNet

In cases where one needs to construct lightweight deep convolutional neural networks, MobileNet becomes handy. MobileNet is an efficient architecture that uses depth wise separable convolutions.

Originally, MobileNet was designed for mobile and embedded vision applications (Howard, Zhu, & Chen, 2017).

For MobileNet, a single filter is applied to each input channel by the depth wise convolution. After that, the pointwise convolution applies a 1x1convolution to combine the outputs the depth wise convolution. In a standard convolution uses both filters and in one step the inputs are combined into a new set of outputs. This is followed by the splitting of depth wise separable convolution into two layers, one layer called filtering layer and another one called separate layer used for combining. This is called factorization and has the impact of lessening the model and computation size (Culfaz, 2018).

The MobileNet architecture has 30 layers with convolutional layer with stride 2 depth wise layer pointwise layer that doubles the number of channels depth wise layer with stride 2 pointwise layer that doubles the number of channels etc. as shown in the figure 31.

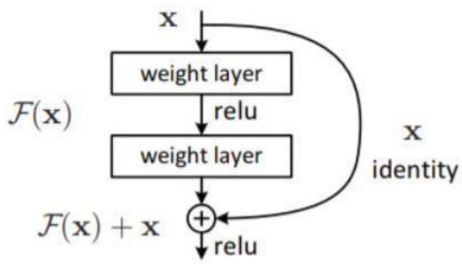
Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5× Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
5× Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Source: Towards Data Science

Figure 1. 3: MobileNet full architecture

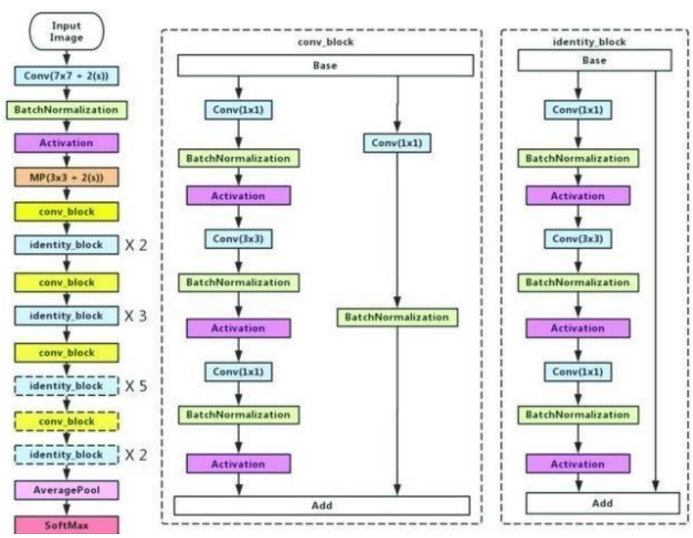
ResNet50

During training, there is a problem of performance decay in the deeper levels of the network. The residual network also known as ResNet has an architecture that was designed to handle this problem (He, Zhang, Ren, & Sun, 2016). In a linear network architecture, relevant information in the data and associated gradients are usually lost because of the down sampling. The short cut connections that are formed are fundamental elements of ResNet and are commonly known as residual blocks (He, Zhang, Ren, & Sun, 2016). A block diagram of the block is shown in Figure 31. ResNet has a number of different common configurations namely based on the amount of convolutional layers within the architecture. The convention is ResNet followed by the number of convolutional layers. In this work ResNet50, the architecture from (He, Zhang, Ren, & Sun, 2016) is shown in Figure 32.



Source: Towards Data Science

Figure 1. 4: Residual block architecture



Source: Towards Data Science

Figure 1. 5: Block diagram of the ResNet50 architecture

1.2 Evaluation Model

Data Science is not simply analyzing and interpreting the data but also evaluating and understanding the results from the analysis. Evaluation is important after carrying out different experiments as it gives results that can be understood by everyone and the performance of the models. There are a number of evaluation approaches that can be used and this section will briefly describe some of the metrics that have been used to evaluate the model performance in classifying Invasive Ductal Carcinoma breast cancer.

1.2.1 Confusion Matrix

One way of evaluating the performance of a classifier is using a confusion matrix. The confusion matrix is a useful tool for analyzing how well a classifier can recognize tuples of different classes. A confusion matrix is commonly presented in an n x n table of predicted versus actual classification where n is the number of different classes that have been used (Visa, Atsushi, & Anca, 2011). A row in a confusion matrix represents an actual class of the data, while each column represents a predicted class.

The columns represent the prediction made by the classifier. Negative prediction in this study is the first column and second column represents a positive prediction. Actual or observation class are represented in the rows with negative prediction and positive prediction in the first and second row respectively.

In the table, there are common terminologies used by the researchers. An observation with a positive prediction and positive prediction is called True Positive (TP). That is positive observations that were correctly classified. True negative (TN) means a negative observation that was correctly predicted by the classifier. False Negative (FN) is a positive observation that were mislabeled as negative. False Positive (FP) is a negative observation that were incorrectly labeled as positive (Visa, Atsushi, & Anca, 2011).

Table 1. 1: Confusion Matrix showing TN, FN, FP and TP

	Predicted 0	Predicted 1
Actual 0	True Negative (TN)	False Positive (FP)
Actual 1	False Negative (FN)	True Positive (TP)

1.2.2 Classification Report

Classification report is also used in understanding the performance of a classifier. It gives a representation of the classification on each and every class used. As such classification report shows a better and more deep perception of a classifier than the overall accuracy that is commonly used that sometimes hide the weakness in one class especially in cases of class imbalance.

Accuracy

Accuracy is a common metric used in deep learning. The accuracy of a classifier is proportion of the observations that have been correctly classified. Thus True positive and True negative divided by the total number of subjects (Jaiwei, Kamber, & Pei, 2012). In this study, a classifier is the neural network model been built to classify cancer images. That is,

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Recall

Recall shows completeness, that is how many positive observations are classified as positive. The higher the recall value the more correct a classifier is. High recall shows a low number of false negatives. Recall is also called sensitivity or true positive rate. It is calculated as shown in the equation 2.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

Precision

Precision is another measure derived from confusion matrix and it is widely used in classification. It can be regarded as an indicator of exactness. It is the number of subjects that are actually positive and have been predicted as positive (Jaiwei, Kamber, & Pei, 2012). High precision demonstrates that the positive classification which means the false positive is low. Equation 3 shows how precision is calculated.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

F1-Score

In order to find balance between recall and precision, F1-score is used. It gives equal weight to precision and recall. It also a good measure where there is a class imbalance where the actual negatives number is large. The F1 combines the properties of both metrics into one (Jaiwei, Kamber, & Pei, 2012). The score uses equation 4 to calculate a value that falls near the values of precision or recall.

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

1.2.3 Validation

When a model is trained, there is need to assess it depending on a given test data. This process is called validation. This is done to test the performance of the trained model a new dataset. It is possible for a model to perform excellent on the train set unlike the test set(Cross-validation (statistics), 2020).

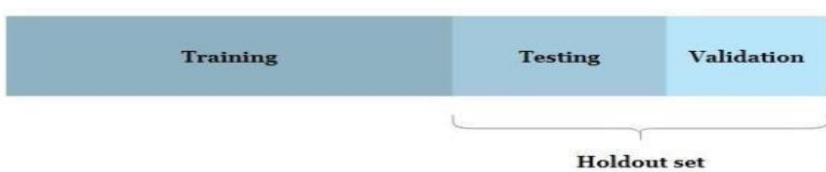


Figure 1. 6: 3-way holdout cross validation

One common method of validating a model is to use the holdout method and is considered as a 3way cross validation. In this method, the dataset is firstly divided into two sets; training and testing sets. A holdout set is taken from the training set for validation. The holdout set is separated from the other sets and is use to tune hyper-parameters. This set is also used to assess the performance of the classifier or predictive on data not seen and this data were not used either training or testing the classifier (DataVedas, n.d.). That is splitting the data set into subdivision for analysis and validating the analysis is part and parcel of validation. One importance of validation is that it decreases the problem of overfitting and bias (Cross-validation (statistics), 2020).

Chapter 2: Related Work

2.1 Literature Review

A number of research studies in the area Machine Learning as well as Artificial Intelligence have worked on image classification using Convolutional Neural Network in medical sector as well as other sectors.

For instance, Ali in 2020 carried a study where two predictive methods of leukaemia (skin cancer) were compared. The two methods are genomic sequencing method and multi-class classification. Genomic sequencing which is figuring out the order of DNA nucleotides, or bases is a binary classification model where as a multi-class classification model is an image-processing method. The two methods used an architecture of a Convolutional neural network (CNN). The two methods on the contrary had different input values. The genome approach performed better compared to the multi-class classification. On one hand, a total accuracy of 98% was achieved by genome method indicating that several numbers of the subjects were classified accordingly with few misclassifications. On the other hand a multi-class classification achieved 81% as a total accuracy (Ali, 2020).

This article is relevant to this thesis by providing related information that has been used in method and discussion by comparing the accuracy of models used in the experiments. The research also has deep learning and neural network information that is relevant to this study.

Musa & Aliyu, 2016 conducted a study on “Application of Machine Learning Techniques in Predicting of Breast Cancer Metastases Using Decision Tree Algorithm, in Nigeria”. In case of breast cancer, the interest is on a classifier that produces less false negative i.e. less number of metastases cases that are classified as not metastases. The false negative of the model that was built was 29 which is a large number. This indicates that the model did not perform well as it failed to capture 29 cases of metastases. This worrisome in the medical field as it can cause 29 patients fail to get treatment in time because they are classified to be non-metastases.

This study contributes to this thesis by giving attention to an important subject of false negatives in classification of IDC breast cancer. Moreover, it also gives direction of the study to look into the reasons of misclassification.

Angel, et al., 2014 discussed in their article on comparing the performance of traditional methods of classifying images versus using Deep Learning. Their method of using Convolutional Neural Network (CNN) produced an F-measure of 71.8% and balanced accuracy of 84.2%. This shows good performance for automatic detection of IDC regions in terms compared to the handcrafted image features “color, texture and edges, nuclear textural and architecture”, and a machine learning classifier for invasive tumor classification using a Random Forest.

This study adds value to this thesis because it discusses the relevance of deep learning techniques and analyzes different automated approaches that can be used in classifying IDC breast cancer images and which methods give best results.

An article by Jessica Olah (2019) from VeryWell explained that using the microscope it is not an easy task to differentiate the breast cells and the normal cells depending on the grade of the tumor. Cancer cells that are in their early stages appear as the normal breast cells. Usually the cells with cancer are arranged in clusters and appear to be attacking the lymphatic vessels and the blood cells. Their nuclei have irregular shape, they are large and they turn darker when dyed with special dyes. Moreover, they are also more than one nuclei.

This article is significant to this thesis because it provides the reason why the models built may have failed to classify cancer images as cancerous. This maybe be due to the fact that images of cancer at early stage are identical to non-cancerous images.

Recently, deep learning has been used widely in several areas especially the medical sector in as far as image classification task is concerned. Xie & Richmond in 2018 conducted a study to see the significance of color in classification of medical images. In this study an Inception-V3 model was pre-trained on ImageNet with colored images and after converting the images to gray scale. The idea was to find out if color is significant in medical image classification. The results showed that there was no any significant difference in the performance of the Inception-V3 model on the colored images and the grayscale images showing that color is not a critical factor in image classification. However, the pre trained models performed better on the grayscale images than on the colored images

This article contributes to this thesis by providing another way of classifying medical images by transforming them into grayscale images. It also shows the relevance of using already available pre-trained models such as Inception-V3 in classification of medical images.

With respect to all related work mentioned above, this research project will build a CNN model and evaluate its performance using IDC breast cancer dataset. It will also use the available pretrained models such as MobileNet and ResNet50 in classifying the IDC breast cancer images. The goal is to achieve the best accuracy with lowest false negative rate in analyzing the breast cancer data.

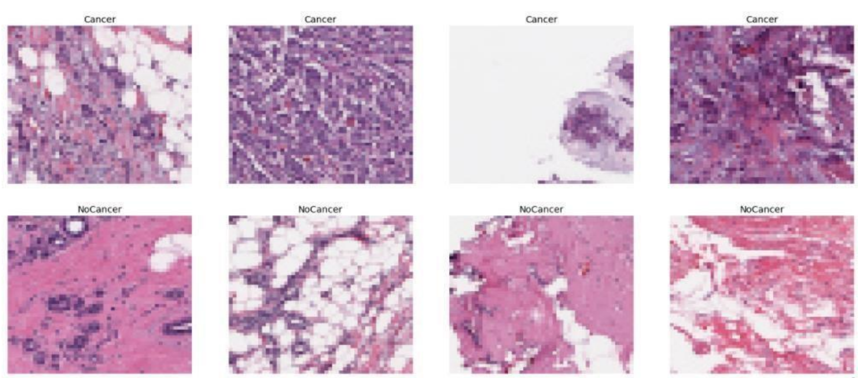
Chapter 3: Data Exploration

3.1 Data Description

The data used in this thesis is from kaggle.com cohort and contains digitized Breast Cancer histopathology slides from 162 women screened for Invasive Ductal Carcinoma(IDC) at the "Hospital of the University of Pennsylvania and The Cancer Institute of New Jersey".

The experts digitized the slides with a scanner at 40x magnification. The images were down to a 50 by 50 pixels. There was a total of 27,750 smaller images that were taken from the original images. An expert pathologist labeled the images in two groups; "IDC positive" or "IDC negative" using a method called manual delineation. The group of "IDC positive" had a total number of 7878 images and 19872 images belonging to the "IDC negative".

For this study, the dataset was further randomly split into 3 subsets comprising: 19,426 (13 911 no cancer images and 5515 cancer images) training, 4,162 validation cases (2981 no cancer images and 1181 cancer images) for parameter exploration and 4,162 test cases (2980 no cancer images and 1182 cancer images) for final evaluation.



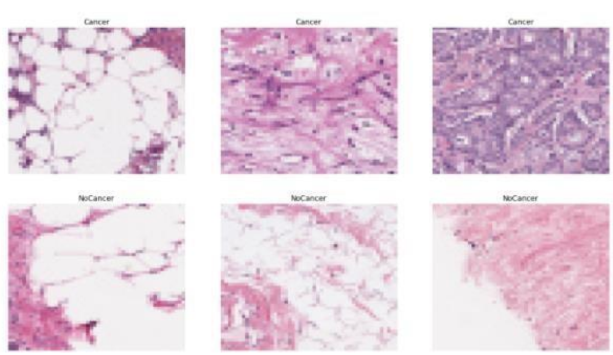
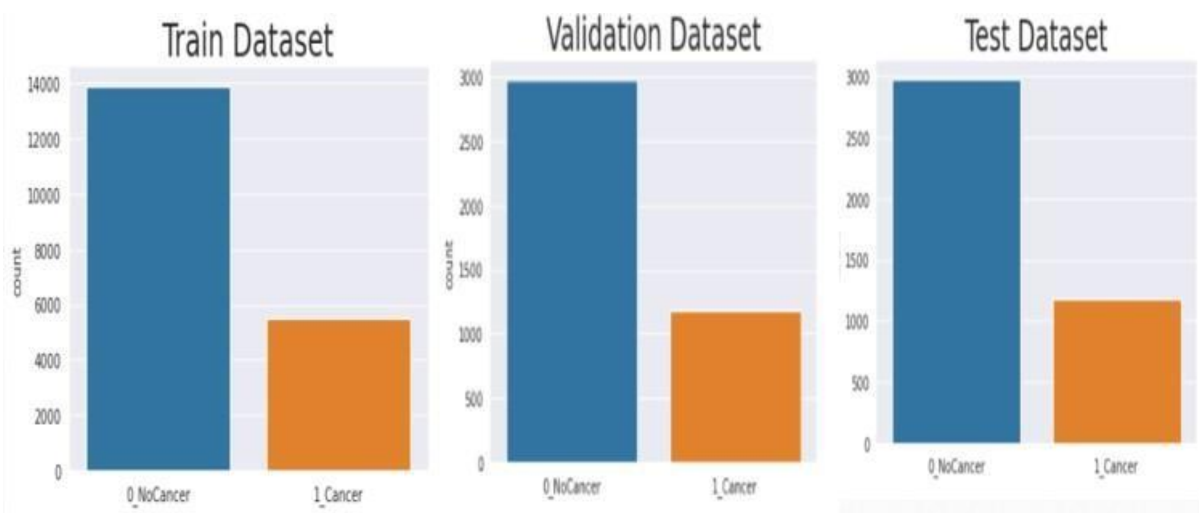


Figure 3. 1: Randomly selected images of cancer and non-cancer images

From the images, there is a considerable variability in the appearance of the cancer and non-cancer images. Most non cancer images are pinkish and cancer look more violet images with no big patches. Although this variability is not in all images. This shows how using physical eyes can be a challenge in identifying the images with IDC breast cancer or not hence the use of deep learning which can learning important features in classifying the images.



(a) (b) (c)

Figure 3. 2: Class Distribution; (a) Train dataset, (b) Validation dataset and (c) Test Dataset

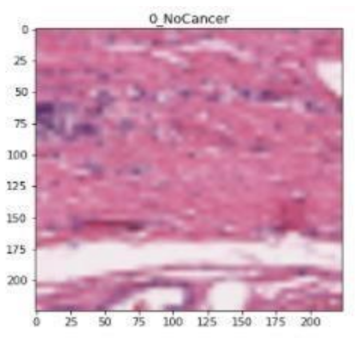
3.2 Pre-processing of image data

Image processing is an important step before image classification. This is important step to improve the appearance of the images and enable to get information that is helpful and useful for the task of classification. Image pre-processing also involves defining that an image is the output and the input is the features or characteristics related to the images. In this case the output is the class of the image, that is the image being cancerous or non-cancerous. Images are converted into a two-dimensional arrays of pixels. An x,y math function is used where x and y denotes the coordinates of the images (Canuma, 2018). The pixel values in the 2-d arrays are in a 0 to 255 range. The image color scale, image height, width and number of pixel/levels are parameters used in the image input (Canuma, 2018).

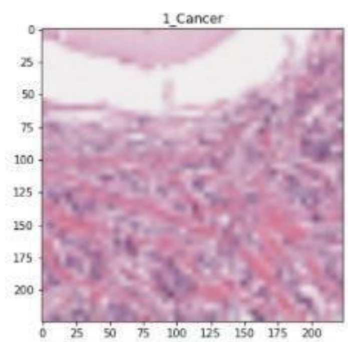
There are two color scale used in this study. The first experiments used the three color channel of Red, green, Blue(RGB) and the other experiments used grayscale also known as one color channel to see the significance of color in image classification.

The first step in the pre-processing stage was to put the images in the same base dimension for easy classification. In case of this dataset, the images were already in the same base dimensions. The next step was to resize all the images to 224x224 to make sure that all have the same target size. In order to create a similar distribution of data, the images were normalized. In normalization the values of the pixels are converted to a range of 0 and 1. Normalization ensures that the network learning process is fast because the network uses the values of the weights to process the inputs which are the features (Canuma, 2018).

Processed images of sizes 224x224



(a)



(b)

Figure 3. 3: Processed images of sizes 224x224. (a) Non-cancer image and (b) Cancer image

Chapter 4: Methods

This chapter describes all the methods and process that have been selected in this study. The process of this thesis aims to explore how deep learning is applied in medical sector in as far as image classification is concerned.

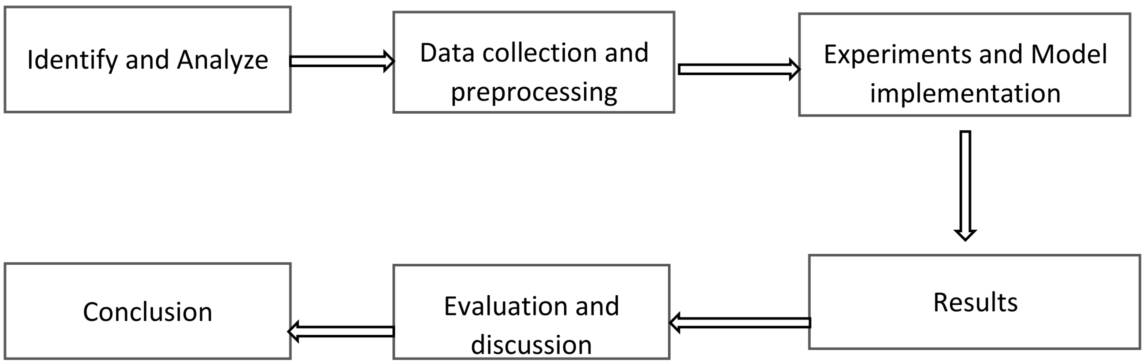


Figure 4. 1: System methodology of the thesis

4.1 Identify and Analyze

In the method of identifying and analyzing is crucial because it is this step that an idea is created from a problem. Using related works and literature study on breast cancer image classification using deep learning helps identifying major and interesting problems. The problems that have been identified are further used to formulate objectives of the study.

4.2 Data Source and preprocessing

The data to be used depends on the problem identified in the initial step. In this case it is classification of Invasive Ductal Carcinoma Breast Cancer as such breast cancer images were collected and sourced and this has been presented in chapter 3. This chapter has expounded how data has been sourced. It has further clarified how the images sourced have been preprocessed to the point they are ready for classification.

4.3 Experiments and model implementations

This step includes all the experiments that have been carried in the thesis. The experiments originated from the specific objectives of the thesis. The experiments include building a CNN model from scratch as well as using transfer learning to classify the IDC images with both colored images and images on grayscale. The experiments have been detailed in chapter 5.

4.4 Results

This step provides clarity on the results obtained from the experiments that were carried in the experiment and model implementation step. This is also detailed in Chapter 5

4.5 Evaluation and discussion

This stage involves observing and evaluating the results that have been achieved from all the experiments that have been carried out in this study. Evaluation and discussion gives an opportunity to compare the results and performance of the models used. The results of all the experiments have been presented in chapter 6, and it is explained using learning curves, accuracy curve and confusion matrices. The results that are achieved are used as answers to the research questions for the study. In most of the experiments, the confusion matrix, learning rate and classification report have been used to assess, compare and summarize how the models have performed.

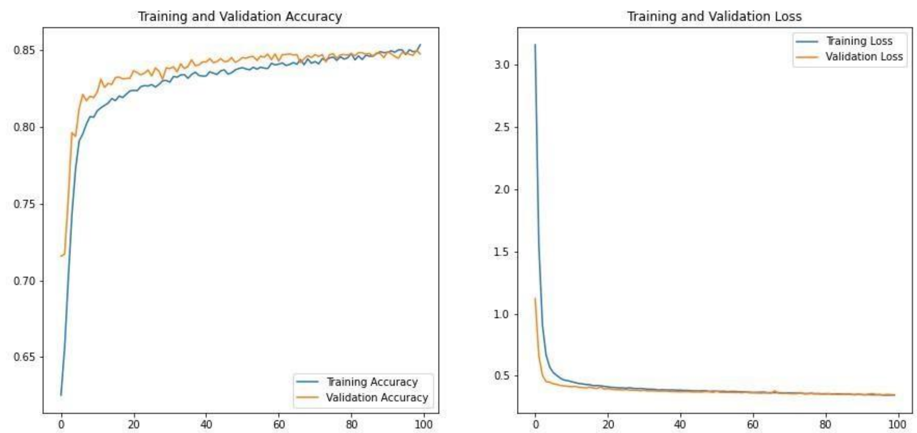
4.6 Conclusion

The conclusion section shows the comparative analysis on the performance of all the models and experiments that have been carried out in the study. This has been achieved by using the evaluation metrics such as accuracy, precision, True positive, True negative, False Positive, False Negative and other metrics discussed in the theory chapter. Evaluation has been done on their performances and values of the results are compared. Then conclusion section is where the research comes to an end and conclusions are drawn from the results. Chapter 6 presents the conclusion section.

Chapter 5: Experiments and Results

5.1 Convolutional Neural Network (CNN)

The system used a 5-layers CNN architecture with 32 neurons for the first layer, 64 for the second convolutional layer, 128 for the third layer, 64 neurons for the fourth convolutional-pooling layers and 32 neurons for the fully-connected layer. A fixed convolutional kernel of size 3×3 was incorporated with a pool kernel of size 2×2. In order to reduce the number of parameters that are trainable, drop out layers were used. Max-pooling layer is used to have a smaller size of the featured map as such it was used. It is also important to convert the feature map dimension to 1D array which follows a fully connected layer accompanied by Softmax activation function and two neurons and this achieved by using a flatten layer. Different learning rates were used with Adam optimizer because Adam optimizer achieved better results compared to SDG. Since there are two classes; IDC positive and IDC negative a loss function of binary cross entropy was regarded. The number of epochs was 100 in most of the experiments with 32 as the size of the batch



(a) A plot of accuracy against number of epochs and (b) a plot of loss against number of epochs for CNN model

Figure 5. 1: (a) Loss plot and (b) Accuracy plots for CNN model

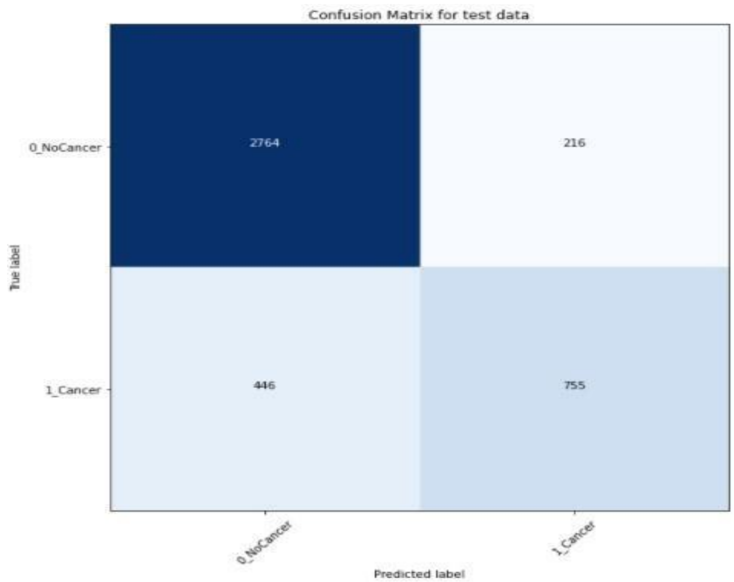


Figure 5. 2: Confusion matrix results on test set

Table 5. 1: Classification report for CNN model

Class	Precision	Recall	F1-Score	Accuracy
IDC(-)	0.86	0.93	0.89	0.84
IDC(+)	0.78	0.63	0.70	

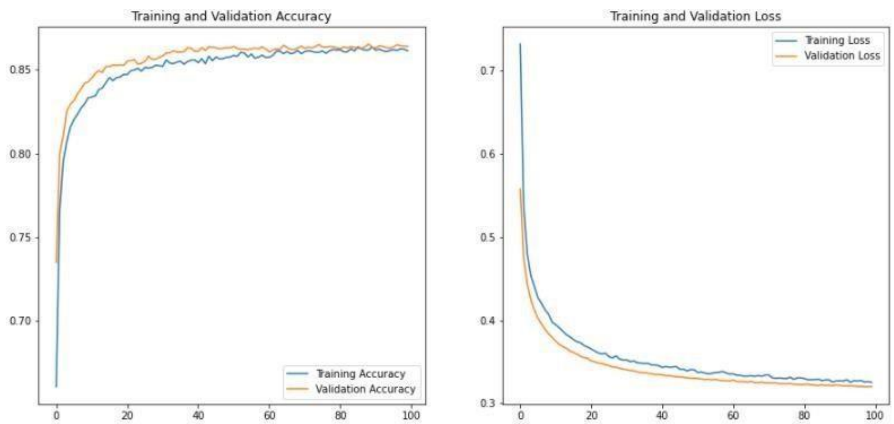
The confusion matrix and the classification report shows that the model achieved an accuracy of 84 percent. This means the model correctly classifies 84 the IDC images in every 100 images. The TP is 63% of correctly classifying positive IDC images. The TN is 93% of correctly classifying non IDC images.

5.2 Transfer Learning

Transfer learning is mostly used in machine learning task where a model that was developed and trained for a particular task is reused for a similar task as a second model. The first model is considered as a starting point.

5.2.1 MobileNet

Carried on with the pre-trained/original weights i.e. weights= “imagenet”. The very first layer is input layer which accept image size = (224, 224, 3). The image sizes are different hence changed the parameter - image_size in the first layer to be (224,224, 3). Used all the pre-trained convolutional layers and freezing the top layer. The last layer of MobileNet has 1000 classes hence had to be changed for this project where there are two classes.



(a) (b)

Figure 5. 3: (a) A plot of accuracy against number of epochs and (b) a plot of loss against number of epochs for MobileNet model

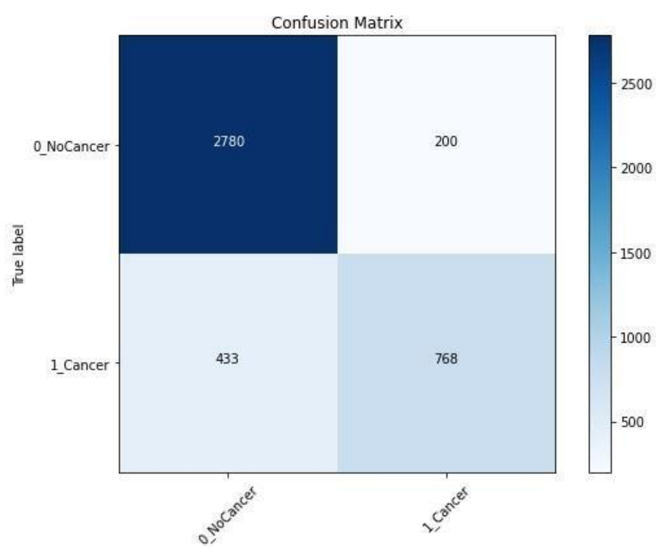


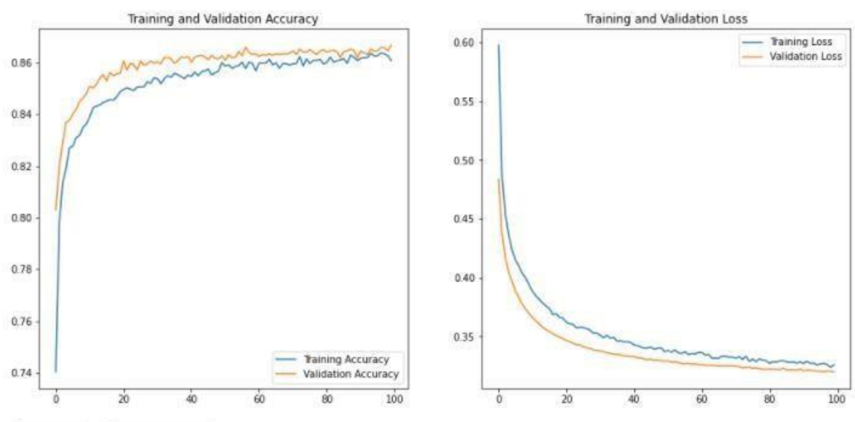
Figure 5. 4: Confusion matrix for MobileNet Model

Table 5. 2: Classification report for the MobileNet model

Class	Precision	Recall	F1- Score	Accuracy
IDC(-)	0.87	0.93	0.90	0.85
IDC(+)	0.79	0.64	0.71	

5.2.2 ResNet50

In this project ResNet50 with pre-trained imagenet weights was used. Used all the layers except the fully connected layer at the top of the Network in the original model. This is regarded as freezing the top layer of the ResNes50 and train the top layer. The dropout layer was also included.



(a) (b)

Figure 5. 5: A plot of accuracy against number of epochs and (b) a plot of loss against

number of epochs for ResNet50 model

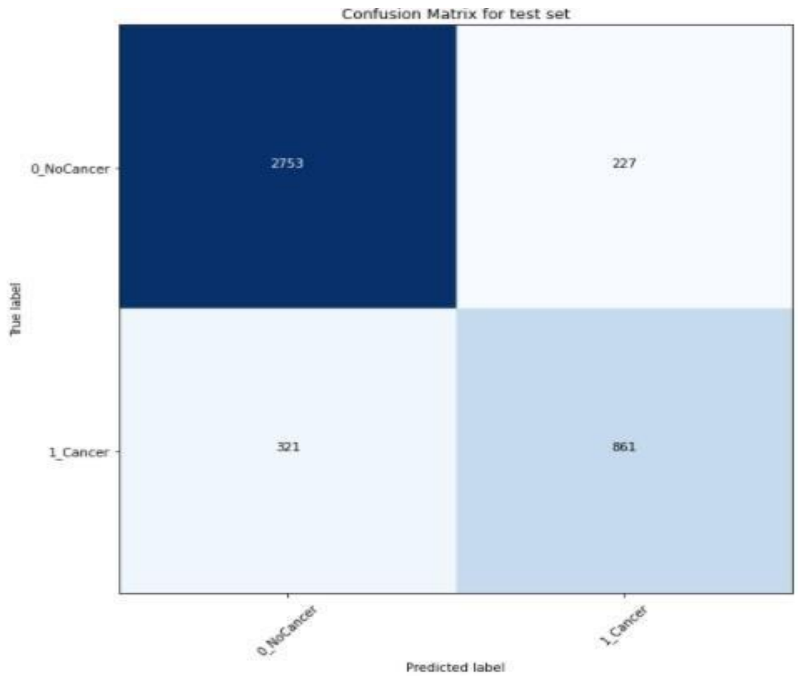


Figure 5. 6: Confusion matrix for the ResNet50 model

Table 5. 3: Classification report for ResNet50 model

Class	**Precision**	**Recall**	**F1- Score**	**Accuracy**
IDC(-)	0.90	0.92	0.91	0.87
IDC(+)	0.79	0.73	0.76	

Table 5. 4: Summary of the results: Classification results for all the models

Model	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
CNN	63%	7%	93%	37%
MobileNet	64%	7%	93%	36%
ResNet50	73%	8%	92%	27%

Table 5. 5: Summary of experiment results

Model	Architecture	F-Score	Precision	Accuracy
CNN	CNN	0.79	0.82	0.84
MobileNet	CNN	0.80	0.83	0.85
ResNet50	CNN	0.83	0.84	0.87

5.3 Misclassification

5.3.1 Understanding Misclassification in ResNet50 using Saliency Maps

In regard to recall, precision and accuracy, ResNet50 model outperformed the other models (CNN from scratch and MobileNet). Moreover, ResNet50 has less False Negatives which is the main concern in this study compared to the other models. This study further analyzed why the model was misclassifying many cancer images as negative by using saliency maps.

Saliency maps are used to measure the spatial support of a particular class in each image. It is a common way of interpreting the predictions of a Convolutional Neural Network. The saliency map is built using gradients of the output over the input. This highlights the areas of the images which the model found to relevant for the classification. In some cases, the model concentrate on areas that are not relevant to classification which may lead to misclassification. As such it is important to understand the areas that the models used in this thesis regarded to be relevant for classification.

Saliency maps on True positive images

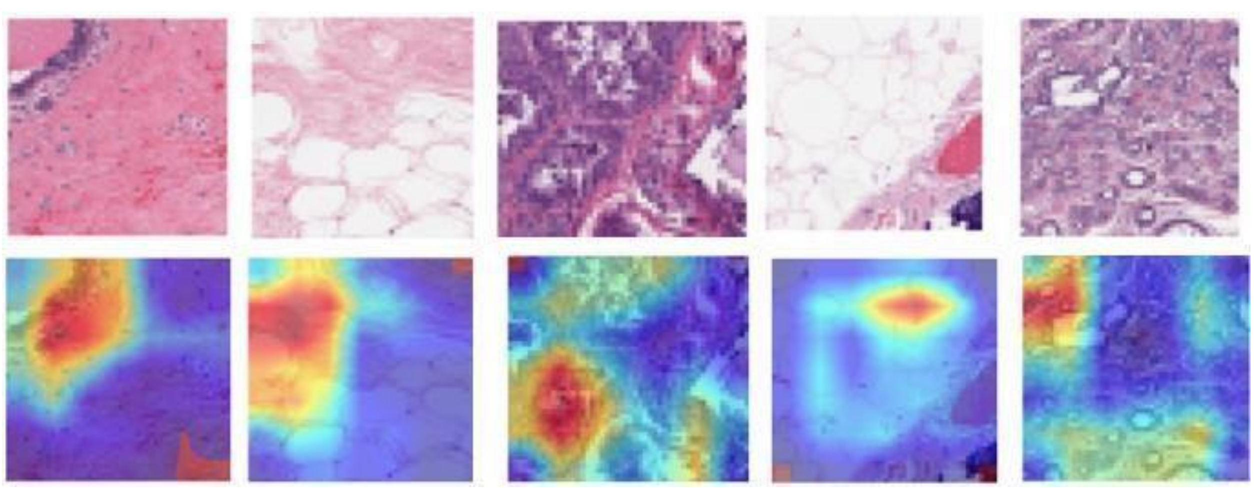


Figure 5. 7: Saliency maps on True positive images

Saliency maps on False Negative images

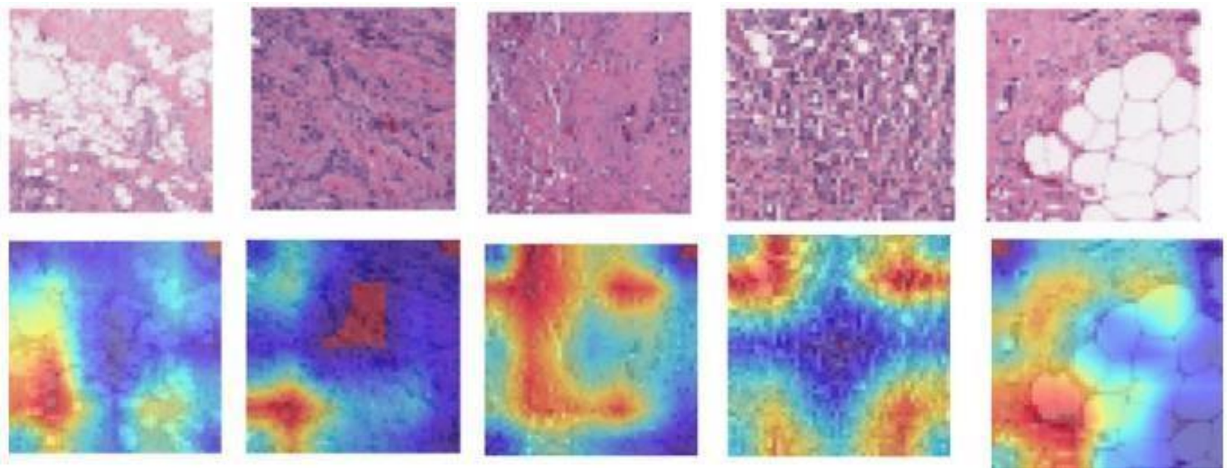


Figure 5. 8: Saliency maps on False Negative images

Saliency maps on True Negative images

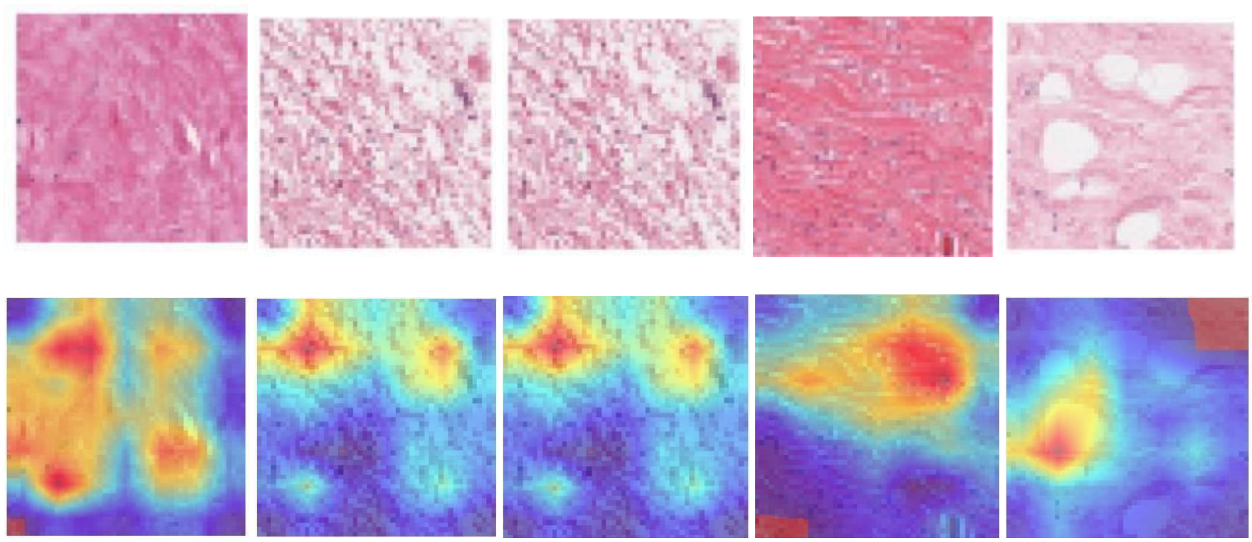


Figure 5. 9: Saliency maps on True Negative images

Saliency maps on False Positive images

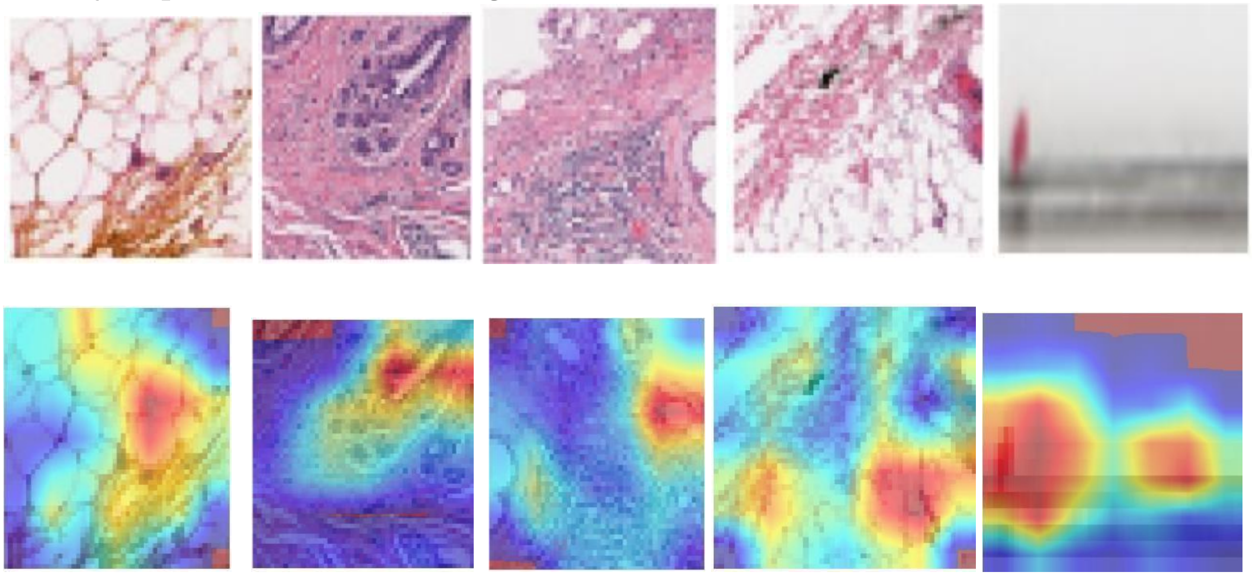


Figure 20: Saliency maps on false positive images

There is no trend or pattern that can be seen on saliency maps for both the true positive and false negative. However, the Saliency maps on the false negative images shows that most of them are paying attention to many regions which is leading to confusion to the model hence misclassifying the images. While for the true positive, the model is paying much attention mostly on one region.

5.3.2 Understanding color intensity in the Cancer and Non Cancer Images

Image histogram is widely used in image classification to give a graphical representation of how the pixel intensities were distributed in a digital image. Large peaks show low intensity while small peaks on the histograms show high intensity. Smaller numbers of close to zero of the pixels represent black and larger numbers closer to 255 represent white.

Pixel Intensity Histograms for cancer images

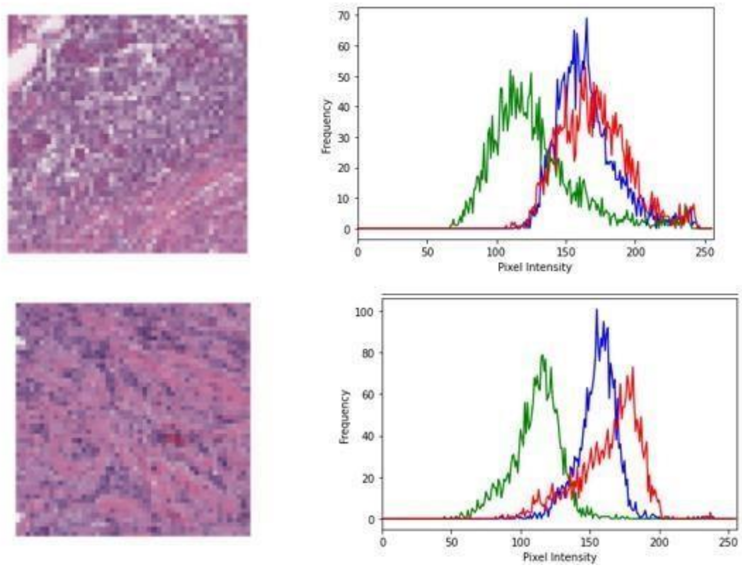


Figure 5. 10: Pixel Intensity Histograms for cancer images that have been selected

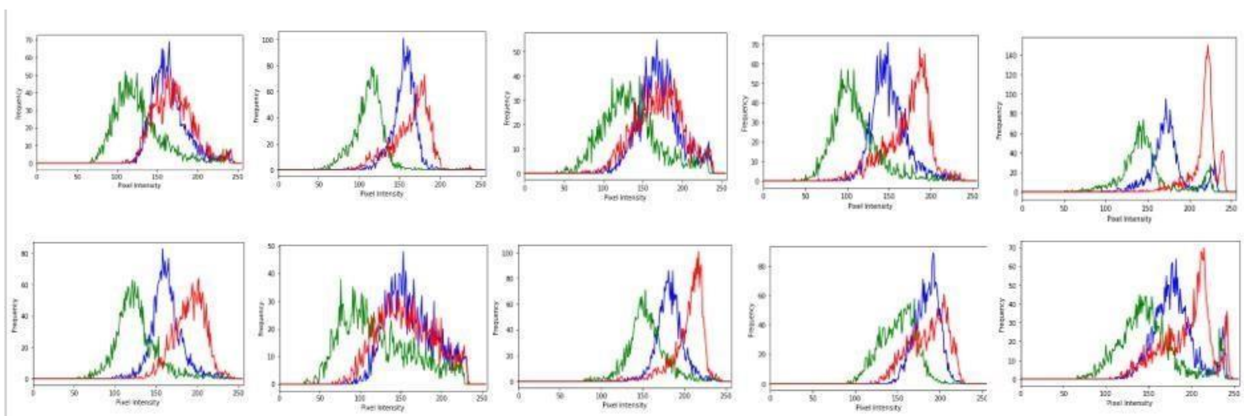


Figure 5. 11: Visualizing more pixel histograms for cancer images to understand the pattern of the Red Green and Blue Colors

Pixel Intensity Histograms for Non-cancer images

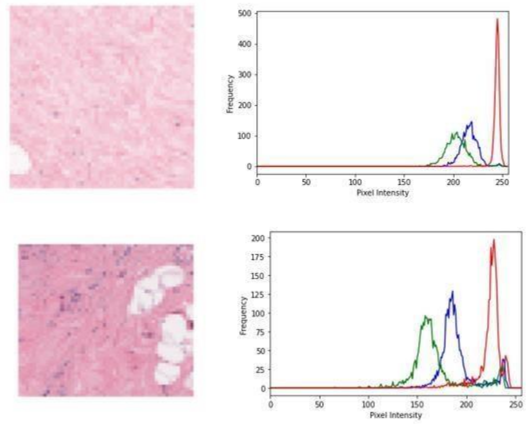


Figure 5. 12: Pixel Intensity Histograms for Non-cancer images that have been randomly selected

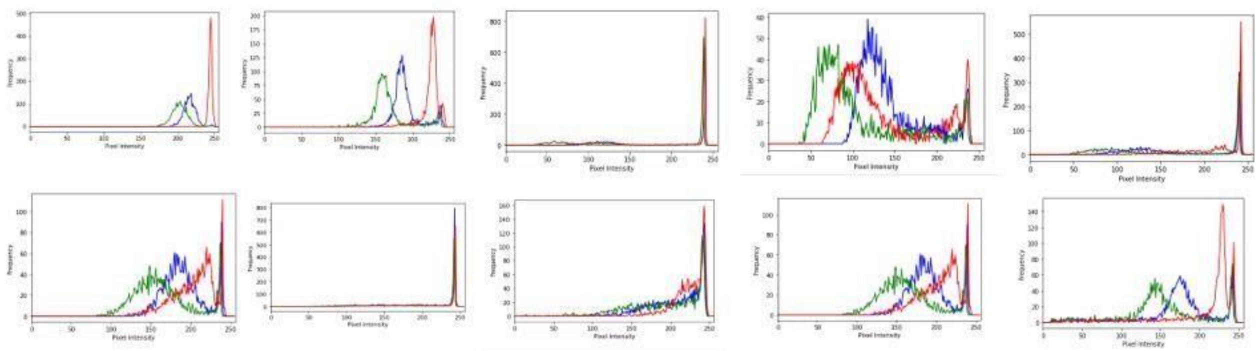


Figure 5. 13: Visualizing more pixel histograms for Non-cancer images to understand the pattern of the Red Green and Blue Colors

Visualizing the average of Pixels

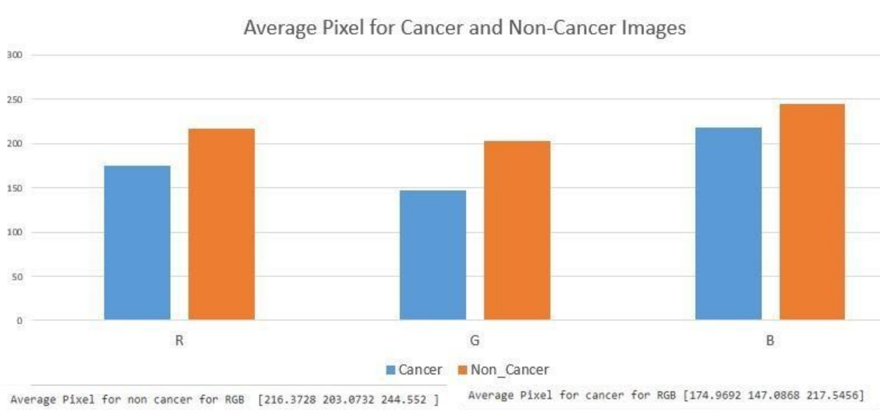


Figure 5. 14: Bar graphs showing the average pixels of Red, Green and Blue for the Cancer and Non-Cancer group

Discussion on pixel images

On one hand most of the cancer images have large peaks between 100 and 200 with all the colors roughly having the same height, on the other hand the non-cancer images have high peaks between 150 and 250 but green appears to have the highest peak compared to red and blue.

This shows that the non-cancer images are lighter than cancer images. However, there are also other cancer images that appear to be darker. Nevertheless, this does not show if the variation in color has an effect on the classification of the IDC breast cancer images. In order to verify if color variation has an impact in the classification task, few models have been built on the grayscale images.

5.4 Models on Grayscale images

In order to understand the significance of color on IDC breast cancer image classification, models were built using grayscale images. The images were changed from 3 color channel to one color channel and changed the input size to (224,224,1).

The same CNN architecture that was trained from scratch was used. This means all the optimization parameters remained constant after converting the images to grayscale in order to see the difference in the performance of the models.

For ResNet50, used the same architecture as well and modify the original model (ResNet50), changing the input format to deal with grayscale and of course training the classifier to have two outputs; cancer or non-cancer.

Visualizing Images in Grayscale scale

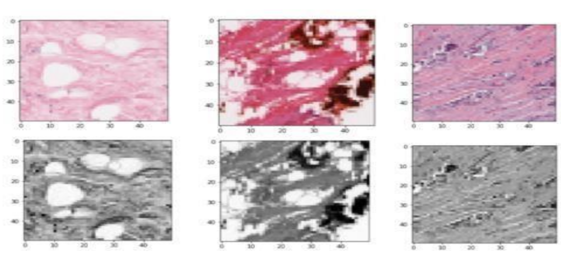


Figure 5. 15: Showing images that have been transformed from color scale to grayscale

5.4.1 CNN model on Gray Scale images

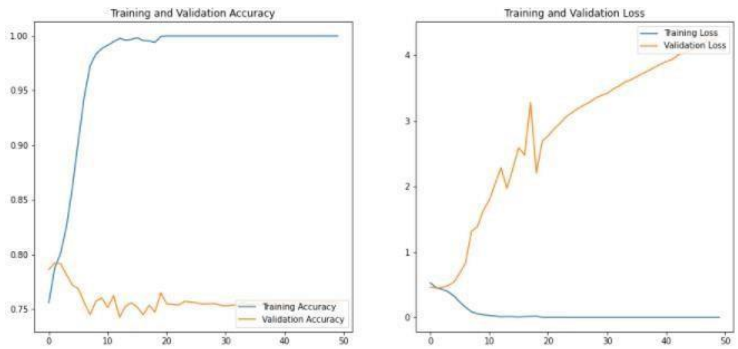


Figure 5. 16: Loss and Accuracy plots against number of epochs of the CNN Model on grayscale images

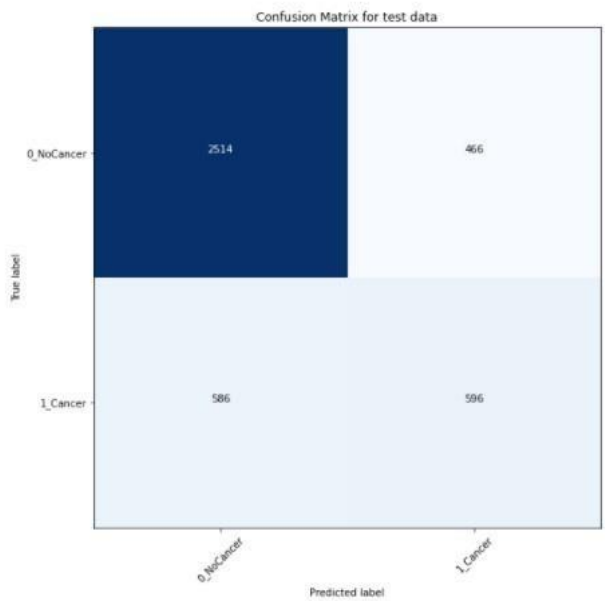


Figure 5. 17: Confusion matrix for CNN model on grayscale images

Table 5. 6: Classification report of CNN model on grayscale images

Class	**Precision**	**Recall**	**F1- Score**	**Accuracy**
IDC(-[redacted])	0.81[redacted]	0.84[redacted]	0.83[redacted]	0.75
IDC(+)	0.56	0.50	0.53	

5.4.2 ResNet50 on grayscale images

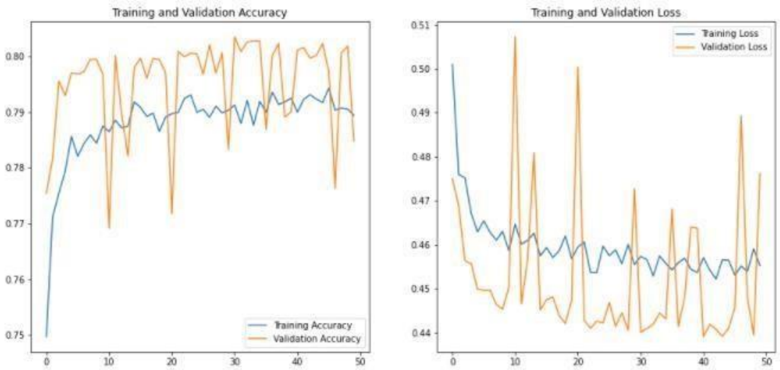


Figure 5. 18: Loss and accuracy against epochs plots for ResNet50 on grayscale images

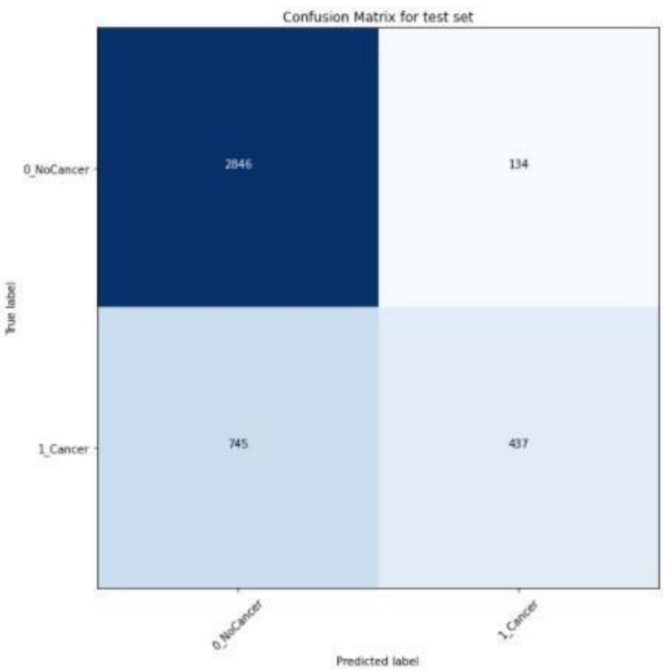


Figure 5. 19: Confusion matrix for ResNet50 on grayscale images

Table 5. 7: Classification Report for ResNet50 model on grayscale images

Class	**Precision**	**Recall**	**F1- Score**	**Accuracy**
IDC(-)	0.79	0.96	0.87	0.79
IDC(+)	0.77	0.37	0.50	

5.4.3 Using Regularization to improve the Results of CNN Model on gray scale

It is clear that there is a problem of overfitting in the CNN model on grayscale images. Regularization is one of the techniques of dealing with overfitting as such the technique was used in this scenario. Applying L2 regularization to a Dense layer.

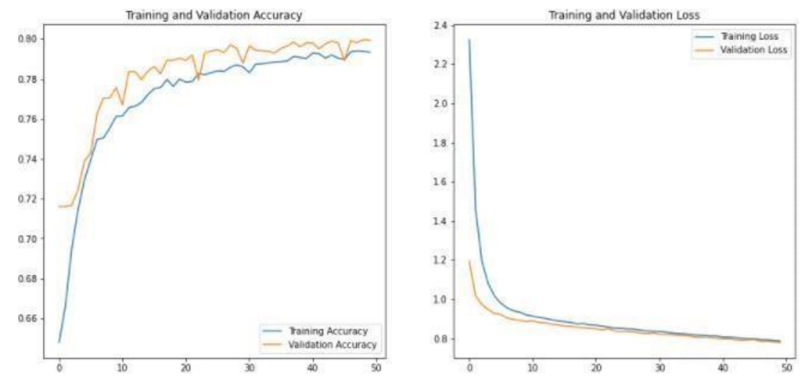


Figure 5. 20: Loss and Accuracy against epochs of CNN model on grayscale using regularization

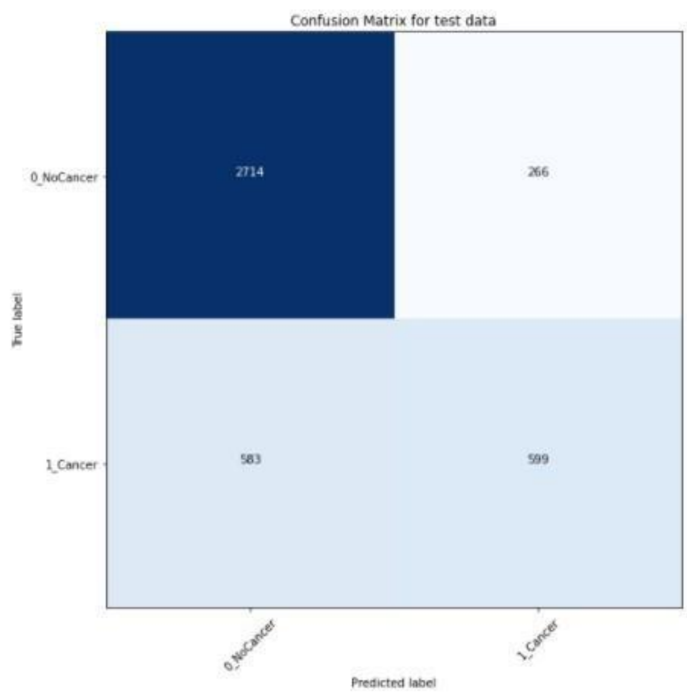


Figure 5. 21: Confusion matrix of CNN model on grayscale using regularization

Table 5. 8: Classification report of CNN model on grayscale using regularization

Class	Precision	Recall	F1-Score	Accuracy
IDC(-)	0.82	0.91	0.51	0.80
IDC(+)	0.69	0.51	0.59	

Choice of regularization parameter

Accuracy and Loss vs L2 regularization

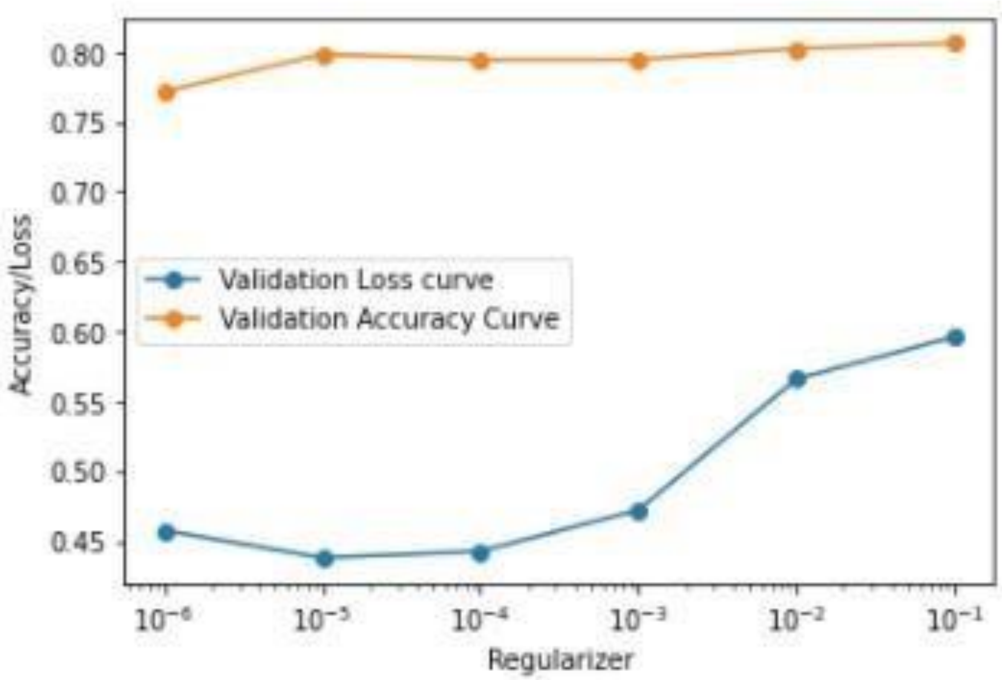


Figure 5. 22: A plot of accuracy and loss versus regularization parameter

Discussion

Using the line plot in figure 5.22, the plot shows the validation accuracy increases with larger weight regularization parameter values to a point (0.001) where the accuracy begins to drop. The largest value of regularization of 0.001 (10^{-3}) results in a large drop until a point where regularization value is 0.1(10^{-1}), the accuracy in the validation set starts to increase again.

From the loss plots, the loss values are higher where the regularization parameter values are also high. Gradually, the loss value increases as the regularizer starts to increase from 0.1(10^{-1}).

The loss plot and accuracy plot are complimenting each other. The loss value is higher (at regularizer value of 0.01(10^{-2})), where the accuracy is the lowest. And when the loss value starts to decrease, the accuracy starts to increase. As such the regularization parameter chosen in the grayscale model is 0.1(10^{-5})

Chapter 6: Discussion

6.1 Summary of the results on the three models

In this thesis of classifying Invasive Ductal Carcinoma Breast Cancer, a CNN based approach was introduced. The thesis also introduced the application of transfer learning in classification of IDC breast cancer images. Two different types of training thus training from scratch and using transfer learning for IDC Breast Cancer classification were examined.

In the first method of a building a new CNN model, a convolutional neural network with four layers was developed and trained. Using the testing set, the model achieved an accuracy of 84%. Table 5.1, Table 5.5 and Table 5.6 shows the quantitative results of the CNN classifier. Among the cancer images in the test set 63% of them were correctly classified as cancer images (TP) while 37% of them were misclassified as non-cancer images (FN). Furthermore, out of all the non-cancer images, 93% of them were correctly classified as non-cancer images (TN) while 7% of them were misclassified as cancer images (FP). The confusion matrix in figure also show the TP, FN, TN and FP.

According to figure 5.1, figures (a) and (b) show the performance of the CNN model using accuracy plot and loss plot. The model loss plot shows that the CNN model had some uncertainties and accuracy measured the model prediction performance. The distance between the training and validation line figure 5.1 (a) of the accuracy plot is small, and almost no distance in the model loss plot in figure 5.1 (b).

In the second method, transfer learning was used to classify images. Firstly, MobileNet CNN model was used as a pre-trained model for feature extraction. The overall accuracy of the test set was 85% after training the model. It also achieved prediction results as: TP=64%, FN=36%, TN=93% and FP=7%. Secondly, ResNet50 CNN model was used and achieved 87% as an overall accuracy. The prediction results for ResNet50 was TP=73%, FN=27%, TN=92% and FP=8%.

The model accuracy and model loss for MobileNet in figure 5.3 (a) and (b) shows a small distance between the validation line and training line. However, around 10 epochs, the line distance starts to separate from each other. The distance between the validation and train line gets smaller again from around 60 epochs.

In figure 5.5 (a) and (b) it shows the performance of the ResNet50 model using model accuracy and loss plots. The distance is small between the validation line and training line. However, the distance is wider compared to the CNN model built from scratch and MobileNet.

Table 5.1, table 5.2 and table 5.3 show the classification reports of the CNN model built from scratch, MobileNet and ResNet50 respectively. The classification reports measured the models' prediction based on the confusion matrix obtained from figure 5.2, 5.4 and 5.6 from CNN model built from scratch, MobileNet and ResNet50 respectively. The calculations are based on the equations 1-6 and compiled into tables.

6.2 Analysis result

The main methods that have been used in this thesis to classify IDC cancer images used a Convolutional Neural Network (CNN) architecture. The evaluation of the models has been presented in chapter 5 and chapter 6.1 using a confusion matrix, classification report and learning curve. The different evaluation metrics were employed to examine the performance of all the models and usde the results as a base for comparison.

In order to answer the research question/objective of identifying which method is suitable for this data set, the results from the evaluation tools have been used to analyze the performance of all the models that have been used.

Using accuracy, precision and F1-score from table 5.5 that has been calculated from each model's confusion matrix, ResNet50 has performed well compared to the other two methods. ResNet50 obtained 87% as an overall accuracy while CNN from scratch obtained 84% and MobileNet50 obtained 85%. And the precision was 0.82 for CNN from scratch, 0.83 for MobileNet and 0.84 for ResNet50. The F1-score was 0.79, 0.80 and 0.83 for CNN from scratch, MobileNet and ResNet50 respectively.

Precision and recall results from the cancer and non-cancer class has been analyzed in order to understand the performance of the models in the two classes of the dataset.

Using table 5.1 which shows the classification report for CNN from scratch. The non-cancer class obtained 0.86 and 0.93 for precision and recall respectively while the cancer class obtained 0.78 and 0.63 for precision and recall respectively. This means that the model was more exact and complete in the non-cancer class. Precision and recall show the cost of False Positive for the precision and False Negative for the recall. In this case recall is lower in cancer group compared to non-cancer group. The error of misclassifying the subjects which can lead to misdiagnosis of patients is a big challenge if implemented in real world and has detrimental consequences.

Table 5.2 shows the classification report for MobileNet. The model has 0.87 and 0.93 precision and recall result respectively in the non-cancer class. In the cancer class there is 0.79 and 0.64 precision and recall results respectively. This also has the same interpretation as to the CNN model results. The recall has increased by 0.1 in the MobileNet.

ResNet50 shows its classification report in table 5.3. According to the classification report, precision had 0.90 and recall had 0.92 for the non-cancer class. However, in the cancer class, precision is 0.79 and recall is 0.73. Precision and recall rate is still lower in the cancer class compared to the non-cancer group, however, the results have increased compared to CNN and MobileNet.

In all the methods, precision and recall is high in the non-cancer class compared to the cancer class. This means that the models are doing better in the non-cancer group. This results can cause harm in real life application.

The false negative was one of the part that mattered most in this thesis. False negatives in the confusion matrices used in this thesis show the number of IDC breast cancer images that have been classified as negative by the model. This is crucial in medical sector because the results indicated by the false negatives would determine if a model can be useful or can kill people. A model with a lot of false negative would be questionable before it is used because it means a lot of patients will not be able to receive treatment in time having being identified negative. As a result, a lot of patients would be diagnosed at a late stage leading to increase in the number of deaths caused by IDC breast cancer which would have been prevented if the model had a lower false negative rate.

Convolutional Neural Network built from scratch, MobileNet and ResNet50 had 37%, 36% and 27% respectively as false negative rate. In case of using False negative rate as a measure of performance in the methods, Resnet50 performed better comparing to CNN and MobileNet because it had a lower false negative rate. As such ResNet50 would be recommended to be used in a hospital using this result.

Saliency maps in chapter 5.3 have explained some of the reasons behind higher false negative rates in the models used. The dataset had some junk images that lead to the models being confused in classifying the images. As such there is need for a good procedure in checking the images to be suitable for CNN used and good image processing methods where the images are being digitalized by a scanner.

Literature has also explained the reason of higher false negative rates in medical image classification. In cases where the tumor is small that is early stages of cancer, it not easy to identify an image as to have cancer because the images appear almost the same (Skandalakis, 2009). Depending on the grade of the tumor there can be either poor differentiation or well differentiation of the breast cancer cells and normal breast cells. This can also be the case with the dataset that has been used with different cancer stages. One way to consider is to use multiclass classification. This can also handle the issue of misclassification.

In section 5.5 the thesis presents the results of training a CNN model and ResNet50 using the same dataset but converting the images from color scale to grayscale. The images were changed from 3 color channel to one color channel and changed the input size to (224,224,1). In order to be able to compare the results of the models from images on a color scale and gray scale, the models for grayscale images used the same architecture of the CNN model from scratch and ResNet50 of the images on the color scale.

Figure 5.16 represents the loss curve and accuracy curve of CNN model on grayscale images. The distance between the validation set and trainset is wide showing overfitting. On one hand the model is performing well with its loss curve showing a decrease and converging and the accuracy curve has higher numbers and becoming stable at around 20. On the other hand, the model is poorly performing on the validation set. The loss curve is going up instead of dropping and the accuracy curve is lower than that of the train set.

In figure 5.17 the thesis presents the confusion matrix for CNN model using grayscale images and in table 5.6 it shows the classification report. The two shows an overall accuracy of 75%, recall of 84% in non-cancer class and 50% in cancer group. Recall is 81% in non-cancer class and 56% in cancer class.

The performance of ResNet50 has been presented in the loss plot and accuracy plot in figure 5.18 for the grayscale images. Although the distance between validation curve and test curve is not wide in both the loss and accuracy plot, but the plots show that the model did not converge because the loss plot has not stabilized yet.

In figure 5.19 and table 5.7 shows the confusion matrix and classification report respectively for the ResNet50 of grayscale images. The overall accuracy achieved by this model is 79%. Precision for non-cancer class is 79% and for the cancer class is 77%. The recall is 96% for the non-cancer group and 37% for the cancer group.

According to overall accuracy, precision and recall, CNN model from scratch and ResNet50 trained on colored images is performing better compared to CNN model from scratch and ResNet50 trained On grayscale images. One interesting thing to note is that all the models trained on CNN model from scratch and ResNet50 perform better in non-cancer class compared to cancer class using recall and precision measure.

Table 6. 1: Results from color and grayscale models

	Accuracy	**Precision**	**Recall**
CNN on color images	84%	82%	78%
CNN on grayscale	75%	69%	67%
ResNet50 on color images	87%	84%	83%
ResNet50 on grayscale	79	78%	66%

Table 6.1 shows that after converting the train, validation set and test set to grayscale, the performance reduced for both the CNN models trained from scratch and using ResNet50. This result was in contrary to the expectation that color would be a significant feature for correct and precise classification of the breast cancer images.

6.3. Opportunities and challenges of deep learning in Invasive Ductal Carcinoma Breast Cancer

This research study has also demonstrated the opportunities that deep learning is providing in classification Invasive Ductal Carcinoma Breast Cancer. With computer vision one of the opportunity is the application of deep learning in image processing. After the image processing, the images were classified as cancerous or non-cancerous using deep learning.

Deep learning is less time consuming compared to using handcrafted methods. Using models with CNN architecture one is able to classify a number of images is short period compared to using handcrafted methods which requires human intervention and this is a challenge in most developing countries. In line with that, deep learning reduces human error and bias in identification of cancer patients.

Although there are a number of opportunities and advantages of using deep learning in classification of IDC breast cancer images, there are also some challenges that the field is facing and that has been identified in this research.

To begin with, deep learning requires large dataset to perform well. In most developing countries, finding large and quality dataset is a challenge as a result building deep neural network model is not easy. Using deep learning in classification of IDC breast cancer images is computationally expensive. This is because deep learning models like CNN have millions of parameters to be trained.

Chapter 7: Conclusion

The prediction or classification of Invasive Ductal Carcinoma Breast Cancer images using Convolution Neural Network architecture was fascinating as well as challenging since the subject area of deep learning is not straight forward and appears to be complicated to be implemented. However, the subject area is still interesting as it incorporates new aspects of data analysis than the traditional analysis. Where possible the neural network and hyper-parameters used in this thesis are similar. The models accomplished good results, which were presented in chapter 5.4 and analyzed in section 6.1.

The three main research questions that have been responded in this thesis include "what are the opportunities and challenges that comes with using deep learning in IDC breast cancer detection?", "what method would be recommended to use to classify the images using the data set that was used in this thesis?" and "Does color affect classification of medical images?"

Research question of understanding opportunities and challenges that arises with using deep learning in IDC breast cancer prediction has been answered in section 6.3. Some of the opportunities pointed out in this research include reducing human intervention in classifying medical images. This results in eliminating human error and bias that may rise using handcrafts methods. However, there are also challenges that this thesis has demonstrated in using deep learning to classify IDC breast cancer images and other medical images. Some of the challenges highlighted include the need for large dataset which is not easy and deep learning models like CNN are computationally expensive.

The answers to questions of which method is recommended to use to classify the images is in section 6.1 and 6.2. Section 6.1 and 6.2 summarizes the results from the main evaluation methods that have been used to analyze the performance of the models. Using accuracy and precision, ResNet50 performs well with highest scores in all the three matrices compared to CNN from scratch and MobileNet. ResNet50 obtained an overall accuracy of 87% and 84% and 85% for CNN from scratch and MobileNet respectively. The highest precision was also obtained by ResNet50 with a score of 84% while CNN from scratch and MobileNet achieved 82% and 83% respectively. Moreover, analyzing false negative rate, ResNet50 performed better because of its lowest score of

27% while CNN from scratch and MobileNet achieved 37% and 36% respectively. Therefore, using this analysis, ResNet50 can be recommend to be used in this dataset.

The results from grayscale models where the results have been presented in figure 5.16, 5.17, 5.18 and 5.19 and in tables answered the research question of finding out if color has significance in classifying IDC images. The models trained on grayscale images to classify the IDC images were evaluated and showed color is not significant in cancer images specifically this dataset that was used in this thesis.

Chapter 8: Future Works

In this thesis, a Convolutional Neural Network and Transfer Learning were implemented to classify Invasive Ductal Carcinoma Breast Cancer images. ResNet50 performed well compared to MobileNet and CNN model from scratch. Future work in this area can be using multiclass classification. The IDC breast cancer images can have different stages or class of cancer as such using multiclass classification would be interesting to look at and would tackle the problem of misclassification of the models. Furthermore, exploring the effects of using deeper architectures such as more layers and more neurons in CNN models would be fascinating to look on. Using larger cohorts on validation can also be an interesting work in the future. Deeper architectures including hyper-parameter tuning and increasing the size of the dataset can improve the classification performance.

References

Carstensen, A., & Bernhard, J. (2016). *Design Science Research as an Approach for Engineering Education Research.* Retrieved from https://www.researchgate.net/publication/303940785_Design_Science_Research_as_an_Approach_for_Engineering_Education_Research

Convolutional neural network. (2021, May 15). Retrieved from https://en.wikipedia.org/wiki/Convolutional_neural_network

Cross-validation (statistics). (2020, May 11). Retrieved from https://en.wikipedia.org/wiki/Crossvalidation_(statistics)

(2021, 6 4). Retrieved from https://www.scikityb.org/en/latest/api/classifier/classification_report.html

Aggarwal, C. (2018). Convolutional Neural Networks. In: Neural Networks and Deep Learning. *Springer, Cham*.

Ali, N. O. (2020). *A Comparative study of cancer detection models using deep learning.*

Angel, C.-R., Ajay, B., Fabio, G., Hannah, G., Michael, F., Shridar, G., . . . Anant, M. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks. *Medical Imaging*, 2-4.

Bell, J. (2014). *Machine learning: Hands-on for developers and technical professionals.*

Canuma, P. (2018, October 10). *Towards Data Science*. Retrieved from https://towardsdatascience.com: https://towardsdatascience.com/image-preprocessingc1aec0be3edf

Culfaz, F. (2018, November 6). *Transfer Learning using Mobilenet and Keras*. Retrieved from Towards Data Science: https://towardsdatascience.com/transfer-learning-usingmobilenetand-keras-c75daf7ff299

DataVedas. (n.d.). Retrieved from HOLDOUT CROSS-VALIDATION: https://www.datavedas.com/holdout-cross-validation/

Donges, N. (2020, Sepetember 3). *BuiltIn*. Retrieved from https://builtin.com/datascience/transferlearning

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

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.

Howard, A. G., Zhu, M., & Chen, B. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications.

Jaiwei, H., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* Waltham: Morgan Kaufmann.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., . . . Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology. 230-243.

Lakshmanamoorthy, R. (2021, May 5). Getting Started With Deep Learning Using TensorFlow Keras.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 436–444.

Millie. (2021). Data Mining. *UR*, 20-21.

Musa, A., & Aliyu, U. M. (2016). *Application of Machine Learning Techniques in Predicton of Breast Cancer Metastases Using Decision Tree Algorithm.*

Nahid, A.-A., & Kong, Y. (2017). *Involvement of Machine Learning for Breast Cancer.* Australia.

Quang, D., & Xie, X. (2016). A hybrid convolutional and recurrent deep neural network. *Nucleic Acids Research*.

Rosenblatt, F. (2018). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 386-408.

Shalev-Shwartz, Shai, S., & Ben-David. (2015). *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press.

Skandalakis, J. (2009). *Embryology and Anatomy of the Breast.* Berlin: Springer Berlin Heidelberg.

Vaishanavi, K., & Kuechler Jr., W. (2008). *Design Science Research Methods and Patterns.* Boca Raton: CRC Press, Taylor & Francis Group.

Visa, S., Atsushi, I., & Anca, R. (2011). Proceedings of the Twentysecond Midwest Artificial Intelligence and Cognitive Science Conference. *University of Cincinnati* (pp. 25-30). Ohio: Omnipress – Madison, WISCONSIN.

Xie, Y., & Richmond, D. (2018). Pre-training on Grayscale ImageNet Improves Medical Image Classification.

Plagiarism report from turnitin

Sunge

ORIGINALITY REPORT

14%	10%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ls00012.mah.se Internet Source	3%
2	Mohammad Abdolahi, Mohammad Salehi, Iman Shokatian, Reza Reiazi. "Artificial intelligence in automatic classification of invasive ductal carcinoma breast cancer in digital pathology images", Medical Journal of The Islamic Republic of Iran, 2020 Publication	1%
3	www.hindawi.com Internet Source	1%
4	Submitted to University of Thessaly Student Paper	<1%
5	downloads.hindawi.com Internet Source	<1%
6	www.irjet.net Internet Source	<1%
7	"Proceedings of the International Conference on Artificial Intelligence and Computer Vision	<1%

(AICV2020)", Springer Science and Business Media LLC, 2020
Publication

8	Submitted to Liverpool John Moores University Student Paper	<1%
9	towardsdatascience.com Internet Source	<1%
10	Sharmistha Nayak, Nurzaman Ahmed, Sudip Misra. "Deep-learning-based reliable routing attack detection mechanism for industrial Internet of Things", Ad Hoc Networks, 2021 Publication	<1%
11	library.iugaza.edu.ps Internet Source	<1%
12	pdfs.semanticscholar.org Internet Source	<1%
13	www.ijert.org Internet Source	<1%
14	Submitted to Monash University Student Paper	<1%
15	Submitted to B.V. B College of Engineering and Technology, Hubli Student Paper	<1%
16	link.springer.com Internet Source	<1%

17	theses.gla.ac.uk Internet Source	<1 %
18	Submitted to Mae Fah Luang University Student Paper	<1 %
19	"Neural Information Processing", Springer Science and Business Media LLC, 2017 Publication	<1 %
20	"Biometric Recognition", Springer Science and Business Media LLC, 2017 Publication	<1 %
21	Submitted to University of Surrey Student Paper	<1 %
22	silo.pub Internet Source	<1 %
23	www.longdom.org Internet Source	<1 %
24	Submitted to University of Warwick Student Paper	<1 %
25	hdl.handle.net Internet Source	<1 %
26	Submitted to CITY College, Affiliated Institute of the University of Sheffield Student Paper	<1 %
27	repository.tudelft.nl Internet Source	<1 %

28	"Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012", Springer Science and Business Media LLC, 2012 Publication	<1%
29	Submitted to CSU, Long Beach Student Paper	<1%
30	"Soft Computing and Signal Processing", Springer Science and Business Media LLC, 2021 Publication	<1%
31	Gurcan, Metin N., Anant Madabhushi, Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, and Natalie Shih. "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks", Medical Imaging 2014 Digital Pathology, 2014. Publication	<1%
32	Jennifer Batamuliza, Damien Hanyurwimfura. "Identity based encryption with equality test", Information Security Journal: A Global Perspective, 2020 Publication	<1%
33	Ojas A. Ramwala, Smeet A. Dhakecha, Antriksh Ganjoo, Divyanshu Visiya, Jignesh N.	<1%

34	www.ncbi.nlm.nih.gov Internet Source	<1%
35	"Proceedings of the Future Technologies Conference (FTC) 2020, Volume 3", Springer Science and Business Media LLC, 2021 Publication	<1%
36	Yusuf Celik, Muhammed Talo, Ozal Yildirim, Murat Karabatak, U Rajendra Acharya. "Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images", Pattern Recognition Letters, 2020 Publication	<1%
37	lup.lub.lu.se Internet Source	<1%
38	spectrum.library.concordia.ca Internet Source	<1%
39	www.mdpi.com Internet Source	<1%
40	Mu, Yang, Wei Ding, Dacheng Tao, and T.F. Stepinski. "Biologically inspired model for	<1%

41	Yibin Wang, Haifeng Wang, Zhaohua Peng. "Rice diseases detection and classification using attention based neural network and bayesian optimization", Expert Systems with Applications, 2021 Publication	<1%
42	dokumen.pub Internet Source	<1%
43	erepo.uef.fi Internet Source	<1%
44	escholarship.org Internet Source	<1%
45	infonomics-society.org Internet Source	<1%
46	kclpure.kcl.ac.uk Internet Source	<1%
47	"Computer Vision – ECCV 2020", Springer Science and Business Media LLC, 2020 Publication	<1%
48	"Intelligent Computing and Optimization", Springer Science and Business Media LLC, 2020 Publication	<1%

49 "World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China", Springer Science and Business Media LLC, 2013 <1%
Publication

50 Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, Ma Yunsheng, Songqing Chen, Peng Hou. "A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure", IEEE Transactions on Services Computing, 2018 <1%
Publication

51 Lecture Notes in Computer Science, 2016. <1%
Publication

52 Luís Carlos Silva, Maricéia B. S. Pádua, Leonardo M. Ogusuku, Marcelo Keese Albertini, Renato Pimentel, André R. Backes. "Wild boar recognition using convolutional neural networks", Concurrency and Computation: Practice and Experience, 2020 <1%
Publication

53 Pakanat Decharatanachart, Roongruedee Chaiteerakij, Thodsawit Tiyarattanachai, Sombat Treeprasertsuk. "Application of Artificial Intelligence in Chronic Liver Diseases: A Systematic Review and Meta-analysis", Research Square, 2020 <1%

54	S. Don, Duckwon Chung, K. Revathy, Eunmi Choi, Dugki Min. "A New Approach for Mammogram Image Classification Using Fractal Properties", Cybernetics and Information Technologies, 2012 Publication	<1%
55	Submitted to University of Birmingham Student Paper	<1%
56	Submitted to University of Sheffield Student Paper	<1%
57	dspace.cc.tut.fi Internet Source	<1%
58	scitepress.org Internet Source	<1%
59	uknowledge.uky.edu Internet Source	<1%
60	www.ijitee.org Internet Source	<1%
61	Alessandra Lumini, Loris Nanni, Gianluca Maguolo. "Deep learning for plankton and coral classification", Applied Computing and Informatics, 2019 Publication	<1%
62	Joseph R. England, Jordan S. Gross, Eric A. White, Dakshesh B. Patel, Jasmin T. England,	<1%
63	"Hybrid Intelligent Systems", Springer Science and Business Media LLC, 2021 Publication	<1%