UNIVERSITY of RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY

Regional Centre of Excellence in Biomedical Engineering and e-Health (CEBE)

Topic: An Artificial Intelligence Model for Mammogram Lesions Detection

By:

Michel UTAZIRUBANDA
Reference Number: 220020526

A Dissertation Submitted to the Regional Centre of Excellence in Biomedical Engineering and e-Health (CEBE), University of Rwanda as partial fulfilment of the requirements for the Master's Degree in Biomedical Engineering.

Supervised by:

Dr. Philibert NSENGIYUMVA
and
Prof Yunjie TONG

DECLARATION

I, Michel UTAZIRUBANDA, declare that this dissertation entitled "An Artificial Intelligence Model for Mammogram Lesion Detection" is my original work based on research and a prototype and has not been submitted for any other degree or professional qualification.

Student Name: Michel UTAZIRUBANDA

Reference Number: 220020526

Student Signature: 

Date: November 21, 2023

CERTIFICATE

This is to certify that the project entitled "An Artificial Intelligence Model for Mammogram Lesions Detection" is a record of original work done by Michel UTAZIRUBANDA (Reference number: 220020526), an MSc. Degree student in Biomedical Engineering. This work has been submitted under the guidance of Dr Philibert NSENGIYUMVA and Prof Yunjie TONG.

Main Supervisor:

Dr. Philibert NSENGIYUMVA

Co-Supervisor:

Prof. Yunjie TONG

Biomedical Engineering Master's Program Coordinator

Dr. Gerard RUSHINGABIGWI

ACKNOWLEDGEMENTS

First and foremost, I want to express my gratitude to the Almighty God for carrying me and keeping me healthy and protected throughout my research period.

I am deeply grateful to Prof. Celestin TWIZERE and Dr. Gerard RUSHINGABIGWI for their continuous guidance, direction, and coordination in supporting my academic advancement. While all my lecturers have been valuable to me, my supervision team, consisting of Dr Philibert NSENGIYUMVA and Prof. Yunjie TONG, has truly stood out for their exceptional effort, insight, enthusiasm, passion, and willingness to engage.

I express my gratitude to the King Faisal Hospital review board for conducting an ethical review and providing permission to collect data from their organization. Additionally, I extend my appreciation to Clement MANZI and Patrick NZABAMWITA for coaching me on how to use the ANACONDA navigator and Jupiter notebook environment.

I am extremely thankful to my classmates, friends, and relatives for their kind comments during the review of this dissertation. I would also like to express my appreciation to CEBE for its financial support and the invaluable skills it has imparted to me throughout my postgraduate program.

ABSTRACT

Introduction: Breast cancer is a type of cancer that affects the fibroglandular cells in the breast. According to a 2018 estimate, the number of new cases of breast cancer worldwide is expected to rise to 29.5 million by 2040. In Rwanda, breast cancer is the most common cancer among women, and delays in diagnosis and treatment are believed to contribute to the high number of cases. Mammography screening is typically performed on women of a certain age to detect early signs of cancer. However, this method places a heavy burden on radiologists who are already scarce in number. Moreover, each person's ability to detect lesions in mammography images varies, and if a radiologist fails to detect the cancerous growth, additional exams may prove to be expensive in terms of both finances and lives.

Aim. This study aimed to develop a deep-learning AI program that classifies medical images as normal or abnormal, aiding radiologists in their diagnoses.

Methodology. Our project utilized Python's built-in methods and packages to create, train, and evaluate a deep-learning model. The model was trained on 1200 images from King Faisal Hospital, all of which were captured in 2022. The Residual Neural Network 50 was the focal point of our approach.

Results. Our results showed a sensitivity probability of 1.00, an accuracy of 0.78, a precision of 0.69, and an F1 score of 0.52.

Conclusion. The model is currently performing well. However, it is highly recommended to continue the project by comparing it to other deep learning models or increasing the dataset size to achieve optimal performance. This is especially important in a medical setting, where even a small error can result in the loss of life. Therefore, it's crucial to ensure that the model is thoroughly tested and validated before deploying it.

Keywords: Mammogram, Breast Cancer, Artificial Intelligence; Deep learning; Neural Network Classifier

LIST OF ACRONYMS

AI: Artificial intelligence

ANN: artificial neural network

BCS-DBT: Breast Conserving Surgery-Digital Breast Tomosynthesis

BI-RADS: Breast Imaging Reporting and Data System

CEBE: Regional Centre of Excellence in Biomedical Engineering and e-health

CNN: Convolutional neural network

DICOM: Digital Imaging and Communications in Medicine

DL: Deep learning

IRB: Institutional Review Board

KHF: King Faisal Hospital

KNN: k-nearest neighbor

ML: Machine Learning

MoH: Ministry of Health

PACS: Picture Archiving and Communication System

SVM: Support vector machine

UR: University of Rwanda

YOLOv4: You Only Look Once Version 4

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

This chapter mainly focuses on the research problem and goals. It also highlights important research questions that need to be addressed to achieve those goals. The chapter also includes the overall structure of the dissertation and the expected long-term benefits of the study.

1.2 Problem Statement

Cancer is one of the most significant non-communicable diseases today, and it is causing a considerable increase in global death rates [1]. In 2018, around 18.1 million new cases of cancer and 9.5 million related deaths were reported worldwide. If cancer control does not improve, it is projected that the annual incidence might climb to 29.5 million new cases by 2040 [2]. In Rwanda, there were 10,704 new cancer cases and 7,662 cancer-related deaths in 2018 [2]. The top five cancers in males are prostate, liver, stomach, colorectal, and Kaposi sarcoma, while cervix, breast, colorectal, stomach, and liver are the most common in females [2]. Studies indicate that breast cancer is the most frequently detected cancer among Rwandan women, and delays in diagnosis and treatment are thought to be responsible for a high percentage of breast cancer-related deaths [2].

The poor contrast and extensive similarities between lesions and breast masses make mammograms difficult to interpret, even for qualified radiologists [3]. Therefore, there is a high demand for more precise interpretation methods.

In Rwanda, there is a growing number of mammography images produced daily, but there is a limited number of qualified radiologists to interpret them. This leads to lower accuracy and performance, resulting in the misclassification of breast lesions and delayed diagnoses. These delays in decision-making can result in the development of cancer and lower survival rates. To address this problem, studies are currently exploring the use of machine learning algorithms. However, none have been implemented in Rwanda yet. Therefore, there is a need to develop a better and faster algorithm that can detect breast lesions more accurately using the case studies of Rwanda's healthcare institutions.

1.2 Research Questions

To guide this study, we will search for answers to three basic questions:

a) How effectively does the model determine the presence or absence of breast cancer?

b) How much time does the model require to process the input data?

c) Does this model improve decision-making in breast cancer imaging?

1.4 Objectives

1.4.1 General Objective

The long-term goal of the research is to develop an AI-based algorithm to assist radiologists in analyzing screening mammograms.

1.4.2 Specific Objectives

The following goals have been established as benchmarks to ensure that the project achieves its overall objective:

a) Develop an innovative CNN model that can accurately classify mammograms based on the presence of lesions.

b) Verify the model's performance through rigorous testing.

c) Evaluate the model's effectiveness and underlying principles in terms of sensitivity, specificity, and accuracy.

1.5. Study Scope

This study aimed to classify digital full mammography images as normal or abnormal, without regard to the stage or type of cancer. The study is focused on binary classification and includes both normal images and images showing suspicious masses related to breast cancer from national healthcare institutions.

1.6 Significance of Study

The implementation of the study is expected to open up new avenues for the use of AI in radiography. Here are the possible benefits:

a) Lowering the risk of breast cancer fatality.

b) Reducing the bias in subjective image analysis and interpretation.

c) Reducing the duration of treatment and the financial expenses for the patients.

d) Standardizing the skills of radiologists across the country.

1.7 Organization

In chapter one of the research study, the study is introduced. The following chapters are arranged in the following order: chapter two identifies gaps and presents a review of relevant literature. Chapter three describes the research approach, including the research design, research tool, and ethical considerations. Additionally, it explains the model design, core principles, and relevant

aspects. Chapter four is dedicated to discussing the results and analysis of the research findings. Finally, chapter five concludes and highlights the study's recommendations.

1.8 Summary

In this chapter, we introduce the problem description and highlight the need for improving the current model to achieve higher levels of performance and accuracy. The benefits of this research are detailed in the following chapter, which provides a comprehensive overview of our study.

CHAPTER 2. STATE OF THE ART

2.1 Introduction

This chapter presents a comprehensive review of the existing studies on breast cancer and the application of deep learning for cancer detection. The literature review is primarily based on a selection of carefully chosen published papers, textbooks, articles, reports, and other relevant sources.

2.2 Literature Review

Breast cancer is the most common malignant tumour in women and the leading cause of cancer death in women around the world, however, if discovered early, breast cancer patients can boost their 5-year survival rate from 25% to 99%[3]. That is early detection, leads to early treatment and avoids metastasis, which can help reduce cancer mortality and morbidity[2].

Medical imaging plays a crucial role in cancer treatment. Breast cancer is the most frequent malignancy among women in Sub-Saharan Africa, and access to surgeries, radiotherapy, and other treatments is limited based on resource constraints [4]. The modern way of life is becoming more important in the development of breast cancer and breast image classification techniques by using AI have been used significantly around today's world [5].

Breast cancer is the most common cancer in Rwanda, accounting for 14 per cent of new cases and 10.5 per cent of fatalities[6]. Almost 10,704 new cancer diagnoses were confirmed in 2018, with 6,184 cases in women and 4,520 cases in men. Those cancers include cervical, prostate, stomach, liver, non-Hodgkin lymphoma, lung, leukaemia, ovary, rectum, oesophagus, central nervous system, colon, Kaposi sarcoma, kidney, penis, lip, larynx, Hodgkin lymphoma, Thyroid, Vulva, Multiple myeloma, bladder, pancreas, skin Melanoma, corpus uteri, oropharynx, nasopharynx, salivary glands, anus, testis, vagina, gallbladder, hypopharynx, and mesothelioma [6].

Rwandan health care services are into four main sectors primary, secondary, and tertiary and referral hospitals where secondary comprises district hospitals that rely on radiography and ultrasound modalities even if there is a limited number of specialized radiologists [7]. Medical imaging has progressed over time and today plays an essential role in cancer screening, early detection, diagnosis, staging, and follow-up, providing cancer management continuity[8]. Pathological diagnosis and medical imaging technologies in Rwanda are still primitive and need to be developed to achieve quality standards for cancer diagnosis and staging[2].

Anatomically, the breast is made up of glands, ducts, connective tissue, and fat. Upon being activated by the central nervous system, sets of nodes called lobules produce milk, which is then transported to the breast's nipple through ducts. Additionally, those supporting components like

fat and ligaments maintain the breast's form, size, and shape. Likewise, the centre of the areola's nipple, which is made up of fibre muscles, serves as a location for releasing milk. Also, the areola glands that emit an oily material that serves as a lubricate. Figure 2.1 shows the side view cross-section of the normal image[9]

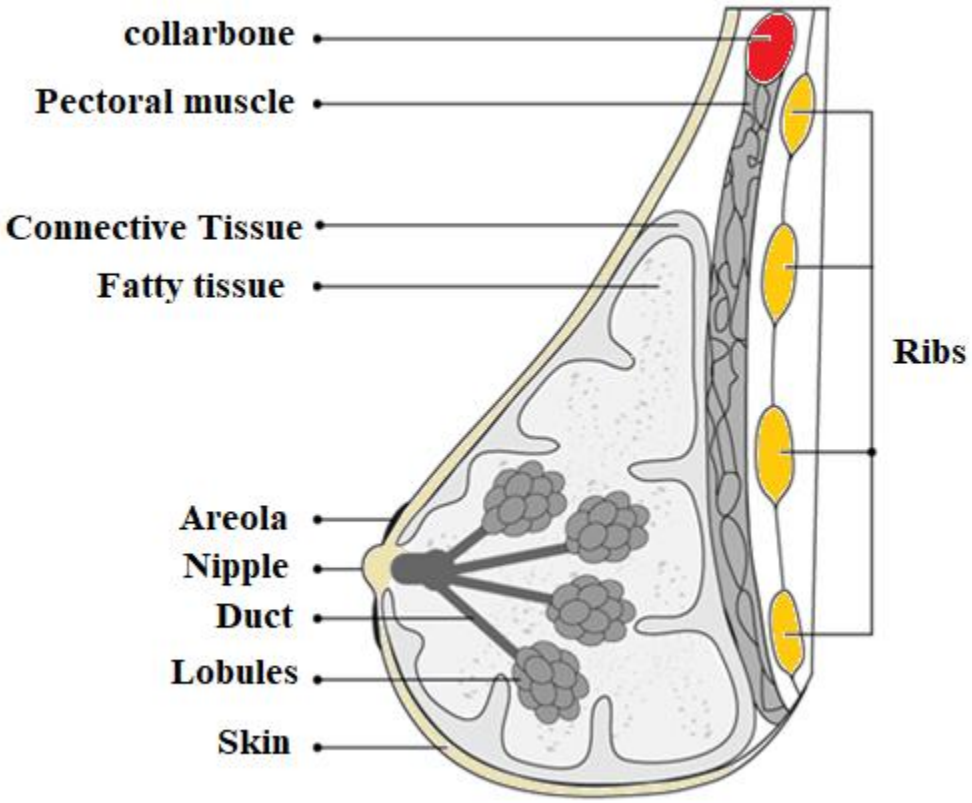

Table 2. 1. Side view cross-section of the normal breast.

When there is uncontrolled growth of cells within the breast then the cancer is developed and once it is not diagnosed and treated well it will lead to catastrophic problems including death. The common symptoms of abnormal breasts include breast thickening or lumping in one area, changes in size and appearance, and redness of the skin above the breasts.[9] . Although there are other ways to classify breast cancer, the current study looked at it in terms of growth rate. When a cancer is non-invasive toward neighbouring organs or does not spread, we do classify it as benign; yet, when cancer spreads through metastasis, we do classify it as malignant.[10]

Radiologists use the BI-RADS standard system to describe mammogram readings and findings. This system categorizes data into four groups labelled 0 through 6 depending on the proportion of fibrous and glandular tissue to fatty tissue in the breasts. The results need to be confirmed by additional tests because the denser the breast, the harder it is to detect lesions in that breast [11]. Table 2.1 displays how radiologists evaluate images and offer comments on categories. [12]

Category	Decision	Comments
0	Mammography: Incomplete	Need additional evaluation and or prior mammograms for comparison
	Ultrasound & MRI: Incomplete	Need additional imaging evaluation
1	Normal	-
2	Benign	-
3	Probably benign	-
4	Suspicious	4A: Low suspicion for malignancy
		4B: Moderate suspicion of malignancy
		4C: High suspicion for malignancy
5	Highly suggestive of malignancy	-
6	Known biopsy-proven malignancy	-

Table 2. 2. BI-RADS Assessment Categories

One method for finding lesions and breast mass density is imaging testing. There are different imaging technologies such as X-rays, MRI, ultrasound, and positron emission tomography but X-ray mammography has become popular as it provides high-quality images at low radiation doses and low cost.[13]. While MRI is advised for people without ferrous metal implants or those with a family history of breast cancer, ultrasound is useful in determining whether solid lumps visible on a mammogram are not abnormal[14]. The fact that each modality has its pros and cons, there doesn't seem to be a single imaging technique that can now identify and diagnose breast cancer with high sensitivity and high specificity.

All of the above modalities require the patients to return for additional testing, but recently digital breast tomosynthesis has been introduced to cut down on those visits. However, there is still a problem with subjecting patients to excessive amounts of X-rays even if they are in the safety range, as well as inconsistent reconstruction algorithms and variation of images due to changes in X-ray[14]. There are computer-aided systems in use today, but they show a large percentage of false-positive results and hence many mammography radiologists avoid employing them due to the reason above not leading to an improvement in diagnostic accuracy[15].

The opportunity to use deep-learning-based medical image analysis to improve the accuracy and efficiency of so many diagnostic and treatment processes has sparked fresh research and development initiatives in computer-aided screening Software such as Python has been used as a

tool for machine learning but the networks need a clear understanding of network algorithm for the cycle of improving the performance[16]. There is also the chance of classifying the mammogram types based on the distance to training samples, boundaries, back propagation neural networks, support vector machines, clustering, and multiple instance learning however some lack the system of extracting features and hence continue to overload the radiologists and they prefer not to deploy them [17]. An example of radiomic algorithms was used to detect lesions, but the constraint of utilizing conventional or deep learning algorithms was the identification of the biological underpinnings of the most informative radiomic properties [18].

SVM, ANN, KNN, binary decision trees, and simple logistic classifiers are some of the most often used classifiers in breast cancer categorization [19]. ANN models and LightGBM classifiers are more likely to detect anomalies in the form of bulk or calcification than architectural asymmetry or distortion however they don't employ deep learning [20], [21].

Convolutional nets, recurrent nets, or recursive neural tensor networks are great options for object recognition when using deep learning, whereas deep belief networks or convolutional nets perform best when used for image recognition[21], [22]. The most frequently used metrics for a standardized clinical test are accuracy, sensitivity and specificity, with accuracy being understood as the ability to classify the mammograms correctly and specificity being defined as the proportion of true negative results while sensitivity is the percentage of true positive outcomes [23].

The use of AI in different domains showed good prediction and it was a key for removing some errors during the decision-making. Researchers are now interested in this issue and a key example is the use of CNN with thermal imaging which resulted in a medium accuracy of 0.84 while detecting surgical site infections.[25]. Table 2.2 summarizes the key previous studies related to CNN deep learning classifiers.

Work	Year	Technique	Observations
[26]	2020	CNN	Fruit recognition in agro-processing and quality control. The same cases showed a performance of 100%.
[27]	2019	1. CNN-Based Transfer Learning 2. CNN-ORB Oriented Fast and Rotated Binary 3. CNN-SVM Support Vector Methods	Pneumonia classification by using a chest X-ray small dataset. To improve the performance, one can apply the data augmentation but because the augmentation transforms the image too complicated augmentation adds noise.
[28]	2022	CNN-SVM-Random Forest	Use of mammogram images from Kaggle. In terms of accuracy, CNN showed high performance at the rate of 99.67:89.84:90 per cent CNN by SVM respectively.
[29]	2020	InceptionResNet-V2,feedforward CNN, ResNet-50 and YOLO detector	By using DDSM and INbreast datasets the accuracy was 95.32%.
[30]	2023	CNN- Google Net- VGG11 - MobileNetV3_Small.	Online histopathological breast cancer datasets (BreakHis and ICIAR-2018). There was a better accuracy for the BreakHis dataset than the ICIAR-2018 data set. That is 98.24%, 98.67%, and 96.16% versus 96.95%.
[31]	2022	CNN- YOLOv4 Vs Nested Contours	Onsite breast cancer images. Yolov4 CNN model did not influence the radiologist's decision therefore Nested counters algorithms are still superior to this model.
[32]	2020	CNN- Texture Feature Extraction	Used to detect melanoma skin cancer. This study showed that there is a need for more models and more datasets because some reached 96.3 % while others gave 52% sensitivity.
[33]	2022	CNN	This comparison study showed that ResNet deep learning can be used to perform deep feature extraction and multilayer perceptron (MLP) can be used to classify breast cancer and the performance is high.
[34]	2018	Full Resolution Convolutional Network (FrCN), CNN, YOLO	Through the use of the Inbreast dataset, the accuracy became 92.97%

Work	Year	Technique	Observations
[35]	2021	CNN	By using In breast cancer, the system resulted in a good accuracy of 80.30%
[36]	2019	CNN	Through training specific sample size of the images retrieved from the DDSM dataset and by using the area under the curve (AUC=92%) as performance measures it has been detected that the larger the sample size the more advantageous the model is.
[37]	2020	CNN_Deep Learning assisted Efficient Adaboost Algorithm	The suggested CNN-based strategy used the cancer Imaging Archive dataset to demonstrate good accuracy of 97.2%, sensitivity of 98.3%, and specificity of 96.5% when compared to other ML techniques.
[38]-[39]	2022	CNN_ ResNet50, VGG-16, MobileNetV2.	This comparative study showed that deep learning led to good accuracy but the Resnet 50 Model was perfect (100.0%) compared to VGG-16 and Mobile NetV2 97.22%, and 98.61% respectively.

Table 2. 3. CNN- Related state-of-the-art works

2.3 Research Gap

Even if some deep learning models have been successfully predicting breast cancer, the majority are using the online dataset with well-categorized images from a group of expert radiologists. We did not find any CNN model deploying or trying to train the mammograms from the national clinics and hospitals. In addition, none of the models is implemented in clinical systems to evaluate its performance. Therefore, more contribution as the researcher is a necessity for further comparison

2.4 Summary

Categorization of breast imaging with AI will lead to early diagnosis of lesions. Deep learning, as a subset of ML and AI, is important in image categorization since it reduces the cost of computing extremely large parameters. We are going to base on the above literature and make a new model as you find it in the next chapter.

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Introduction

This chapter explains the research methodologies and tools used in this study. The stages taken from the start of the research to the present are highlighted here. You are going to find the research procedure as well as the developed model's fundamental operating principles in this chapter.

3.2 Research Process and Tools

The image binary classification has been used to determine if images are abnormal or not. The literature review and familiarization with the Python tools are the primary research approaches used in this work. To train the model, 1000 images were used, while 200 images helped in testing. Both sets are collected from King Faisal Hospital where half of each set is nora mal full mammogram and the other the abnormal images. There were 600 normal images and 600 abnormal images used in this study, with abnormal images falling into two categories: malignant and benign. Only images in the form of Joint Photographic Experts Group data have been downloaded from PACs (JPEG).

To complete the work, the Python Jupyter Notebook has been used to develop description, CNN feature extractor, learning, validation, testing, and classifier training. Before feature extraction and classification, images have been enhanced and preprocessed with mathematical functions to shorten the processing time. Figure 3.1 shows all processes involved and the related activities.

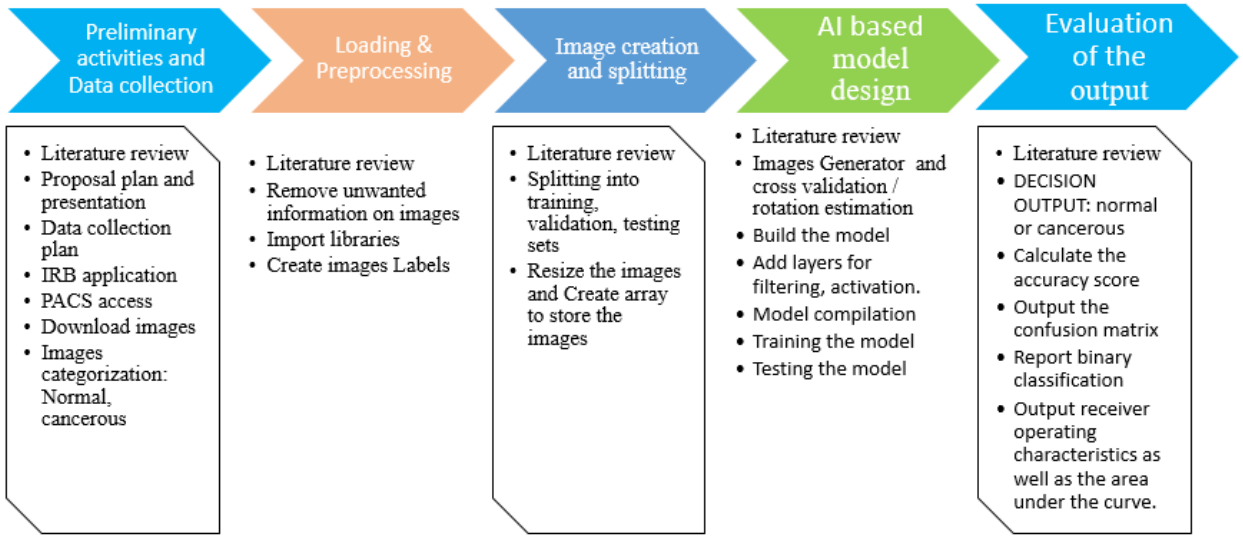

Figure 3. 1. Research Process and corresponding main activities

3.3 Preliminary Activities and Data Collection

3.3.1 Preliminary activities

This period consisted of reading different literature reviews as summarized in chapter two. It was also the time for the researcher to select the research topic, make the research proposal and improve on it based on the social academic comments. There was also the time to visit some hospitals and observe challenges in medical service delivery. The visited hospitals for observation are Rwanda Military Hospital, King Faisal Hospital, Kigali University Teaching Hospital and Butare University Teaching Hospital. Moreover, the time here went for learning the Python tools for deep learning applications.

3.3.2 Data Collection

a. Data requirements

The secondary type of data collection has been used as there was not any physical interaction with the patient but his or her medical records. The data type was full images taken by digital mammography. The images might fall into one of two categories normal images free from lesions and abnormal breast images. Thereafter the sampled data were used for both training and testing as presented in Figure 3.2.

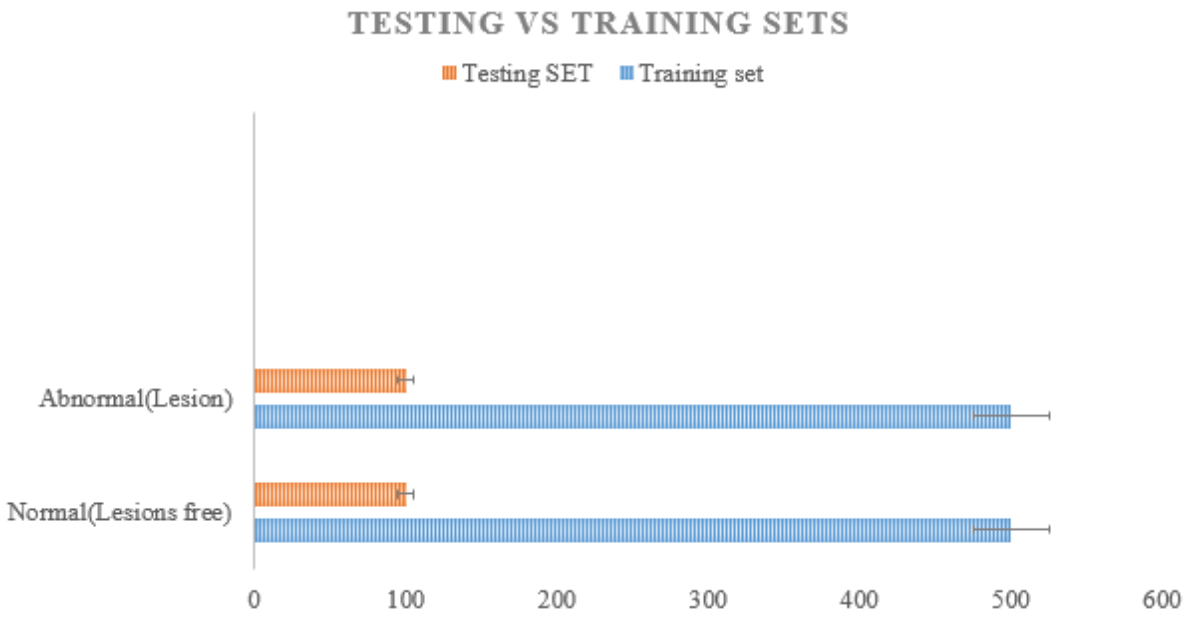

Figure 3. 2. Sampled data

b. Data collection

The data have been downloaded from KFH PACS. On the last, there is one folder with three subfolders one for malignant images, another for benign images and the third for normal images. The next step was to combine benign images with malignant images to form the abnormal set

because malignant images were not enough for the current study and Figure 3.3 displays some random input images to the model.

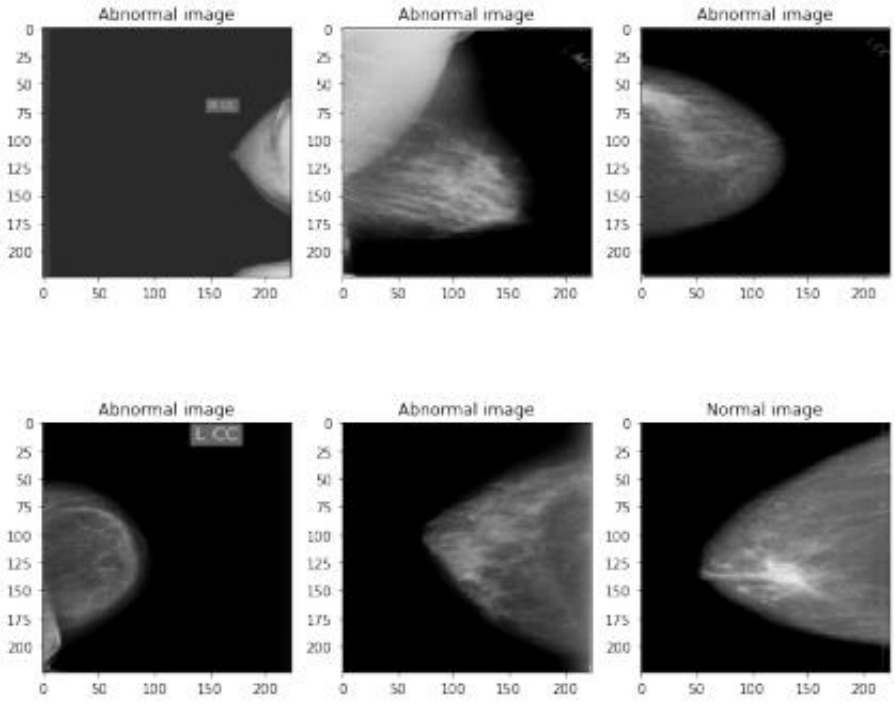

Figure 3. 3 Example of random Input images

c. Data Sampling and cleaning

For getting better accuracy the model needs huge images but for the cycle of this study, 1200 full breast images taken by digital mammography from KFH have been randomly sampled.

e. Data Modeling

Sampled and cleaned data have been modelled, visualized and interpreted by using the Anaconda Python-Jupyter Notebook in the computer with sixteen gigabytes of random access memory and HP ProBook 450 G7 model as the system requirements.

f. Inclusion and exclusion

The inclusion and exclusion criteria do not apply to this study data. Both confirmed breast cancer images and non-confirmed breast cancer were needed for this research regardless of age, sex and cancer stages as well as region.

g. Ethical considerations.

This study research proposal was approved by CEBE, the College of Medicine and Health Science (CMHS) as well as KFH institutional review boards. The data collection has been collected from KFH and the main ethical principle in the research was beneficence. As the research went to binary image classification and manipulation, there was zero harm to the patient and the outcomes might have been beneficial to future healthcare researchers.

h. Vulnerable populations

The study did not involve vulnerable populations.

e. Medical or psychosocial support

There has not been an expected likelihood of our study participants to require medical or psychological support.

f. Information and consent process

As the research objectives above stated, the information collected were both normal and benign or malignant confirmed breast cancer images. Therefore, there was not any process of filling out questionnaires as well as signing the informed consent.

g. Protection of privacy and confidentiality

The research activities have been performed under all ethical principles for the circle of respecting patient data and improvement of future patient health care services. By downloading the images from PACS, we saved only the images and the report has been consulted in the system to not disclose any kind of identification such as ID, age, or name. The downloaded images have been labelled with a new name.

h. Safekeeping of data

All data provided are kept in a password-protected laptop and metadata is in an encrypted document to which only the author can have access. In addition to the aforementioned above, the data will be kept and then destroyed after 10 years as per IRB requirements.

3.4 The AI-based Model Design

3.4.1 System Main Components

We are describing the system as two inputs branching into a block of classifier which on its round takes the decision on the images for being normal or abnormal. The first branch consists of a training dataset, from which, the classifier is trained and the features are extracted later. This classifier will fetch the features from a second testing dataset to evaluate its performance. The judgment to be normal or abnormal is dependent on previously learned features and raw features that were collected from the testing set, as shown in Figure 3.4.

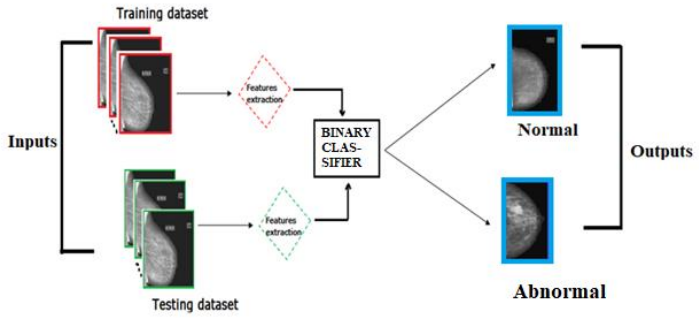

Figure 3. 4. System general parts and operating principle

The training dataset is the collection of mammograms from which the model was trained; these are the images that could have an impact on the model network, whereas the testing dataset is the collection of mammograms used to assess the model but not for model fitting it. The features

3.4.2 CNN Binary Classifier

The core concept of this system is convolution, in which images are multiplied by filters and subsampling is achieved through pooling. The current binary model was developed using an iterative approach in which feedback was gathered after both forward-moving and backward-moving steps as illustrated by Figure 3.5. The features are extracted and completely connected by rotation of the images and iterations, depending on the type of input datasets i.e. training or testing The classification loss function or binary cross entropy is used to illustrate how well or poorly the function is working before repeating the procedure and obtaining the outputs of labels. Choosing a batch size of sixteen and allowing the classifier to use forty epochs made sense.

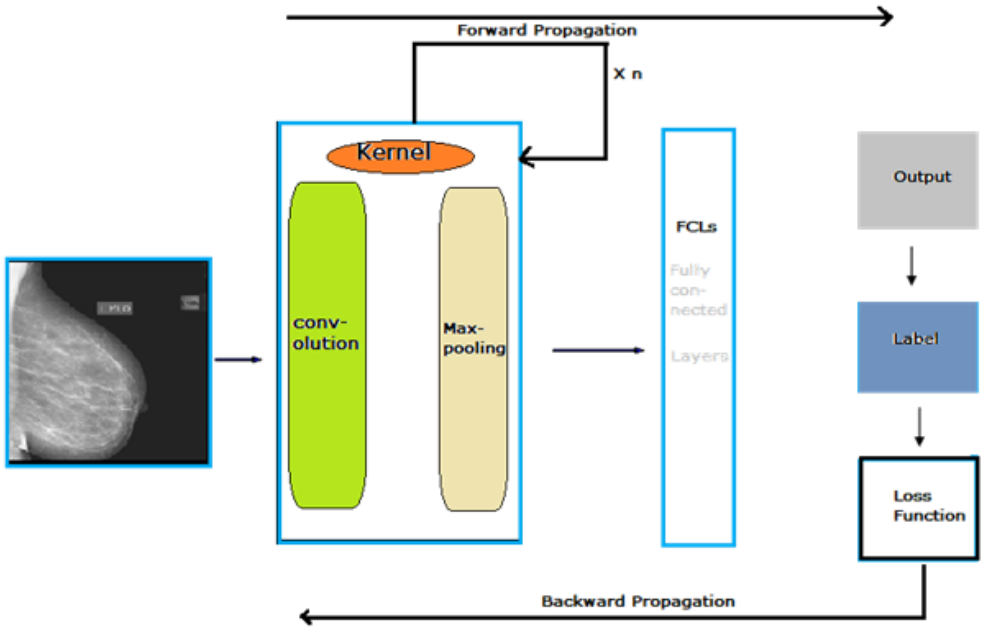

Figure 3. 5. Model architecture

3.4.3 Mathematical Algorithms

Convolution, pooling, and fully connected layers are the three types of layers utilized for feature extraction and classification. As an example, Figure 3. 6 explains how the fully connected layer can be achieved by using a maximum pooling structure every time a convolution filter is applied [40]. A Global average convolutional kernel was used to extract features, and max pooling

produced a down-filtered featured map. To activate the model SoftMax function is utilized to define the output of the neuron as it received a large number of inputs.

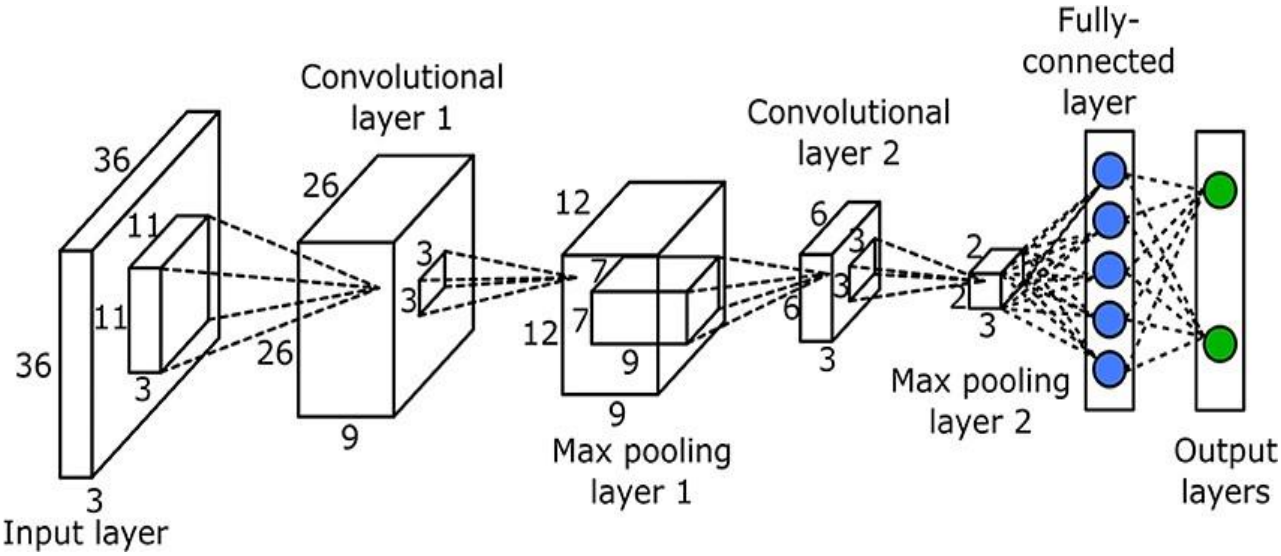

Figure 3. 6. CNN Example with Max Pooling

By applying ResNet50, we used forty-eight convolutional layers for training the dataset, one kernel layer and one maximum pooling layer. The validation of the overall model employed the softmax function which on its round converts the values from the array vector to values between 0 and 1. Those values can now be understood as probabilities. Its mathematical formula is as follows:

$$\sigma\left(\vec{Z}\right)_i = \frac{e^{Z_i}}{\sum_{j=1}^{K} e^{Z_j}} \qquad [41]$$

With $\vec{Z}$: input vector, e^{Z_i}: standard exponential function for input vector, K: number of classes in multi-class classifier and Z_j: standard exponential function for output vector.

This study used the Cross-Entropy loss function for compilation because it does not need any other activation function hence simplifying the learning rate through its mathematical expression here below:

$$CE = \sum_{i=2}^{2} \mathrm{t}_i \log(s_i) = -t_1 \log(s_1) - t_2 \log(s_2) \qquad [42]$$

Where t and s the ground truth and CE – cross-entropy

Using the Adam optimization strategy, the weights were modified during the epoch iterations. Adam, or the adaptive momentum estimation optimizer, was selected over the alternatives in this study because it combines the benefits of gradient descent extensions with the adaptable gradient algorithm[43]. Furthermore, it is a good choice for problems needing large amounts of data since it is easy to use, efficient in terms of computation, memory-light, insensitive to diagonal resizing of the gradients, and invariant to diagonal resizing of the gradients. [44]

3.4.4 Evaluation Tools

The first approach for evaluating the mode was to use validation where during the preparation of the dataset, we divided data into two subsets one reserved for training and the other for testing. Secondary to the outputs in Figure 3.7, the confusion matrix, has been used to examine the accurate performance, misclassification, precision, sensitivity, specificity, as well as model prevalence. The four outputs of the matrix are defined as:

a) True Negative (a): the percentage of situations where a classification is valid, indicating that normal images were easily detected as normal.
b) False positive (b): The proportion of false classification, or when normal images are mistakenly classified as abnormal.
c) False Negative (c): The proportion of times an example is falsely detected as being negative, which results in abnormal images being mistakenly classified as normal.
d) True Positive (d): the percentage of accurately classifying situations in which abnormal images are identified as such.

	Detected normal_0	Detected abnormal_1	c+$b \equiv$ *False classified*
Actual normal_0	a TN	b FP	$a + b \equiv$ *actual normal*
Actual abnormal_1	c FN	d TP	$c + d \equiv$ *actual* abnormal
n = a+b+c+d $\equiv$ *random sampled images*	a+$c \equiv$ *detected normal*	b+$d \equiv$Detected abnormal	$a + d \equiv$ *True classified*

Figure 3. 7. Confusion matrix structure

Figure 3.8 explains the characteristics of the model. Firstly, the model's specificity is the true negative rate or detected true normal to actual normal images, whereas the model's sensitivity is the true positive rate or the ratio detected abnormal to actual abnormal. Furthermore, although the model's accuracy represents the accurate classification rate of the sample, prevalence represents the actual frequency of the abnormal condition. Additionally, precision measures how well abnormal lesions are identified. Finally, the F1 score, which represents performance even for

imbalanced classification, has been produced using the same matrix and it is defined as the harmonic mean of precision and recall. [45]

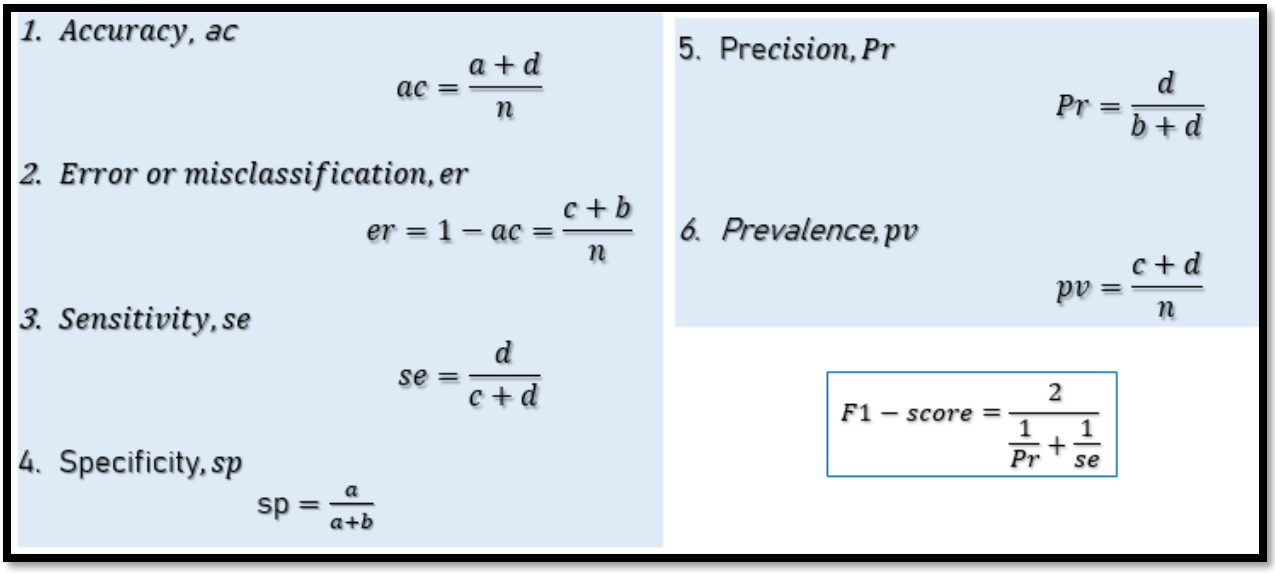

Figure 3. 8. Model's Characteristics

In addition to fore above-mentioned confusion matrix and for easier reading, the performance of this binary classifier across all possible thresholds has been summarized using an ROC plot. It was created by plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis).

3.3 Summary

The methods and techniques utilized in the study to develop our ResNET50 model were highlighted in this chapter. The chapter also included an overview of the tools used to present the findings. In the following chapter, we will go into greater detail about those findings.

CHAPTER 4. THE PROJECT RESULTS

4.1 Introduction

This chapter includes discussions of the results of the study. It also provides the model summary and the essential interpretation of the study findings.

4.2 Model Summary

This model is made up of Resnet50 layers, 2D global average pooling layers, and 2D local average pooling layers. Dropout layers and batch normalization layers can also be turned on and off. The model has been tested and shown to function well without the need for data normalization to rearrange or remove any disorganized or redundant features.

The model consisted of extracting the parameters and there were 23, 600, 002 total parameters whose 57 216 were non-trainable. This means that 23,542,786 parameters are updated during learning through iterations. The training cycle was repeated 40 times until the network's validation accuracy was scored at 0.78.

4.3. Performance Analysis

Based on Figure 3.8 and Figure 4.1, the performance characteristics are evaluated, and Table 4. 1 is generated to highlight key numerical values for those properties as well as their implications.

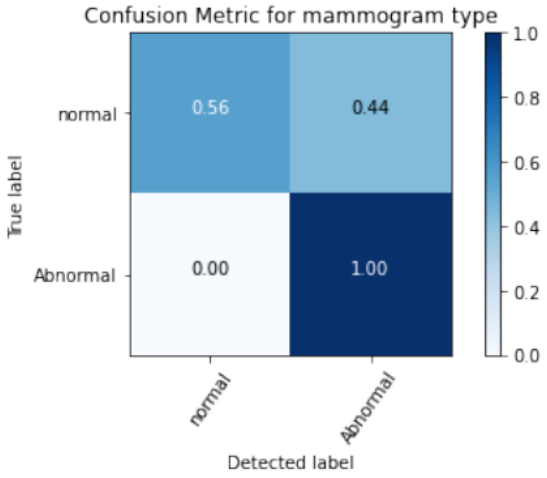

Figure 4. 1 Model Confusion Matric for classifying mammogram lesions type

Figure 4.1 demonstrates that while this model correctly classifies abnormal images, there is a misclassification when classifying normal images. The dataset for typical scenarios and the fitting of the current model design could both be problematic. Table 4.1 demonstrates that although this model is not perfect, it has good accuracy since it is over 0.7. The model is also quite sensitive, which means that it can accurately predict favourable results. Finally, other analyzed qualities are still not at a good level to represent an ideal AI model.

#	characteristics	value	Implication/comments
1	*Accuracy*	0.78	Good
2	Error or misclassification	0.22	Good
3	Sensitivity or recall	1.00	Perfect
4	Specificity	0.56	moderate
5	Pre*cision*	0.69	satisfactory
6	*Prevalence*	0.5	satisfactory
7	F1 score	0.52	satisfactory

Table 4. 1. Model performance main characteristics

The model was rated an area under a curve of 0.780, as displayed in Figure 4.2. This indicates that the model can differentiate between normal and abnormal images with an accuracy rate of approximately 78.0%, which is almost comparable to the model's accuracy. The ideal threshold for the model is represented by the point (x, y) closest to the top left corner, which provides the best balance of true positive and false positive rates. As per the classification options, abnormal images are rated as true (1) and normal images are rated as false (0). It is also important to note that the model demonstrated 100% sensitivity, which is represented by showing sensitivity on the y-axis and 1-specificity on the x-axis, with the ideal threshold having y equal to 1.

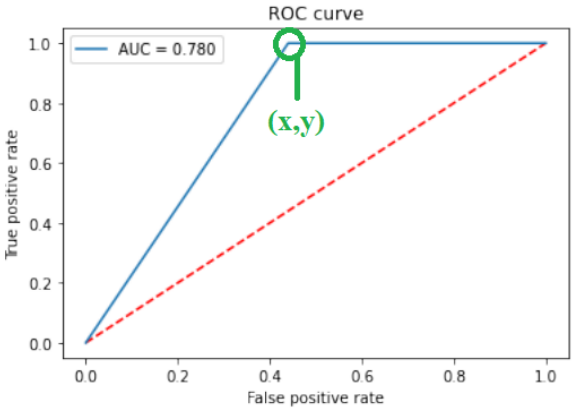

Figure 4. 2. ROC curve and AUC

As shown in Figure 4.3, the random image display revealed the correct mammography lesions detection or binary classification as determined results matched actual pre-verified facts.

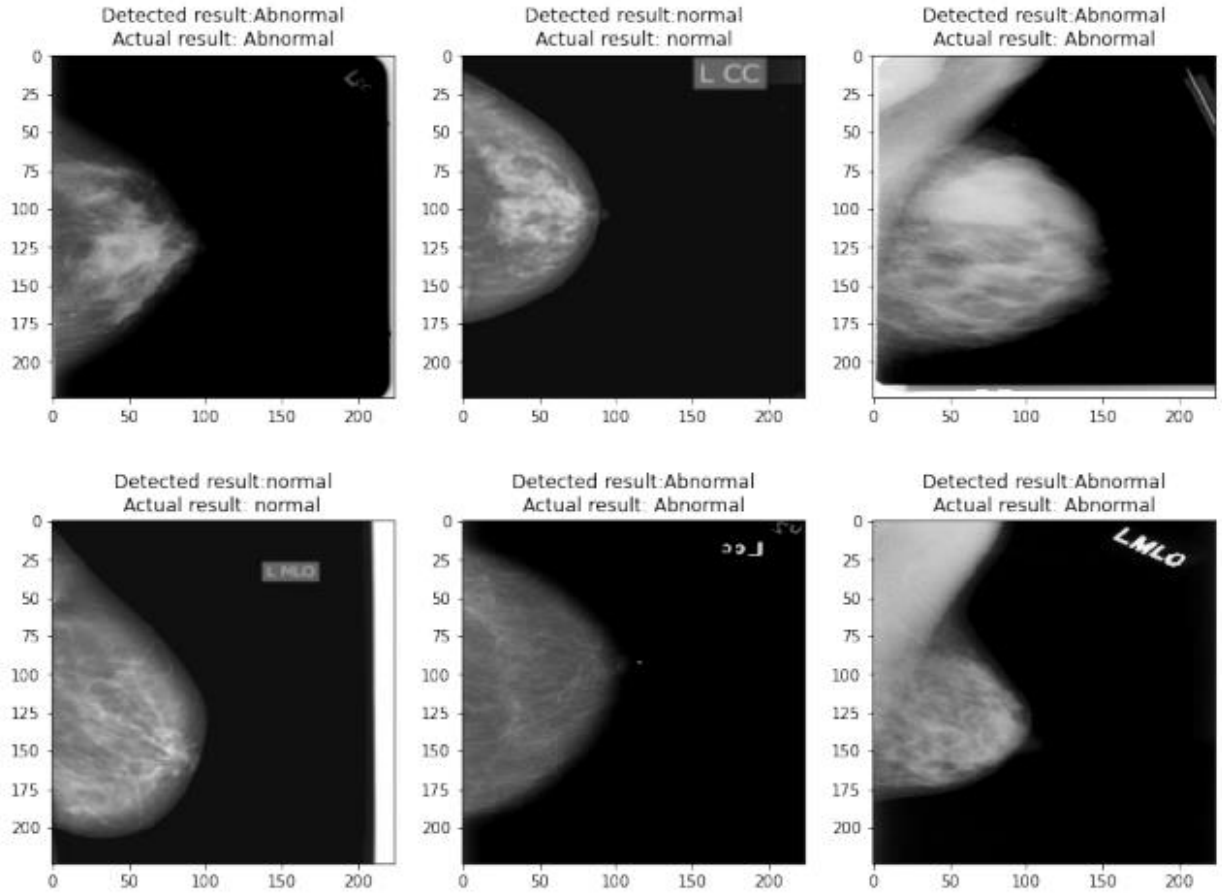

Figure 4. 3. Example of classified images

4.2 Summary

The model performs well overall. The performance characteristics indicate that the research objectives have been achieved at an impressive rate. However, some characteristics could be improved by increasing the learning rate or by expanding the data set. It has also been observed that this model's performance varies depending on the iteration rates and convolutional kernel type used.

CHAPTER 5. CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In general, the goal of this work was to create a deep learning artificial intelligence-based algorithm that would act as a decision support system for radiologists assessing digital screening mammographic images through training and testing the CNN model's performance.

This study proved that as data are loaded into the model, it may learn to classify the normal and abnormal mammograms effectively. When provided with a specific mathematical value at different levels, the capacity to learn from the most significant feature of all layers resulted in a perfectly sensitive system with a good accuracy of 0.78 but some properties are not achieving the implementable phase and hence need more improvement.

5.2 Recommendations

The use of AI deep learning algorithms in medical imaging is a very recent innovation, and we suggest the following:

- → To the cycle of future research try to improve the characteristics of this model by applying the optimizers as there was not high emphasis during this study.
- → For future researchers, make the comparative study and evaluate the performances among different models to decide which one should be implemented.
- → To help future research, the Rwandan government, through the MoH, should give hospital personnel tools for de-identifying images and building public databases.
- → The University of Rwanda should encourage the general population to learn radiology so that the country gets a big pool of radiologists who can help with image annotation and the construction of future large datasets.
- → To CEBE, instil into the biomedical students advanced programming languages so that they can conduct research involving technologies like this application of AI in medical imaging healthcare services.
- → To the national hospitals, plan for introducing AI to facilitate the radiologists in decision making. This should reduce the process of taking more than one exam once one modality is confused and it will minimize the recommendation of another modality.
- → To KHF find ways to store radiographs in the form of DICOM as standard format with the high-performance review.

REFERENCES

[1] S. Patil *et al.*, 'Reviewing the Role of Artificial Intelligence in Cancer', *Asian Pac. J. Cancer Biol.*, vol. 5, no. 4, pp. 189–199, Dec. 2020, doi: 10.31557/apjcb.2020.5.4.189-199.

[2] Non-Communicable Diseases Division, Rwanda Biomedical Centre, Ministry of Health, Rwanda, 'Rwanda National Cancer Control Plan 2020-2024', Ministry of Health, Rwanda Republic, Feb. 2020. [Online]. Available: https://www.rbc.gov.rw/fileadmin/user_upload/result/Rwanda_NCCP_Final_Signed.pdf

[3] Y. Sun *et al.*, 'Deep learning model improves radiologists' performance in detection and classification of breast lesions', *Chin. J. Cancer Res.*, vol. 33, no. 6, pp. 682–693, 2021, doi: 10.21147/j.issn.1000-9604.2021.06.05.

[4] D. S. O'Neil *et al.*, 'Breast Cancer Care Quality in South Africa's Public Health System: An Evaluation Using American Society of Clinical Oncology/National Quality Forum Measures', *J. Glob. Oncol.*, no. 5, pp. 1–16, Dec. 2019, doi: 10.1200/JGO.19.00171.

[5] A. Kolak *et al.*, 'Primary and secondary prevention of breast cancer', *Ann. Agric. Environ. Med.*, vol. 24, no. 4, pp. 549–553, Dec. 2017, doi: 10.26444/aaem/75943.

[6] P. C. Igiraneza, L. A. Omondi, B. Nikuze, M. G. Uwayezu, M. Fitch, and G. Niyonsenga, 'Factors influencing breast cancer screening practices among women of reproductive age in South Kayonza District, Rwanda', *Can. Oncol. Nurs. J.*, vol. 31, no. 3, pp. 251–257, Jul. 2021, doi: 10.5737/23688076313251257.

[7] Harvard Medical School *et al.*, 'Imaging in the Land of 1000 Hills: Rwanda Radiology Country Report', *J. Glob. Radiol.*, vol. 1, no. 1, 2015, doi: 10.7191/jgr.2015.1004.

[8] F. Rubagumya *et al.*, 'State of Cancer Control in Rwanda: Past, Present, and Future Opportunities', *JCO Glob. Oncol.*, no. 6, pp. 1171–1177, Nov. 2020, doi: 10.1200/GO.20.00281.

[9] Canadian Cancer Society, 'The breasts', 2023. https://cutt.ly/B91bvWW (accessed Mar. 02, 2022).

[10] Abuobieda Bala Abusharib, *General Pathology Made Easy*, 2nd ed. 2017.

[11] P. Whelehan, A. Evans, M. Wells, and S. MacGillivray, 'The effect of mammography pain on repeat participation in breast cancer screening: A systematic review', *The Breast*, vol. 22, no. 4, pp. 389–394, Aug. 2013, doi: 10.1016/j.breast.2013.03.003.

[12] Atlas, 'ACR BI-RADS', *quick reference*. https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/BIRADS-Poster.pdf (accessed Jan. 20, 2023).

[13] National Library of Medecine, 'Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer.', 8600 Rockville Pike Bethesda, MD 20894: National Library of Medecine. Accessed: Mar. 02, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK222350/

[14] American Cancer Society, 'Breast Cancer Early Detection and Diagnosis'. American Cancer Society. Accessed: Dec. 01, 2023. [Online]. Available: https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection.html

[15] K. J. Geras, R. M. Mann, and L. Moy, 'Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives', *Radiology*, vol. 293, no. 2, pp. 246–259, Nov. 2019, doi: 10.1148/radiol.2019182627.

[16] M. J. Canty, *Image analysis, classification and change detection in remote sensing: with algorithms for Python*, Fourth edition. Boca Raton, FL: CRC Press/Taylor & Francis Group, 2019.

[17] K. D. Toennies, *Guide to Medical Image Analysis*. London: Springer London, 2012. doi: 10.1007/978-1-4471-2751-2.

[18] R. J. Gillies and M. B. Schabath, ‘Radiomics Improves Cancer Screening and Early Detection’, *Cancer Epidemiol. Biomarkers Prev.*, vol. 29, no. 12, pp. 2556–2567, Dec. 2020, doi: 10.1158/1055-9965.EPI-20-0075.

[19] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, ‘Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review’, *J. Med. Internet Res.*, vol. 21, no. 7, p. e14464, Jul. 2019, doi: 10.2196/14464.

[20] Y. J. Suh, J. Jung, and B.-J. Cho, ‘Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning’, *J. Pers. Med.*, vol. 10, no. 4, p. 211, Nov. 2020, doi: 10.3390/jpm10040211.

[21] E. Michael, H. Ma, H. Li, and S. Qi, ‘An Optimized Framework for Breast Cancer Classification Using Machine Learning’, *BioMed Res. Int.*, vol. 2022, pp. 1–18, Feb. 2022, doi: 10.1155/2022/8482022.

[22] T. Mahmood, J. Li, Y. Pei, F. Akhtar, M. U. Rehman, and S. H. Wasti, ‘Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach’, *PLOS ONE*, vol. 17, no. 1, p. e0263126, Jan. 2022, doi: 10.1371/journal.pone.0263126.

[23] J. Joy, D. Petitti, and E. Penhoet, ‘Institute of Medicine (US) and National Research Council (US) Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer’, in *Saving Women’s Lives: Strategies for Improving Breast Cancer Detection and Diagnosis*, Washington (DC): National Academies Press (US), 2005. Accessed: Jan. 15, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK22311/

[24] G. Ghalachyan, ‘Multidimensional Evaluation of AI Models Performance with Socio-Economic Metrics’, *Messenger Armen. State Univ. Econ.*, no. 6, pp. 126–136, 2021, doi: 10.52174/1829-0280_2021_6_126.

[25] R. R. Fletcher *et al.*, ‘The Use of Mobile Thermal Imaging and Deep Learning for Prediction of Surgical Site Infection’, in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico, Nov. 2021, pp. 5059–5062. doi: 10.1109/EMBC46164.2021.9630094.

[26] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, and A. Valenzuela, ‘A Review of Convolutional Neural Network Applied to Fruit Image Processing’, *Appl. Sci.*, vol. 10, no. 10, p. 3443, May 2020, doi: 10.3390/app10103443.

[27] S. S. Yadav and S. M. Jadhav, ‘Deep convolutional neural network based medical image classification for disease diagnosis’, *J. Big Data*, vol. 6, no. 1, p. 113, Dec. 2019, doi: 10.1186/s40537-019-0276-2.

[28] ‘Breast cancer detection based on thermographic images using machine learning and deep learning algorithms’, *Int. J. Eng. Comput. Sci.*.

[29] M. A. Al-Ansari, S.-M. Han, and T.-S. Kim, ‘Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms’, *Comput. Methods Programs Biomed.*, vol. 196, p. 105584, Nov. 2020, doi 10.1016/j.cmpb.2020.105584.

[30] S. Majumdar, P. Pramanik, and R. Sarkar, ‘Gamma function based ensemble of CNN models for breast cancer detection in histopathology images’, *Expert Syst. Appl.*, vol. 213, p. 119022, Mar. 2023, doi: 10.1016/j.eswa.2022.119022.

[31] A. Kolchev, D. Pasynkov, I. Egoshin, I. Kliouchkin, O. Pasynkova, and D. Tumakov, ‘YOLOv4-Based CNN Model versus Nested Contours Algorithm in the Suspicious Lesion Detection on the Mammography Image: A Direct Comparison in the Real Clinical Settings’, *J. Imaging*, vol. 8, no. 4, p. 88, Mar. 2022, doi 10.3390/jimaging8040088.

[32] S. M. Alizadeh and A. Mahloojifar, ‘Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features’, *Int. J. Imaging Syst. Technol.*, vol. 31, no. 2, pp. 695–707, Jun. 2021, doi: 10.1002/ima.22490.

[33] R. M. Al-Tam *et al.*, ‘A Hybrid Workflow of Residual Convolutional Transformer Encoder for Breast Cancer Classification Using Digital X-ray Mammograms’, *Biomedicines*, vol. 10, no. 11, p. 2971, Nov. 2022, doi 10.3390/biomedicines10112971.
[34] M. A. Al-antari, M. A. Al-masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, ‘A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification’, *Int. J. Med. Inf.*, vol. 117, pp. 44–54, Sep. 2018, doi: 10.1016/j.ijmedinf.2018.06.003.
[35] N. Chouhan, A. Khan, J. Z. Shah, M. Hussnain, and M. W. Khan, ‘Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography’, *Comput. Biol. Med.*, vol. 132, p. 104318, May 2021, doi: 10.1016/j.compbiomed.2021.104318.
[36] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, ‘Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets’, *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 686–696, Mar. 2019, doi: 10.1109/TMI.2018.2870343.
[37] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, ‘Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis’, *IEEE Access*, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/ACCESS.2020.2993536.
[38] A. K. Das, J. Nayak, B. Naik, S. Vimal, and D. Pelusi, Eds., *Computational intelligence in pattern recognition: proceedings of CIPR 2022*. Singapore: Springer, 2022.
[39] B. Dayı, H. Üzen, İ. B. Çiçek, and Ş. B. Duman, ‘A Novel Deep Learning-Based Approach for Segmentation of Different Type Caries Lesions on Panoramic Radiographs’, *Diagnostics*, vol. 13, no. 2, p. 202, Jan. 2023, doi: 10.3390/diagnostics13020202.
[40] SHUBHAM GUPTA, ‘Understanding Image Recognition and Its Uses’, Dec. 11, 2019. https://www.einfochips.com/blog/understanding-image-recognition-and-its-uses/ (accessed May 20, 2022).
[41] L. Bin, W. Liang, and Y. Guosheng, ‘A Graph Total Variation Regularized Softmax for Text Generation’. arXiv, Dec. 31, 2020. Accessed: Jan. 15, 2023. [Online]. Available: http://arxiv.org/abs/2101.00153
[42] Raúl Gómez Bruballa, ‘Understanding Categorical Cross-Entropy Loss, Binary Cross-Entropy Loss, Softmax Loss, Logistic Loss, Focal Loss and all those confusing names’, May 23, 2018. https://gombru.github.io/2018/05/23/cross_entropy_loss/ (accessed Jan. 16, 2023).
[43] Akira, ‘Building AI-First Enterprise’, *Adam Optimization*, 2020. https://www.akira.ai/glossary/adam-optimization (accessed Jan. 26, 2023).
[44] D. P. Kingma and J. Ba, ‘Adam: A Method for Stochastic Optimization’. arXiv, Jan. 29, 2017. Accessed: Jan. 26, 2023. [Online]. Available: http://arxiv.org/abs/1412.6980
[45] ‘F-score’. Wikipedia, Dec. 05, 2022. Accessed: Aug. 01, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=F-score&oldid=1125694227

APPENDIX 1: THE UTILIZED CODES

1 Loading & Preprocessing

```
# Importing Libraries
import tensorflow as tf
import matplotlib.pyplot as plt
import pandas as PD
import numpy as np
import itertools
import scipy
import math
import json
import gc
import os
import cv2
import warnings
warnings.filter warnings("ignore")
from tensorflow.keras.applications.resnet50 import ResNet50
from tensorflow. keras.optimizers import Adam
from keras.callbacks import Callback, ModelCheckpoint, ReduceLROnPlateau,
␣
↪TensorBoard
from keras. preprocessing.image import ImageDataGenerator
from keras. utils.np_utils import to_categorical
from keras.models import Sequential
from keras import backend as K
from keras import layers
from sklearn.model_selection import train_test_split
from sklearn.metrics import cohen_kappa_score, accuracy_score
from sklearn import metrics
from tqdm import tqdm
from functools import
partial from collections
import Counterfrom
PIL import Image

%matplotlib inline
```

```
#Transfer 'jpg' images to an array
IMG#Load images from the drive

def Dataset_loader(DIR,
RESIZE):
IMG = []
read = lambda imname: np.asarray(Image.open(imname).convert("RGB"))
for IMAGE_NAME in
tqdm(os.listdir(DIR)):PATH =
os.path.join(DIR,IMAGE_NAME)
```

```
    _, type = os.path.plaintext(PATH)
    if ftype == ".jpg" or ftype == ".jpeg":img =
    read(PATH)
    img = cv2.resize(img, (RESIZE,RESIZE))
    IMG.append(np.array(img))
    return IMG
normal_train = np.array(Dataset_loader('datasetKFH/train/0_normal',224))
cancer_train = np.array(Dataset_loader('datasetKFH/train/1_abnormal',224))
normal_test = np.array(Dataset_loader('datasetKFH/val/0_normal',224))
cancer_test = np.array(Dataset_loader('datasetKFH/val/1_abnormal',224))
```

2 Create Label

```
# Creating labels: breast Cancer: normal vs.
cancerous

normal_train_label = np.zeros(len(normal_train))
cancer_train_label = np.ones(len(cancer_train))

normal_test_label = np.zeros(len(normal_test))
cancer_test_label = np.ones(len(cancer_test))
# Merge data

X_train = np.concatenate((normal_train, cancer_train), axis = 0)
Y_train = np.concatenate((normal_train_label, cancer_train_label), axis = 0)

X_test = np.concatenate((normal_test, cancer_test), axis = 0)
Y_test = np.concatenate((normal_test_label, cancer_test_label), axis = 0)

# Shuffle train data
s = np.arange(X_train.shape[0])
np. random.shuffle(s)
X_train = X_train[s]
Y_train = Y_train[s]
# Shuffle test data
s = np.arrange(X_test.shape[0])
np. random.shuffle(s)
X_test = X_test[s]
Y_test = Y_test[s]
# To categorical
Y_train = to_categorical(Y_train, num_classes= 2)
Y_test = to_categorical(Y_test, num_classes= 2)
```

3 Train and Evaluation split

```
x_train, x_val, y_train, y_val = train_test_split(X_train, Y_train,
test_size=0.1, random_state=1
)
```

4 Display Some collected Images

```
w=60
h=40
fig=plt.figure(figsize=(15, 15))
columns = 4
rows = 3
for i in range(1, columns*rows +1):
    ax = fig.add_subplot(rows, columns, i)
    if np.argmax(Y_train[i]) == 0:
    ax.title.set_text('Normal image')
    else:
      ax. title.set_text('Abnormal image')
      plt.imshow(x_train[i], interpolation='nearest')
plt.show()
```

5 Data Generator

```
BATCH_SIZE = 16
```

```
# Using original generator
train_generator = ImageDataGenerator( zoom_range=2,
 # set range for random zoom
rotation_range = 90,
horizontal_flip=True,      # randomly flip images
vertical_flip=True,     # randomly flip images
      )
```

6 Creating the Model

```
def build_model(backbone, lr=1e-4): model = Sequential() model.add(backbone)
model.add(layers.GlobalAveragePooling2D())
model.add(layers.Dropout(0.5))
model.add(layers.BatchNormalization())
model.add(layers.Dense(2, activation='softmax'))
model.compile( loss='binary_crossentropy',
optimizer=Adam(lr=lr),
metrics=['accuracy']
)
return model #returning the model
```

```
K.clear_sess
ion()
gc.collect()
resnet =
ResNet50(
weights='imagenet',
include_top=False,
input_shape=(224,224,3)
)
model = build_model(resnet ,lr =
1e-4)model.summary()
```

7 Learning Rate Reduction

```
learn_control = ReduceLROnPlateau(monitor='val_accuracy', patience=3,
verbose=1,factor=0.2, min_lr=1e-7)

# Checkpoint
filepath="weights. best.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='val_accuracy', verbose=1,
↪save_best_only=True, mode='max')
```

8 Training

```
model.load_weights("weights.best.hdf5")

history = model.fit_generator(
train_generator.flow(x_train, y_train, batch_size=BATCH_SIZE),
steps_per_epoch=x_train.shape[0] / BATCH_SIZE,
epochs=40,
validation_data=(x_val, y_val),
callbacks=[learn_control, checkpoint])
```

9 Detection: Normal or Abnormal

```
model.load_weights("weights.best.hdf5")
Y_val_pred = model.predict(x_val)
accuracy_score(np.argmax(y_val, axis=1), np.argmax(Y_val_pred, axis=1))
Y_pred = model.predict(X_test)
tta_steps = 10
predictions = []

for i in tqdm(range(tta_steps)):
preds = model.predict_generator(train_generator.flow(X_test,
↪batch_size=BATCH_SIZE, shuffle=False),
steps = len(X_test)/BATCH_SIZE)

predictions.append(preds)
gc.collect()

Y_pred_tta = np.mean(predictions, axis=0)
```

10 Confusion Matrix

```
from sklearn.metrics import confusion_matrix

def plot_confusion_matrix(cm, classes,
normalize=False, title='Confusion matrix',
cmap=plt.cm.Blues):
"""
This function prints and plots the confusion matrix.
Normalization can be applied by setting
```

```
    `normalize=True`."""
    if normalize:
    cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
    print("Normalized confusion matrix")
    else:
    print('Confusion matrix, without normalization')
    print(cm)
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=55)
    plt.yticks(tick_marks, classes)

    fmt = '.2f' if normalize else 'd'
    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
    plt.text(j, i, format(cm[i, j], fmt),
    horizontalalignment="center",
    color="white" if cm[i, j] > thresh else "black")

    plt.ylabel('True label')
    plt.xlabel('Predicted label')
    plt.tight_layout()

    cm = confusion_matrix(np.argmax(Y_test,
    axis=1), np.argmax(Y_pred, axis=1))
    cm_plot_label =['normal', 'abnormal']
    plot_confusion_matrix(cm, cm_plot_label, title ='Confusion Metric for mammogram
    ↪lesion type')
```

11 **Binary Classification Report**

```
    from sklearn.metrics import classification_report
    classification_report( np.argmax(Y_test, axis=1), np.argmax(Y_pred_tta, axis=1))
```

12 **Receiver Operating Characteristics(ROC) & Area Under the Curve(AUC)**

```
    from sklearn.metrics import roc_auc_score, auc
    from sklearn.metrics import roc_curve

    roc_log = roc_auc_score(np.argmax(Y_test, axis=1), np.argmax(Y_pred_tta,
    ↪axis=1))
    false_positive_rate, true_positive_rate, threshold = roc_curve(np.
    ↪argmax(Y_test, axis=1), np.argmax(Y_pred_tta, axis=1))
    area_under_curve = auc(false_positive_rate, true_positive_rate)

    plt.plot([0, 1], [0, 1], 'r--')
    plt.plot(false_positive_rate, true_positive_rate, label='AUC = {:.3f}'.
    ↪format(area_under_curve)


    plt.xlabel('False positive rate')
```

```
plt.ylabel('True positive rate')
plt.title('ROC curve')
plt.legend(loc='best')
plt.show()plt.close()
```

13 Display some classified Images

```
i=0 prop_class=[] mis_class=[]

for i in range(len(Y_test)):
  if(np.argmax(Y_test[i])==np.argmax(Y_pred_tta[i])):
      prop_class.append(i)
    if(len(prop_class)==8):
       break
i=0
for i in range(len(Y_test)):
  if(not np.argmax(Y_test[i])==np.argmax(Y_pred_tta[i])):
  mis_class.append(i)
   if(len(mis_class)==8):
       break
w=60
h=40
fig=plt.figure(figsize=(18, 10))
columns = 4
rows = 2
def Transfername(namecode):
if namecode==0:
return "normal"
else:
return "abnormal"

for i in range(len(prop_class)):
ax = fig.add_subplot(rows, columns, i+1)
ax.set_title("Detected result:"+ Transfername(np.
↪argmax(Y_pred_tta[prop_class[i]]))
+"\n"+"Actual result: "+ Transfername(np.
↪argmax(Y_test[prop_class[i]])))
     plt.imshow(X_test[prop_class[i]], interpolation='nearest')
plt.show()
```