

**CLASSIFICATION OF BREAST CANCER HISTOPATHOLOGY
IMAGES USING A CONVOLUTIONAL NEURAL NETWORK MODEL**

**SIMONYAN, EMMANUEL OLUWATOBI
(19PCJ01981)**

AUGUST, 2022

**CLASSIFICATION OF BREAST CANCER HISTOPATHOLOGY
IMAGES USING A CONVOLUTIONAL NEURAL NETWORK MODEL**

BY

**SIMONYAN, EMMANUEL OLUWATOBI
(19PCJ01981)**

**B.Eng Electrical and Electronics Engineering,
Michael Okpara University of Agriculture Umudike, Nigeria**

**A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE
STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE AWARD OF THE MASTER OF ENGINEERING (M.Eng.) DEGREE IN
COMPUTER ENGINEERING IN THE DEPARTMENT OF ELECTRICAL
AND INFORMATION ENGINEERING, COLLEGE OF ENGINEERING,
COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA**

AUGUST, 2022

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Master of Computer Engineering (M.Eng.) in Computer Engineering in the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Nigeria.

Mr. Taiwo B. Erewunmi
Secretary, School of Postgraduate

Signature and Date

Prof. Akan B. Williams
Dean, School of Postgraduate Studies

Signature and Date

DECLARATION

I, **SIMONYAN, EMMANUEL OLUWATOBI (19PCJ01981)** declare that this research was carried out by me under the supervision of **Dr. Joke A. Badejo** of the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Nigeria. I attest that this dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

SIMONYAN, EMMANUEL OLUWATOBI

Signature and Date

CERTIFICATION

We certify that this thesis titled "**CLASSIFICATION OF BREAST CANCER HISTOPATHOLOGY IMAGES USING A CONVOLUTIONAL NEURAL NETWORK MODEL**" is an original research work carried out by **SIMONYAN, EMMANUEL OLUWATOBI (19PCJ01981)** in the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria under the supervision of **Dr Joke. A Badejo**. We have examined and found this work acceptable as part of the requirements for the award of Masters of Engineering degree in Computer Engineering.

Dr Joke A. Badejo
Supervisor

Signature and Date

Prof. Emmanuel Adetiba
Head of Department

Signature and Date

Dr. Jonathan Gana Kolo
External Examiner

Signature and Date

Prof. Akan B. Williams
Dean, School of Postgraduate Studies

Signature and Date

DEDICATION

This research work is dedicated to the great father and creator of the world who showed his self-mighty during this program. To my parents Prof. Kayode and Prof. (Mrs.) Judith Simonyan for their investment and help in making this journey a reality.

ACKNOWLEDGMENTS

I want to acknowledge God for His help. God's word is accurate, I have come to experience the tangibility of the word of God and the comforter, helper and spirit of Wisdom, the Holy spirit for this guidance. Inspired by the book of Jeremiah 29 vs 11 which as, “for I, know the thoughts that I think toward you saith the Lord, thoughts of peace, and not of evil, to give you an expected end”. Also, holding the word in Philippians 1 vs 6, “being confident of this very thing, that he which hath begun a good work in you will perform it until the day of Jesus Christ, kept me going and believing that it will end well. May all the glory be unto the Lord”.

It is of great joy and deep appreciation to the visioner and Chancellor of Covenant University, Dr. David O. Oyedepo, for setting up this great institution of learning and for his immersive sacrifice in making this platform to be world-class, and for the vision put in place in raising new generation of leaders in their various fields. Most especially in imbibing the core values, two of which now applies to me, they are having a possibility mentality and building capacity.

My most sincere and gratitude to my supervisor Dr. Joke A. Badejo, for her continuous mentorship, guidance, examination, the contribution of her rich knowledge and keen patience in working with me. I express my profound appreciation to the Head of the Department of Electrical Information Engineering, Prof. E. Adetiba, for being a father and for his immerse contributions and counsel in this research. All department faculty members for their support in bringing the best out of this work. Also, the Vice-Chancellor, Prof. Abiodun H. Adebayo, and the Dean School of Post Graduate School, Prof. Akan B. Williams, for providing an enabling environment for this also, to my long-life mentor Engr. Igiri Onaji for his support and providing comments during my research and guidance on balancing leadership, research and personal development.

I really want to appreciate the input of my colleagues and coursemates during times of hitches and challenges during of my research for their immense contribution.

I give my profound thanks and appreciation to one of the PhD scholars, Engr. Akinrinmade Adekunle, his understanding of deep learning and MATLAB is renowned. Every time I needed help and guidance, he was always open to help out, thank you sir, and Engr. John Wejin for his scrutiny of the write-up.

To all executives and members of the Postgraduate student fellowship for their understanding and support when it looked overwhelming, and when I needed prayers and support, thank you. I will not forget the members of the Postgraduate student council for their understanding and diplomacy in leading the PG Students.

I also appreciate Uncle Emmanuel and Auntie Kelechi Onyedis's support, hospitality, and accommodation during the break periods.

I am grateful to my friends Tolu, Obinna, Ugochukwu, Shola, and Amarachi, for their words of encouragement.

To my parent, Prof. (Engr.) Kayode and Prof. (Mrs.) Judith Simonyan and my sister Itunuoluwa Simonyan for their broad support for sponsoring my research programme and for the decision to invest this in me, for their guidance and counsel. Raising me in the way of the Lord, most especially teaching me to take the crucial life-changing decision of overcoming whatever limit I am confronted with and being strong to overcome by trusting in God and putting efforts into finding solutions to these challenges.

TABLE OF CONTENTS

COVER PAGE	i
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
DEDICATION	vi
ACKNOWLEDGMENTS	vii
TABLES OF CONTENTS	viii
LIST OF FIGURES	xi
LIST OF TABLES	xiv
LIST OF ABBREVIATION	xv
GLOSSARY	xviii
ABSTRACT	xx
CHAPTER ONE: INTRODUCTION	1
1.1 Background of Research	1
1.2 Motivation for the Research	3
1.3 Statement of Problem	3
1.4 Aim and Objectives of the Research	4
1.5 Scope and Limitation of the Research	4
1.6 Dissertation Organisation	4
CHAPTER TWO: LITERATURE REVIEW	5
2.1 Chapter Introduction	5
2.2 Cancer and its Impact	5
2.3 Breast Cancer	6
2.3.1 Characteristics of Breast Cancer	8
2.3.2 Classes of breast cancer	9
2.3.3 Breast cancer diagnosis techniques	10
2.3.4 Types of Breast Cancer	13
2.3.5 Breast cancer Grading & Staging	14
2.3.6 Causes of Breast Cancer	15
2.3.7 Cancer Treatment Methods	17
2.4 Computer Vision in Cancer Diagnosis and Treatment	19
2.5 Machine Learning Approaches to Medical Image Classification	21
2.5.1 Artificial Neural Networks (ANN)	22
2.5.2 Operations of the Neural Networks	25
2.5.3 Synaptic weights	25
2.5.4 Layers	26
2.5.5 Summation Function	26
2.5.6 Activation Function	26
2.5.7 Linear Activation Function	27
2.5.8 Sigmoid Transfer Function	28
2.6 Convolutional Neural Network (CNN)	28

2.6.1 Convolutional Neural Network (CNN) Architecture	31
2.6.2 CNN Layers	31
2.6.3 Pretrained Deep Neural Network	33
2.7 Evaluation Metrics	40
2.8 A Review of Related Works	43
2.9 Chapter Summary	52
CHAPTER THREE: MATERIALS AND METHODS	53
3.1 Chapter Introduction	53
3.2 System Architecture	53
3.2.1 Curation of Histopathology Image Dataset	53
3.2.2 Dataset Description	53
3.3 Data Pre-processing	59
3.3.1 Data Augmentation	60
3.4 Transfer Learning	60
3.5 Dataset Experimental Protocol	60
3.6 Chapter Summary	73
CHAPTER FOUR: RESULTS AND DISCUSSION	74
4.1 Chapter Introduction	74
4.2 Result for Binary classification for the different CNN networks	75
4.2.1 Discussion of the binary classification	77
4.2.2 Training graph	77
4.3 Confusion matrix	94
4.4 Comparison of the different Deep Neural Network	94
4.5 Comparison with the state-of-the-art result	96
4.6 Chapter Summary	97
CHAPTER FIVE: CONCLUSION AND RECOMMENDATION	98
5.1 Summary	98
5.2 Conclusion	98
5.3 Contribution to knowledge	99
5.4 Recommendation	99
REFERENCES	99
APPENDIX A	108
APPENDIX B	118

LIST OF FIGURES

FIGURES	TITLE OF FIGURES	PAGES
1.1	WHO Global Cancer Observatory	1
2.1	Demography of cancer cases around the group as estimated by the WHO report 2021	5
2.2	Overview of medical imaging analysis framework	6
2.3	A schematic illustration of the female Mammary Gland	6
2.4	The Microscopic structure of the normal breast tissue	8
2.5	Generic medical image analysis pipeline	12
2.6	The Inner structure of the duct	14
2.7	A schematic representation of an ANN	23
2.8	Biological Neuron	24
2.9	Showing the most common nonlinear activation functions	28
2.10	Schematic diagram of CNN	31
2.11	An Illustration of the transfer learning technique	34
2.12	Graphical illustration of Transfer learning of pre-trained networks	34
2.13	A WSI image fed as an input into deep Learning (CNN model)	35
2.14	A graphical representation of the ROC curve	41
2.15	Graphical representation of a 2 x 2 confusion matrix plot	43
3.1	Flowchart of the proposed methodology	54
3.2	Adenosis(A) breast cancer slide image at various magnifications	56
3.3	Fibroadenoma (FB) breast cancer slide image at various magnifications	56
3.4	Phyllodes tumour (PT) breast cancer slide image at various magnifications	56
3.5	Tubular adenoma (TB) breast cancer slide image at various magnifications	56
3.6	Ductal carcinoma (DC) breast cancer slide image at various magnifications	57
3.7	Lobular carcinoma (LC) breast cancer slide image at various magnifications	57
3.8	Mucinous carcinoma (MC) breast cancer slide image at various magnifications	57
3.9	Papillary carcinoma (PC) breast cancer slide image at various magnification	57
3.10	Histopathology benign image samples at 200X magnification	58

3.11	Histopathology malignant image samples at 200X magnification	58
3.12	General experimental protocol and the training setup, as well as the final results	62
3.13	The experimental protocol 1 with training set (BreakHis) and the test set (BreakHis)	63
3.14	The second experimental protocol, comprising the training set (BreakHis) and the test set (Histo)	64
3.15	The experimental protocol 3 with the training set (Histo) and the test set (BreakHis)	65
3.16	The experimental protocol four (4) setup training the CNN models on the training set (Histo) and the test set (BreakHis)	65
3.17	The experimental protocol 5 with training set (Mega) and the test set (Mega)	67
4.1	The Densenet201 graph when trained on the 40X magnification factor	78
4.2	A graph showing the outcome of the training of Densenet201 when trained on the magnification factor of 100X	78
4.3	The accuracy and loss curve for the Densenet201 when trained on the 200X magnification factor	79
4.4	Densenet201 when trained on the 400X magnification factor captured output graph	79
4.5	The graph pattern for the Densenet201 when trained on the Histopathology dataset	80
4.6	A graph showing the performance metrics for the Densenet201 when trained on the Mega dataset	80
4.7	Confusion matrices generated by the densenet201 network trained on the various magnification image dataset	81
4.8	The accuracy and loss curve for the Resnet50 when trained on the 40X magnification factor	82
4.9	The accuracy and loss curve for the Resnet50 when trained on the 100X magnification factor	82
4.10	The Resnet50 when trained on the 200X magnification factor displaying the accuracy and loss curve	83
4.11	The output of the Resnet50 training when trained on the 400X magnification factor	83
4.12	The result of the Resnet50 training when trained on the Histopathology dataset	84

4.13	The output graph for the Resnet50 training when trained on the Mega dataset	84
4.14	Confusion matrices generated by the Resnet50 network trained on the various magnification image dataset	85
4.15	The Resnet101 training output graph when trained on the 40X magnification factor	86
4.16	The training output of the Resnet101 when trained on the 100X magnification factor	86
4.17	The output graph of for Resnet101 model when trained on the 200X magnification factor	87
4.18	The accuracy and loss curve for the Resnet101 when trained on the 400X magnification factor	87
4.19	The result of the training for the Resnet101 when trained on the Histopathology dataset	88
4.20	The Resnet101 training output graph when trained on the Mega dataset	88
4.21	Confusion matrices generated by the Resnet101 network trained on the various magnification image dataset	89
4.22	The output the mobilenet-v2 training when trained on the 40X magnification factor	90
4.23	The output curve of the mobile-v2 when trained on the 100X magnification factor	90
4.24	The training curve for the mobile-v2 when trained on the 200X magnification factor	91
4.25	The training result of the mobile-v2 when trained on the 400X magnification factor	91
4.26	The output mobilenet-v2 model training when trained on the Histopathology dataset	92
4.27	The accuracy and loss curve for the mobilenet-v2 when trained on the mega dataset	92
4.28	Confusion matrices generated by the mobilenet-v2 network trained on the various magnification image dataset	93

LIST OF TABLES

TABLES	TITLE OF TABLES	PAGES
2.1	Break down of the Resnet 50 architecture	34
2.2	Break down of the Resnet 101 architecture showing details of the convolution layer as well as the output size and no of units	35
2.3	The brief summary of the DenseNet201 architecture	36
2.4	MobileNet-v2 model structure summary	38
2.5	Deep Learning Methods for Breast, Colon, Lung Cancer Classification and the research results	47
3.1	Distribution Data of BreakHis Dataset according to Subclasses and Magnifications	55
3.2	Distribution Data of Breast Histopathology Dataset according to Subclasses and Magnifications	59
3.3	The distribution of the dataset used in the experiment	68
3.4	Parameters Adopted in the CNN Networks for the experimental setup	69
3.5	Data Augmentation Parameters	73
4.1	Performance metrics of the various CNN models on the 40X magnified image	75
4.2	Performance metrics of the various CNN models on the 100X magnified image	76
4.3	Performance metrics of the various CNN models on the 200X magnified image	76
4.4	Performance metrics of the various CNN models on the 400X magnified image	76
4.5	Performance metrics of the various CNN models on the Histopathology dataset	77
4.6	Performance metrics of the various CNN models on the Mega dataset	77
4.7	Comparative performance analysis of the deep neural network	95
4.8	BreakHis and Breast Histopathology Image dataset benchmarked with other related works	96

LIST OF ABBREVIATIONS

A	Adenosis
ACC	Accuracy
AI	Artificial Intelligence
AUC	Area under the Curve
BreakHis	Breast Cancer Histopathological Image Classification
Breasthisto	Breast Cancer Histopathology Image
CAD	Computer-aided detection
CNN	Convolutional Neural Network
CT	Computed Tomography
DC	Ductal carcinoma
DenseNet	Densely connected network
DL	Deep Learning
DNN	Deep Neural Network
F	Fibroadenoma
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GPU	Graphical Processing Unit
LC	Lobular carcinoma
MC	Mucinous carcinoma
MIB	Magnification Independent Binary
MIM	Magnification independent Multiclass

ML	Machine learning
MRI	Magnetic Resonance Imaging
MSB	Magnification Specific Binary
MSM	Magnification Specific Multiclass
PC	Papillary carcinoma
PET	Positron emission tomography Cytopathology
PT	Phyllodes tumour
Relu	Rectified Linear Unit
Resnet	Residual neural network
RGB	Red-Green-Blue channel
ROC	Receiver Operating characteristics
SEN	Sensitivity
SGDM	Stochastic gradient decent
SPE	Specificity
SPECT	Single-photon emission computed tomography
SVM	Support Vector Machines
TA	Tubular adenoma
TN	True Negative
TP	True Positive
TPR	True Positive Rate
US	Ultrasound
VGG	Visual Geometry Group Network
WHO:	World Health Organization

WSI	Whole slide Image
X-ray	Radiography
XRM	X-ray mammography

GLOSSARY

Breast Cancer: is a disease condition that occurs as several cells in the breast tissue become cancerous, reproduce indiscriminately, and spreads, resulting in the formation of a tumour.

Benign: Non-cancerous conditions such as tumours and growths are described as benign. There is no spread to other sections of the body. Thus, it is contained. It does not spread to other tissues in the area. In medical terminology, the term benign denotes the absence of risk or seriousness of a condition.

CAD: A Computer-Aided Diagnosis is a tool, algorithm, or system whose output can help inform a user's decision-making or diagnosis.

Cancer: Cancer is a heterogeneous group of disorders that can begin in almost any organ or tissue of the body and spread throughout the body when precancerous cells grow spontaneously, invade adjacent parts of the body, and spread to other organs. Cancer is characterized by uncontrolled cell growth, invasion of adjacent parts of the body, and spreads throughout the body system.

CNN: is a variant of the learning network in (Artificial Intelligence) AI. For CNN to function, it must have a convolutional layer as its foundation. The convolution layer is located at the bottom of the stack. High-level aspects of the input signal are extracted using this technique. After the convolution layer comes the pooling layer. Depending on the applications, the pooling actions are predetermined. A variety of pooling operations are available, including maximum, minimum, and average pooling. Dimensionality reduction and selection of the most important feature are two common uses for the pooling method. It feeds these properties to the layer that includes activation functions, making it a fully integrated system.

Deep Learning: Features are learned using deep learning algorithms, which combine features from low levels of the hierarchy to produce higher ones. The potential is that a system can gain knowledge of multidimensional features linking inputs to outputs without relying only on handcrafted features as a result of this characteristics, it utilises the use of features that automatically identify at several abstraction levels.

Histopathology: Involves extracting a biopsy to study tissues and cells under a clinical microscope to diagnose and research disorders of the tissues. Studying tissues under a microscope to diagnose various diseases, including kidney, colon cancer, prostate cancer, and

breast cancer, is known as Histopathology. The pathologist or laboratory technologist performs this investigation by removing tissue via surgery, biopsies, or needle piecing.

Medical Imaging: various medical modality image output or results such as mammogram, endoscopy, and Histopathology.

Malignant: are abnormal cells that infiltrate and kill normal body cells are referred to as cancers that can spread to other parts of the body (metastasize) or infiltrate neighbouring tissues (locally). Due to genetic abnormalities, malignant cells develop rapidly and do not die naturally.

ABSTRACT

Among the different cancers that exist, breast cancer has been identified to account for about 2.26 million new cases among women globally in 2020 according to WHO. The early diagnosis of breast cancer can help reduce the mortality rate. Due to the large volume of breast cancer cases and the limited availability of histopathologists and clinicians, the available ones can be subjective which can lead to misjudgement. An intelligent system that can assist the limited histopathologist is crucial to help in optimal diagnosis. It has been identified that dataset usually came from one source, and a custom CNN was required to train. Therefore, this dissertation aims to employ Convolutional Neural Network models for accurate classification of breast cancer histopathology images curated from different dataset sources. This work utilised two datasets at different magnifications, the BreakHis and the Breast Histopathology dataset. A hybrid dataset was created from these two datasets and divided into 70% and 30% for training and testing. Four pre-trained Convolutional Neural Network (CNN) models (DenseNet201, ResNet50, ResNet101 and MobileNet-v2) were used for the analysis after preprocessing and rescaling. The findings show that DenseNet201 achieved the highest classification accuracy of 88.17%, 87.73%, 92.2% and 91.4% for BreakHis Dataset at 40X, 100X, 200X, 400X magnification factors respectively; 83.67% for Breast Histopathology Dataset at 200X and 85.78% for the Hybrid dataset at 200X. The models were able to classify the images between benign and malignant images, with DenseNet201 giving the best performance in terms of Specificity and Sensitivity at 100%. The implication is that the DenseNet201 model can be used to accurately differentiate between benign and malignant histopathology breast images thus serving as a decision support system in the early diagnosis of breast cancer.

Keywords: Convolutional Neural Network (CNN), histopathology images, breast cancer, image classification, intelligent system, diagnosis.