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

LIST OF FIGURES

LIST OF TABLES

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.