

LB/TH/41/2025
TH6004

**EXPLAINABLE AI FOR BREAST CANCER
DETECTION IN MAMMOGRAPHY**

L.L.M. Wickremesinghe

229406T

Degree of Master of Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

January 2025

EXPLAINABLE AI FOR BREAST CANCER DETECTION IN MAMMOGRAPHY

L.L.M. Wickremesinghe

229406T

Dissertation submitted in partial fulfillment of the requirements for the
degree
Degree of Master of Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

January 2025

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 30-06-2025

The supervisor should certify the Dissertation with the following declaration.

The above candidate has carried out research for the Degree of Master of Computer Science Dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Thanuja D. Ambegoda

Signature of the Supervisor:

Date: 30-06-2025

ACKNOWLEDGEMENT

I am grateful to Dr. Thanuja D. Ambegoda for his invaluable assistance in identifying an engaging research topic and for his unwavering support and encouragement. His guidance was pivotal in setting goals and enhancing my study engagement. I extend my heartfelt thanks to the Department of Computer Science and Engineering at the University of Moratuwa for their assistance in overcoming challenges. Lastly, I want to express deep appreciation to my family for their steadfast support throughout this journey.

ABSTRACT

Breast cancer remains a significant global health concern among women. This research introduces an explainable AI-assisted breast cancer detection system aimed at improving both the accuracy and interpretability of mammogram-based diagnoses. The study utilizes high-quality mammographic datasets, CBIS-DDSM and the RSNA Screening Mammography dataset, to train and validate the models.

The system uses two powerful deep learning models: Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The InceptionResNetV2 CNN achieved an accuracy of 92%, while the ViT model reached 96% accuracy by effectively focusing on important regions in the mammogram images. To make the system more transparent, several Explainable AI (XAI) methods were applied, including Grad-CAM, SIDU, Attention Maps, and Ablation-CAM. Among these, SIDU provided the clearest and most accurate visual explanations, which are valuable for medical decision-making.

To further improve the reliability and clinical value of the system, this study introduces a Dual-Stage Ensemble Diagnosis and Decision Fusion Framework. This approach combines the diagnostic strengths of both models to deliver a more confident and balanced final decision, supported by detailed visual explanations. The platform consists with a user-friendly web application that allows doctors and patients to easily upload mammogram images and receive AI-based predictions with clear and interpretable outputs. This research helps advance the development of trustworthy AI tools for breast cancer detection in real clinical settings.

Keywords: Explainable AI(XAI), Breast Cancer, Convolutional Neural Networks(CNN), Vision Transformers (ViT), Mammogram, Medical Imaging

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Acknowledgement	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
1 Introduction	1
1.1 Motivation and Background	1
1.2 Problem Statement	7
1.3 Research Gap	9
1.4 Research Objectives	10
2 Literature Review	12
2.1 Breast Cancer and Screening Methods	12
2.2 Evaluation of Deep Learning Techniques for Breast Cancer Detection Using Mammography	14
2.3 Explainable Artificial Intelligence (XAI) Methods	21
3 Methodology	28
3.1 System Architecture	28
3.2 Dataset Description	28
3.3 Data pre-Processing	30
3.4 Feature Extraction and Training Framework	33
3.4.1 Feature Extraction	33
3.4.2 Training Configuration Settings	33
3.4.3 Model Setups	34
3.5 Experiment Training and Model Selection	38
3.6 Transformer Model-Building	41
3.7 Validation and Testing	42
3.8 Applying XAI Technique	43

3.8.1	Applying XAI Techniques for Vision Transformers (ViT) Model	43
3.8.2	Applying XAI Techniques for CNN-based Model	44
3.9	Interpretability Analysis	51
3.10	Dual-Stage Ensemble Diagnosis and Decision Framework	52
3.11	Integration with Medical System	57
3.12	Tools and Technologies	58
4	Results and Discussion	60
4.1	Results and Discussion for CNN Architecture	60
4.1.1	Results of Experiment 01	60
4.1.2	Results of Experiment 02	62
4.2	Comparison of Experiment 01 & 02	65
4.3	Results and Discussion for ViT Architecture	67
4.4	Results and Discussion for ViT Based XAI Applicaton	70
4.5	Results and Discussion for CNN Based XAI Applicaton	74
4.6	Discussion of the Integration with Medical System	79
5	Conclusion	83
5.1	Limitations and Future Work	85
	References	87

LIST OF FIGURES

Figure	Description	Page
Figure 1.1	Bar chart presents comparing age-standardized incidence and mortality rates (per 100,000 person-years) for the 15 most common cancer types in 2018, contrasting countries with high or very high Human Development Index (HDI) against those with low or medium HDI, separately for women (top) and men (bottom).	1
Figure 1.2	Mammography-based CAD system Process	4
Figure 1.3	AI vs XAI	8
Figure 2.1	Mammographic X-Ray of Human Breast	13
Figure 2.2	ML and DL Approach	15
Figure 3.1	System Architecture of the research	29
Figure 3.2	A Mammogram image after applied transformations	31
Figure 3.3	General structure of the VGG16 network.	35
Figure 3.4	General structure of the VGG19 network.	35
Figure 3.5	General structure of the DenseNet Architecture	38
Figure 3.6	vit_b_16 Model	41
Figure 3.7	Workflow of Attention Map Generation for ViT Model	45
Figure 3.8	Workflow of the Generation of Feature Activation Image Masks Step	48
Figure 3.9	Workflow of the Computing Feature Importance Weights Method	49
Figure 3.10	Visual comparison of saliency maps generated from Grad-CAM, Ablation-CAM and SIDU for cancerous images on InceptionResNetV2 model.	50
Figure 3.11	Conceptual Work-flow Diagram for Dual-Stage Ensemble Diagnosis	56
Figure 4.1	Classification Report for Experiment 02 InceptionResNetV2 Model: Demonstrates the model's precision, recall, and F1-scores across both classes, highlighting strengths and weaknesses in cancer detection.	62
Figure 4.2	Confusion Matrix for Experiment 02 InceptionResNetV2 Model: displays the distribution of true positives, true negatives, false positives, and false negatives, emphasizing areas for improvement in misclassification reduction.	63
Figure 4.3	Classification Loss Plot for Experiment 02 InceptionResNetV2 Model: Illustrates the reduction in model loss during training and validation, showing stable learning progression with minor fluctuations	64
Figure 4.4	Classification Accuracy Plot for Experiment 02 InceptionResNetV2 Model: Shows the progression of training and validation accuracy, indicating improved performance with minimal risk of overfitting.	64

Figure 4.5	Classification Report for <code>vit_b_16</code> Model: Highlights ViT’s balanced precision, recall, and F1-scores, demonstrating consistent performance in classifying cancerous and non-cancerous cases.	68
Figure 4.6	Confusion Matrix for <code>vit_b_16</code> Model: Visualizes the ViT model’s ability to minimize false positives and negatives, showcasing accurate classification	68
Figure 4.7	Classification Accuracy Plot <code>vit_b_16</code> Model: Depicts consistent training and validation accuracy trends, reinforcing the model’s stability.	69
Figure 4.8	Classification Loss Plot <code>vit_b_16</code> Model: Displays stable loss reduction during training, confirming ViT’s effective convergence with minimal overfitting.	70
Figure 4.9	ViT model-generated attention maps, showing focused regions in cancerous mammogram images for improved model interpretability	71
Figure 4.10	ViT model-generated attention maps overlaid on non-cancerous mammogram images, highlighting key non-cancer regions.	72
Figure 4.11	Grad-CAM Results — Illustrates the Grad-CAM-generated heatmaps over mammogram images, showing important regions the Inception-ResNetV2 model focused on for both cancerous and non-cancerous predictions.	74
Figure 4.12	Ablation-CAM Results — Displays Ablation-CAM-generated heatmaps that emphasize key visual patterns influencing the InceptionResNetV2 model’s predictions.	75
Figure 4.13	SIDU Results — Demonstrates SIDU-generated saliency maps, emphasizing high-accuracy feature selection both cancerous images for improved transparency in AI predictions.	77
Figure 4.14	SIDU Results — Demonstrates SIDU-generated saliency maps, emphasizing high-accuracy feature selection in non-cancerous cases for improved transparency in AI predictions.	78
Figure 4.15	Clinical Validation Output: High-Risk Case	79
Figure 4.16	Clinical Validation Output: Low-Risk Case	80