

ResViT: A Transformer–CNN-based Hybrid Model for Robust Breast Cancer Detection

By
Faiaz Bin Rhaman
211-15-4016

FINAL YEAR DESIGN PROJECT REPORT

**This Report Presented in Partial Fulfillment of the
Requirements for the Degree of Bachelor of Science in
Computer Science and Engineering**

Supervised by

Mr. Md. Asaduzzaman
Assistant Professor
Department of Computer Science and
Engineering, Daffodil International
University

Co-Supervised by

Mr. Mohammad Jahangir Alam
Assistant Professor
Department of Computer Science and
Engineering, Daffodil International
University



**DAFFODIL INTERNATIONAL
UNIVERSITY**


Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled “ResViT: A Transformer–CNN based Hybrid Model for Robust Breast Cancer Detection”, submitted by Faiaz Bin Rhaman, ID No: 211-15-4016 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17 September, 2025.

BOARD OF EXAMINERS


17.09.2025

Dr. Fernaz Narin Nur
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



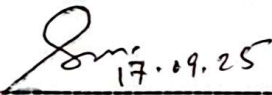
Dr. Abdus Sattar
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner


17.09.25

Mohammad Jahangir Alam
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner


17.09.25

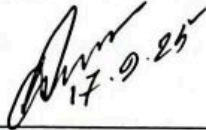
Dr. Sajeeb Saha
Associate Professor
Department of Computer Science and Engineering
Jagannath University

External Examiner

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Mr. Md. Asaduzzaman**, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:



Mr. Md. Asaduzzaman

Assistant Professor

Department of Computer Science and
Engineering, Daffodil International
University

Co-Supervised by:



Mr. Mohammad Jahangir Alam

Assistant Professor

Department of Computer Science and
Engineering, Daffodil International
University

Submitted by:



Faiaz Bin Rhaman

Student ID: 211-15-4016

Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. I am deeply grateful to everyone who has assisted me in one way or another.

First, I express my heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for me to complete the **Final Year Design Project (FYDP)** successfully.

I am grateful and wish my profound indebtedness to **Mr. Md. Asaduzzaman, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of my supervisor in the field of **Computer Vision** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartfelt gratitude to the Head of the Department of Computer Science and Engineering for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

I would like to thank my entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

Breast cancer remains one of the leading causes of mortality among women worldwide. Early and accurate detection is essential for effective treatment and improved survival rates. This thesis presents a robust and efficient deep learning-based breast cancer detection system by integrating Vision Transformer (ViT), ResNet, and ResViT models into a hybrid ensemble framework. While CNNs like ResNet are effective in capturing local image features, they often fail to represent long-range dependencies. Conversely, ViTs are capable of capturing both local and global features but are computationally expensive. The proposed ensemble combines the strengths of these architectures to improve classification accuracy while maintaining deployment efficiency. The dataset undergoes preprocessing steps such as DPI adjustment, resizing, normalization, and contrast enhancement to improve model input quality. To address class imbalance, data augmentation and class-weighted loss functions are applied. The trained model is converted to TensorFlow Lite format and deployed in a Flutter-based mobile application, enabling real-time, offline diagnosis. Experimental results show that ResNet achieved 82% test accuracy and 0.90 AUC, ViT reached 94% test accuracy with 0.94 AUC, and the ResViT model outperformed all with 97% test accuracy and 0.99 AUC. These findings highlight the effectiveness of the proposed hybrid model in breast cancer classification tasks. By combining high performance with mobile accessibility, the system offers a practical and scalable solution for early detection in low-resource clinical environments, contributing significantly to mobile health innovation.

Table of Contents

Approval	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Introduction	1
1.2 Motivation.....	3
1.3 Objectives.....	4
1.4 Methodology.....	5
1.5 Project Outcome	6
1.6 Organization of the Report	7
2 Background	9
2.1 Introduction	9
2.2 Literature Review	10
2.2.1 Similar Applications	15
2.2.2 Related Research.....	16
2.3 Gap Analysis.....	18
2.4 Summary.....	20
3 Research Methodology	21
3.1 Methodology/Requirement Analysis & Design Specification	21
3.1.1 Overview.....	21
3.1.2 Proposed Methodology/ System Design.....	22
3.1.3 Functional and Nonfunctional Requirements	23
3.1.4 Data Flow Diagram Level 1	24
3.1.5 UI Design	25
3.2 Detailed Methodology and Design.....	27

3.3	Project Plan	34
3.4	Task Allocation.....	36
3.5	Summary.....	37
4	Implementation and Results	38
4.1	Environment Setup.....	38
4.2	Testing and Evaluation/Performance/ Comparative Analysis ..	40
4.3	Results and Discussion	42
4.4	Summary.....	55
5	Engineering Standards and Design Challenges	56
5.1	Compliance with the Standards.....	56
5.1.1	Software Standards.....	57
5.1.2	Hardware Standards	57
5.1.3	Communication Standards	57
5.2	Impact on Society, Environment and Sustainability	58
5.2.1	Impact on Life.....	58
5.2.2	Impact on Society & Environment	59
5.2.3	Ethical Aspects.....	60
5.2.4	Sustainability Plan	61
5.3	Project Management and Financial Analysis	62
5.4	Complex Engineering Problem.....	63
5.4.1	Complex Problem Solving.....	64
5.4.2	Engineering Activities	64
5.5	Summary.....	65
6	Conclusion	67
6.1	Summary.....	67
6.2	Limitation	68
6.3	Future Work	68
	References	69

List of Figures

3.1	The Methodological Flowchart	22
3.2	Data Flow Diagram level 1	24
3.3	The Two-Tier Architecture and UI for Application	27
3.4	Train-Validation-Test Splitting	28
3.5	Sample Image of Each Class	28
3.6	The Architecture of ResNet50	31
3.7	The Architecture of Vision Transformer Model	33
3.8	The proposed Architecture of the Res-ViT Hybrid Model	34
4.1	Loss and accuracy curve of the ResNet50 model over 80 epochs ...	42
4.2	Confusion matrix on the validation set of the ResNet50 model	43
4.3	Confusion matrix on the test set of the ResNet50 model	45
4.4	The ROC-AUC score for each class of the ResNet50 model	46
4.5	Loss and accuracy curve of the ViT model over 80 epochs	47
4.6	Confusion matrix on the validation set of the ViT model	47
4.7	Confusion matrix on the test set of the ViT model	49
4.8	The ROC curve and AUC score for each class of the ViT model	50
4.9	Loss and accuracy curve of the ResViT model over 80 epochs (early stopped at epoch 30)	51
4.10	Confusion matrix on the validation set of the ResViT model	52
4.11	Confusion matrix on the test set of the ResViT model	53
4.12	The ROC curve and AUC score for each class of the ResViT model	54

List of Tables

2.1	Research Matrix for Breast Cancer Detection Techniques.....	14
2.2	Similar Applications – Research Matrix.....	16
2.3	Related Research – Matrix Table	17
2.4	Gap Analysis Table – Existing Research vs. Proposed Work	19
3.1	Dataset Specifications	29
3.2	Preprocessing Properties and Description	29
3.3	GANTT Chart of Project Timeline	35
3.4	Task Allocation Table.....	36
4.1	Common parameter table for all experimented ViT models.....	39
4.2	Common data split for all experimented ViT models.....	40
4.3	Classification report on the validation set of the ResNet50 model	44
4.4	Classification report on the test set of the ResNet50 model.....	45
4.5	Classification report on the validation set of the ViT model	48
4.6	Classification report on the test set of the ViT model.....	49
4.7	Classification report on the validation set of the ResViT model	52
4.8	Classification report on the test set of the ResViT model.....	53
4.9	Performance analysis among all the experimented ViT models	55
5.1	Details of tools and platform used.....	63
5.2	Estimated Cost and Financial Analysis.....	63
5.3	Mapping with Complex Problem Solving.....	64
5.4	Mapping with Knowledge Profile	65
5.5	Mapping with Complex Engineering Activities.....	65

Chapter 1

Introduction

This chapter introduces the foundation of the research by presenting the background, motivation, and significance of early breast cancer detection using deep learning. It outlines the main objectives of the study, describes the proposed hybrid methodology combining ViT, ResNet, and ResViT, and highlights the expected outcomes. The chapter also provides an organized structure of the entire report, setting the stage for the technical and analytical discussions in the following chapters.

1.1 Introduction

Breast cancer is the most common cancer in women all over the world and is the greatest cause of cancer death. As the incidence of the disease all over the world has been increasing and breast cancer is currently the most diagnosed cancer in women, there has been an urgent need to detect this disease early on in order to increase the survival rates and the outcomes of the treatment [1]. The use of advanced screening techniques, mammography, ultrasound and histopathology images are more of central role in early disease detection, but they need extensive adjustments to accuracy and diagnostic efficacy because there are usually frequent limitations such as high false positive rates with dense breast tissue and variation among radiologists [2-3]. Machine learning (ML) and Deep Learning (DL) methods have become exceptionally useful in improving the rate of breast cancer detection over the last few years. The original studies were on the conventional ML algorithms, including k-Nearest Neighbors (kNN), Support Vector Machines (SVM), and Random Forest, which employ the structured datasets, i.e., the Wisconsin Breast Cancer Dataset [4-5]. These models are praised however they are limited in terms of feature extraction, hence their ability to analyze high-dimensional data used in medical images is limited [6-7].

With the increased complexity of medical imaging, deep learning became a sophisticated solution, and it can be motivated by Convolutional Neural

Networks (CNNs) due to the capacity to identify complex patterns within images. CNN models have been found to be significantly improved in image classification, particularly when they are combined with feature selection methods including Recursive Feature Elimination [6]. However, CNNs predominantly focus on local features, which, while effective, lack the ability to analyze complex spatial relationships in images comprehensively [8]. Vision Transformers (ViTs), introduced more recently, address this limitation by capturing both local and global spatial information, allowing for better feature representation, especially in densely textured breast tissue. Studies show that ViT models, as well as hybrid architectures that combine CNNs and ViTs, consistently outperform traditional models, achieving higher accuracy in histopathological image classification tasks [9-10].

To further enhance breast cancer diagnostic accuracy, recent studies have explored the integration of ensemble and hybrid models, combining CNNs, ViTs, and Generative Adversarial Networks (GANs). These hybrid models aim to leverage the strengths of each component, yielding models that are both accurate and computationally efficient [2][8]. Despite the promising results, the computational demands of ViTs, which are known to be resource-intensive, continue to limit their applicability in real-time diagnostic contexts, especially in resource-constrained environments [11-12].

While existing studies demonstrate the potential of hybrid models, several key gaps remain unaddressed:

- The majority of datasets suffer from class imbalances, where malignant cases are underrepresented, leading to biased model predictions. Existing resampling techniques provide partial solutions but do not fully eliminate the issue [13-14].
- ViTs and hybrid models, while highly accurate, require significant computational power, making real-time deployment difficult, particularly in resource-constrained environments. Efficient model optimization techniques are needed.
- Although CNN-ViT hybrids have shown improvements, there is still a lack

of structured frameworks integrating multiple architectures like ResNet, ViT, and ResViT in a unified ensemble model for robust performance.

To address these challenges, this study makes the following contributions:

- This study introduces a novel ensemble approach combining ViT, ResNet, and ResViT to leverage both CNN-based feature extraction and transformer-based spatial attention mechanisms, improving breast cancer classification accuracy.
- By incorporating optimized feature selection and model fusion strategies, this study enhances computational efficiency, making the proposed approach more suitable for real-time diagnostic applications.
- This study implements data augmentation, and class-weighted loss functions to mitigate the impact of class imbalance in breast cancer datasets. Additionally, different preprocessing techniques such as noise reduction, and contrast enhancement are applied to improve image quality, ensuring more robust and unbiased feature extraction for accurate classification.

While ML and DL techniques have significantly advanced breast cancer detection capabilities, challenges such as class imbalance, dataset limitations, and computational demands remain. To overcome these shortcomings, further code creation of hybrid models with effective CNNs and transformers fusion, as well as to diversify datasets and lower computational expenses, in particular, in real-time use, is needed. ViT-ResNet-ResViT ensemble model is proposed as a solution to these gaps and provide a more accurate, efficient, and generalizable way of detecting breast cancer.

1.2 Motivation

Breast cancer is a type of cancer that is very widespread and deadly to all women in the world. Early diagnosis is highly significant as it aids in an increase in the survival rates and the efficiency of the treatment. Despite the various medical imaging technologies such as histopathology and mammography being utilized in diagnosing the condition, the modalities continue to experience problems of false findings and disagreement among the physicians.

Deep learning, particularly Convolutional Neural Network (CNNs), has proved to be very promising in the analysis of medical pictures. CNNs are effective in detecting detailed patterns, however, generally, they pay attention to small areas of the picture. This is a constraint on their capacity to comprehend the entire structure of complicated tissue samples. Vision Transformers (ViTs) on the other hand are more effective in terms of capturing both local and global information and this can assist in the detection of cancer cells in histopathology images. ViTs are however generally more demanding in terms of computing power which makes it hard to apply in real-time or in a mobile environment.

To address these problems, this paper is inspired to create a hybrid model through a combination of ResNet (CNN), ViT (Transformer), and ResViT (CNN-Transformer) into a single ensemble. The objective of this model is to enhance precision and minimize the cost of computation. The trained model will be optimized to a lightweight format (.tflite) and integrated into a mobile application based on Flutter to be used in real-time. The system proves to be particularly helpful in distant regions or hospitals with a limited number of resources as the availability of fast and reliable diagnostic tools is limited.

1.3 Objectives

The main aim of this study is to enhance the precision, effectiveness, and feasibility of breast cancer detection in terms of histopathological images. Although conventional CNN models and machine learning methods have had important contributions to the field, they do not always have the capacity to learn intricate spatial arrangements or can be practiced in real-time systems. ViTs are more useful in extracting global features, but their application is constrained by high computational costs. To solve these problems, the present study suggests a hybrid and ensemble architecture that incorporates CNNs and Transformers in order to achieve powerful and real-time diagnosis.

The specific objectives of this study are:

- To create a hybrid ensemble model combining the strengths of ResNet (CNN), Vision Transformer (ViT) and ResViT (CNN-Transformer fusion)

models to improve feature extraction, classifier accuracy, and offer a more detailed insight into histopathological image patterns.

- To enhance the quality of the data and overcome the problem of class imbalance with the help of a set of preprocessing methods such as DPI correction, normalization, contrast-enhancement methods, and image-resizing techniques and more advanced methods such as data augmentation and class-weighted loss functions to make the learning process balanced across the various types of cancer.
- In order to streamline the proposed model to be used in real-time and convert the trained end ensemble into a smaller format (.tflite) to support mobile devices and low resource settings, without significantly affecting the accuracy of some diagnostic tasks.
- To create a convenient mobile application in Flutter, which enables medical personnel and healthcare professionals to identify and categorize breast cancer on histopathological images immediately and directly, thereby providing a solution to the problem of diagnosing breast cancer early and remotely or in underserved locations.

1.4 Methodology

This paper adheres to a systematic deep learning approach of precise and effective detection of breast cancer by use of histopathological images. The proposed solution incorporates the advantage of both Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) and combines the ResNet, ViT and ResViT architecture in a hybrid ensemble model. It aims at creating a good diagnostic system that is not only effective in classification but is light enough to do real-time on a mobile platform.

The methodology is divided into the following key phases:

Image Preprocessing: In order to guarantee uniform and good input to the model, a number of steps are pre-processed. These include DPI (Dots Per Inch) adjustment, resizing all images to a uniform size, and normalization to scale pixel values. Contrast enhancement is also used to improve the visibility of histological features. These steps help the model learn more effectively from the

image data.

Model Design and Training: Three deep learning architectures are used. ResNet, a CNN-based model known for efficient feature extraction using residual blocks. ViT (Vision Transformer), which captures both local and global spatial features using self-attention mechanisms. ResViT, a hybrid model that combines CNNs and ViTs for enhanced representational power. These models are individually trained and then fused into an ensemble using a weighted averaging technique, improving robustness and generalization.

Handling Class Imbalance: The dataset often contains uneven distribution among different classes (e.g., fewer malignant samples), which can lead to biased predictions. To address this, techniques like data augmentation and class-weighted loss functions are applied. Augmentation artificially increases the number of samples per class, while class-weighting ensures the model pays equal attention to all classes during training.

Model Optimization and Deployment: Once the ensemble model is trained, it is converted to TensorFlow Lite (.tflite) format to reduce file size and computational demand. This makes the model suitable for mobile environments. The optimized model is then deployed into a real-time Flutter-based mobile application, allowing healthcare workers to use the system on smartphones for quick and accessible breast cancer screening.

This multi-stage methodology ensures that the proposed system is not only accurate but also practical and deployable in real-world scenarios, especially in remote or under-resourced medical settings.

1.5 Project Outcome

This research aims to develop a reliable and efficient breast cancer detection system using a hybrid deep learning framework. The expected outcomes of this study are as follows:

- Development of a robust hybrid ensemble model that combines ResNet, Vision Transformer (ViT), and ResViT architectures. This model will achieve improved classification accuracy by effectively extracting both local and global features from histopathological images.
- A well-preprocessed and balanced dataset with enhanced image quality, addressing issues like class imbalance, noise, and uneven illumination through preprocessing techniques and augmentation strategies.
- A lightweight and optimized deep learning model converted into TensorFlow Lite (.tflite) format, which significantly reduces computational load and enables the system to be used in real-time on mobile devices.
- A Flutter-based mobile application that integrates the trained model to allow real-time breast cancer detection directly from histopathological images, making the solution accessible, fast, and practical for healthcare professionals, especially in low-resource or rural areas.

These outcomes will contribute to both academic research and real-world healthcare by offering a scalable, portable, and high-performing solution for early breast cancer diagnosis.

1.6 Organization of the Report

This report is organized into six chapters, each focusing on a specific component of the research work. The structure of the report is as follows:

Chapter 1: Introduction

This chapter presents the background of the study, the motivation behind the research, the main objectives, and the proposed methodology. It also highlights the expected outcomes and provides a summary of how the report is structured.

Chapter 2: Background

This chapter provides an overview of the existing techniques in breast cancer detection using machine learning and deep learning. It includes a detailed literature review of related works focusing on CNN, ViT, hybrid models, and ensemble techniques. The gap analysis section highlights the research limitations in current studies and explains the need for the proposed solution.

Chapter 3: Research Methodology

This chapter explains the overall system design and technical requirements of the project. It presents the proposed methodology in detail, including preprocessing, model training, ensemble integration, and model optimization. It also includes system diagrams such as the context diagram, data flow diagram, UI design, and a detailed project plan with task allocation.

Chapter 4: Implementation and Results

This chapter focuses on the implementation of the proposed ViT-ResNet-ResViT ensemble model. It explains the development environment setup, model training process, evaluation metrics, and a detailed performance analysis. Results are discussed through comparative evaluation to demonstrate the effectiveness of the proposed system.

Chapter 5: Engineering Standards and Design Challenges

This chapter discusses the engineering considerations of the project, including compliance with software and hardware standards, and the impact of the proposed system on society, environment, and sustainability. It also covers ethical concerns, sustainability planning, financial analysis, and the challenges faced in solving a complex engineering problem.

Chapter 6: Conclusion

The final chapter summarizes the key findings of the research, discusses existing limitations of the current work, and outlines possible directions for future improvement and development of the system.

Chapter 2

Background

The chapter begins with an in-depth literature review, divided into two key areas: Similar Applications that use traditional ML approaches, and Related Research that explores deep learning, Vision Transformers, hybrid models, and ensemble learning for breast cancer detection. Following the review, a gap analysis is conducted to identify the limitations of current studies, including lack of unified hybrid models, poor handling of class imbalance, and the absence of mobile-deployable solutions. The chapter concludes by establishing the motivation for the proposed ViT-ResNet-ResViT ensemble model and Flutter-based mobile system.

2.1 Introduction

Breast cancer is a major global health concern and one of the leading causes of cancer-related deaths among women. Since the incidence has increased in all parts of the world, early and accurate diagnosis is a paramount objective in the medical circles. The common screening techniques that are used in diagnosing breast cancer include mammography, ultrasound, and histopathological imaging. Nonetheless, such methods are mostly associated with drawbacks such as reduced sensitivity in dense breast tissue, inter-observer error, high false positives or false negatives and all these can result in delayed treatment or unnecessary intervention.

To overcome such difficulties, machines learning (ML) and deep learning (DL) methods have been actively studied by researchers, as this provides automated, precise, and data-driven methods of detecting cancer. The first activities were on classical ML classifiers, including Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Decision Trees and Random Forests and mostly trained on structured data like the Wisconsin Breast Cancer Dataset. These models proved useful in classification tasks such that their performance was limited due to the manual feature engineering and the inability to model the complicated spatial

patterns of medical images.

The invention of Convolutional Neural Networks (CNNs) represented a tremendous breakthrough to the analysis of medical images, especially in the classification of breast cancer. CNNs automatically obtain deep features and spatial data of histopathological images with great accuracy compared to the traditional ML models. Nevertheless, CNNs are locally oriented and cannot identify long-range dependencies in the image, which is essential to the interpretation of the intricate structure of cancerous tissue.

Vision Transformers (ViTs) have recently received attention to address this limitation. ViTs can represent global and local features via self-attention mechanisms, and hence they can be used effectively in medical imaging. Several studies have proved the efficiency of ViTs and ViT-based hybrid models in enhancing the accuracy of diagnosis of breast cancer. Nevertheless, they are frequently computationally costly and not optimized to be deployed in real time or for a mobile system.

In light of this, hybrid models that combine CNNs and ViTs have emerged as a promising solution, offering the strengths of both architectures. Ensemble techniques have further been applied to improve robustness and generalization. Yet, there remains a lack of structured and lightweight ensemble frameworks that integrate multiple advanced models like ResNet, ViT, and ResViT into a single system optimized for real-world deployment.

This chapter presents a detailed review of existing literature, focusing on similar ML applications and related deep learning research in breast cancer detection. It also identifies critical research gaps, particularly in the areas of model generalization, efficiency, and real-time deployment, which are directly addressed by the proposed ViT-ResNet-ResViT ensemble model integrated into a Flutter-based mobile application.

2.2 Literature Review

Breast cancer, the most prevalent cancer among Indian women and one with a

50% fatality rate in diagnosed cases, has driven research efforts toward improved predictive and diagnostic models. Sharma [4] focused on comparing popular machine learning (ML) algorithms, such as Random Forest, kNN, and Naïve Bayes, using the Wisconsin Diagnosis Breast Cancer dataset. Their analysis, emphasizing accuracy and precision metrics, showed competitive results in detection, thereby supporting these techniques as beneficial in early diagnosis and treatment. Expanding on this, Mohammed [13] analyzed classifiers including Decision Tree, Naïve Bayes, and Sequential Minimal Optimization on benchmark datasets, addressing class imbalance through resampling methods that improved performance. Their findings indicated that SMO excelled on the WBC dataset, while J48 performed best on a Breast Cancer dataset, underlining the importance of tailored approaches to data diversity and imbalance in predictive modeling.

Vaka [3] proposed a novel ML-based approach, achieving high accuracy and efficiency in experimental analyses, which demonstrates the potential of tailored ML techniques to enhance the detection of early-stage breast cancer. Breast cancer is a global health issue as it kills more than 10 million people annually; furthermore, the incidences of the disease are expected to rise in the next several decades. Abunasser [15] performed a systematic review, and the following techniques were defined as the best methods of breast cancer detection: Artificial Neural Networks (ANN), Support Vector Machines (SVM) and k-Nearest Neighbor (kNN). ANN and SVM particularly stood out, with a range of accuracy of 83.45 to 99.30 being obtained in 32 studies that were reviewed.

Amethiya [16] investigated the biosensors-ML technique integration, as well as evaluating a wide range of algorithms, among them ELM-RBF, SVR, RVM, and ANN. Their comparative review focused on accuracy, precision and recall when using datasets of different imaging and biosensor sources. On the same note, Agarap [17], compared six ML algorithms using the wdbc dataset with MLP recording the highest accuracy of. which validates the strength of particular architectures in the context of mammographic data analysis to classify a tumor.

The use of supervised learning model is also seen in the research by Alarabeyyat

and Alhanahnah [18], which proposed a system of processing and extracting features to identify breast cancer with a high accuracy. Their analysis compared two models, Back Propagation Neural Network (BPNN) and Logistic Regression (LR), and their results indicate that LR was more beneficial than BPNN when using a large number of features and this fact supports the importance of model selection in depending on the complexity of the features used and the specifics of the data set.

The prevalence of breast cancer in the UAE highlights the importance of effective and reliable diagnostic procedures in the initial stages. Bazazeh and Shubair [5] made a comparison between SVM, Random Forest and Bayesian Networks as the main metrics of model efficacy included accuracy and ROC area. Their results offer a useful comparison of old ML techniques to practitioners who are interested in having powerful algorithms in clinical diagnostics. Ak [7] investigated the use of visualization alongside logistic regression and other ML-based visualization tools and obtained the accuracy of the classification of 98.1%. The paper has highlighted the promise of visual analytics and synergy between visual analytics and ML in improving the quality of diagnosis and clinical decision-making. Sindhvani [19] paid attention to the application of ML to enhance the prognosis and recovery of breast cancer, noting that the method allows identifying the necessary characteristics in large data volumes and thereby minimizes the number of needless treatment. Their approach included the simplest development of ML models with the help of Python, which proves the efficiency of pattern recognition and predictive algorithms in improving the quality and confidence of breast cancer detection.

Bhise [6] examined CNN-based model that included Recursive Feature Elimination (RFE) to optimize the selection of features. As they comparatively studied using classifiers such as SVM, random forest and Naive Bayes on the BreakHis 400X data, the accuracy and precision of CNN in breast cancer detection was identified, indicating the effectiveness of CNN as a medical image data classifier. Making the same argument, Mangukiya [14] compared seven ML algorithms, such as Adaboost and XGBoost, on the Wisconsin Breast Cancer Dataset. Their results showed XGBoost got a significant accuracy of 98.24, and

the lowest error rate of all models, which highlighted its strength and the accuracy of cancer categorization. Stating the drawbacks of mammographic data which is subject to false positives, Tahmooreesi [11] introduced a hybrid ML model based on SVM, ANN, KNN and Decision Trees. This model was applied in various data types demonstrating how the model can be adjusted to various data, a strategy that is useful in achieving reliable breast cancer detection not based on conventional imaging strategies.

Research studies of the recent years have also been concentrated on image quality and data diversity to improve better classification. Rautela et al. [2] introduced a distinctive model with the combination of GANs and vision transformers (ViTs) to enhance image resolution, where a two-channel GAN model is used to enhance GAN strength and efficiency. This model with high sensitivity and accuracy (0.932 and 0.988, respectively) on the INbreast dataset has shown the importance of the high-resolution image generation in the breast cancer detection. The role of ViT in the medical image analysis is being developed by other studies. Hassan [8] proposed a mass detection CAD system based on YOLOv4 and ViT on CESM images with a mean Average Precision (mAP) of 98.69 on INbreast, indicating that ViT can detect intricate spatial features to detect a mass and classify it. On the same note, Alotaibi [10] discussed an ensemble model that combines ViT with Data-Efficient Image Transformers (DeiT) and the accuracy was very impressive at 98.17 per cent on the BreakHis dataset. Their results confirm the possibility of ViT-based ensembles to classify histopathology images with high accuracy in order to enhance the accuracy of the diagnosis. Additionally in favor of ViT, Abimouloud [9] assessed lightweight hybrid ViT-CNN networks, and it was shown that they are able to address the complexity of the data with reduced computational load on the BreakHis dataset with accuracy up to 98.64 percent. The paper highlights the potential of hybrid models to be used as a possible method of breast cancer classification by providing computationally-compromised alternatives to standard CNN-heavy models.

Also, ensemble models have become the subject of interest in improving the accuracy of diagnosis. Desai [20] contrasted CNN and MLP in detecting

malignancy in the breast cells and discovered that CNN was more effective than MLP, yet they proposed that more research should be done under similar conditions in order to consolidate these findings. Similarly, Melekoodappattu [21] evolved an experienced ensemble procedure of CNN and texture-based feature extraction with a maximum accuracy of 98.3 on DDSM data. Their approach leverages Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction, thereby improving classification accuracy through enhanced feature interpretability. Sahu [12] contributed to ensemble model research with a hybrid CNN-based framework that combines ShuffleNet and ResNet for breast cancer detection on mammogram and ultrasound datasets. Their work achieved accuracy rates up to 99.17% for DDSM data, highlighting the advantages of hybrid models in delivering consistent results even on limited datasets. Ahad [22] assessed D-CNN architectures, transfer learning, and ensemble models, with the latter achieving the highest accuracy (99.94%) for breast cancer classification. However, transfer learning did not improve accuracy across models, suggesting a limited role in certain medical imaging applications.

Table 2.1: Research Matrix for Breast Cancer Detection Techniques

Author (Year)	Models Used	Accuracy (%)	Dataset Info
Sharma et al. (2018)	Random Forest, kNN, Naïve Bayes	Varied (Competitive)	Wisconsin Diagnosis Breast Cancer
Mohammed et al. (2020)	Decision Tree, Naïve Bayes, SMO	Varied	Wisconsin Breast Cancer (WBC), Breast Cancer dataset
Vaka et al. (2020)	Machine Learning Techniques	High Accuracy	Custom Dataset
Abunasser et al. (2023)	ANN, SVM, kNN, Group Classifiers	83.45% - 99.30%	32 Selected Studies (2002-2020)
Amethiya et al. (2022)	ELM-RBF, SVM, RVM, ANN	Varied	Various Imaging and Blood Data Sources
Ak (2020)	Logistic Regression, kNN, SVM, Naïve Bayes, Decision Tree	98.1%	Dataset from Univ. of Wisconsin Hospital
Mangukiya et al. (2022)	SVM, Decision Tree, Naïve Bayes, kNN, Adaboost, XGBoost, Random Forest	98.24%	Wisconsin Breast Cancer
Tahmooresi et al. (2018)	SVM, ANN, kNN, Decision Tree	Effective Performance	Various Data Types

Rautela et al. (2024)	GANs, Vision Transformers	0.932 Sensitivity, 0.988 Accuracy	INbreast
Alotaibi et al. (2023)	ViT, DeiT	98.17%	BreakHis
Abimouloud et al. (2024)	ViT-CNN Hybrids	98.64% (ViT), 96.99% (CCT), 97.52% (MVIT)	BreakHis
Melekoodappattu et al. (2023)	CNN, Texture Feature Extraction	97.8% (MIAS), 98.3% (DDSM)	MIAS, DDSM
Sahu et al. (2023)	ShuffleNet-ResNet	99.17%	mini-DDSM, BUSI, BUS2
Sahu et al. (2023)	ShuffleNet-ResNet Hybrid Framework	98.13%	mini-DDSM, BUSI, BUS2

2.2.1 Similar Applications

Several studies have explored machine learning and classical deep learning methods to develop breast cancer detection systems. These works primarily focus on structured datasets, biosensor data, and traditional feature extraction methods for early diagnosis.

Sharma et al. applied Random Forest, kNN, and Naïve Bayes on the Wisconsin Diagnosis Breast Cancer dataset and achieved competitive accuracy, supporting their use in early detection models. Similarly, Mohammed et al. applied SMO, Naïve Bayes, and Decision Tree models on benchmark datasets, where SMO performed best on WBC, and J48 was most effective on another Breast Cancer dataset. These results emphasized the importance of choosing models according to dataset characteristics and imbalance issues. Vaka proposed an ML-based breast cancer detection system that demonstrated high accuracy on a custom dataset, while Abunasser et al. reviewed 32 studies and found that ANN and SVM were among the most effective classifiers, achieving accuracy ranging from 83.45% to 99.30%.

In another significant contribution, Amethiya et al. integrated biosensor data with ML algorithms like ELM-RBF, SVR, RVM, and ANN, emphasizing multi-source dataset compatibility and comparative accuracy. Ak's study compared several ML algorithms and reported that logistic regression with visual analytics

achieved a 98.1% classification accuracy on the University of Wisconsin dataset, showcasing how traditional models still offer value in data-driven diagnosis. Tahmooresi introduced a hybrid ML model using SVM, ANN, and Decision Trees, evaluated across diverse datasets, underlining the importance of model adaptability. These studies demonstrate that machine learning has consistently served as a foundation for breast cancer prediction systems, although they often lack the spatial interpretability and global attention that newer transformer-based methods offer.

Table 2.2: Similar Applications – Research Matrix

Author (Year)	Models Used	Accuracy (%)	Dataset Info
Mohammed et al. (2020)	Decision Tree, Naïve Bayes, SMO	Varied	WBC, Breast Cancer dataset
Vaka et al. (2020)	Tailored ML Techniques	High Accuracy	Custom Dataset
Abunasser et al. (2023)	ANN, SVM, kNN	83.45% 99.30%	– 32 Reviewed Studies
Amethiya et al. (2022)	ELM-RBF, SVR, RVM, ANN	Varied	Various biosensor and imaging sources
Ak (2020)	LR, kNN, SVM, Naïve Bayes, Decision Tree	98.1%	Univ. of Wisconsin Hospital Dataset
Tahmooresi et al. (2018)	SVM, ANN, kNN, Decision Tree	Effective Performance	Various Data Types

2.2.2 Related Research

Recent advancements in deep learning have significantly influenced breast cancer detection using histopathological and mammographic images. Several researchers have explored hybrid architectures, transformer-based models, and ensemble techniques to improve accuracy and computational efficiency.

Bhise integrated a CNN with Recursive Feature Elimination (RFE) and compared it with SVM, Random Forest, and Naïve Bayes, demonstrating CNN's superior accuracy in analyzing medical images on the BreakHis 400X dataset. Similarly, Mangukiya evaluated multiple classifiers including XGBoost, which achieved 98.24% accuracy, showing the effectiveness of boosting methods for structured medical datasets. Focusing on image quality and spatial detail,

Rautela proposed a GAN-ViT-based framework that significantly improved classification performance (0.988 accuracy on INbreast) by enhancing image resolution. Alotaibi introduced a ViT-DeiT ensemble model that achieved 98.17% accuracy on BreakHis, highlighting the potential of transformer-based models in histopathology.

Abimouloud proposed a lightweight ViT-CNN hybrid, achieving 98.64% accuracy on BreakHis while maintaining computational efficiency. These approaches reflect a growing shift towards transformer-based architectures for their ability to model complex spatial features in medical images. In terms of ensemble models, Sahu developed a ShuffleNet-ResNet hybrid, achieving 99.17% accuracy on DDSM and ultrasound datasets, demonstrating that hybrid CNN models can perform robustly even with limited data. Melekoodappattu also combined CNN with texture-based feature extraction, improving performance on DDSM and MIAS datasets. Ahad further validated ensemble approaches by comparing D-CNN, transfer learning, and ensemble strategies, with the latter showing the highest accuracy (99.94%).

These studies form the foundation for the present research, which proposes a unified ensemble integrating ViT, ResNet, and ResViT, combining the spatial strength of transformers with the feature extraction capabilities of CNNs for real-time, mobile-compatible breast cancer detection.

Table 2.3: Related Research – Matrix Table

Author (Year)	Models Used	Accuracy (%)	Dataset Info
Bhise et al. (2022)	CNN + RFE, SVM, RF, NB	High Accuracy	BreaKHis 400X
Mangukiya et al. (2022)	XGBoost, Adaboost, SVM, RF, etc.	98.24%	Wisconsin Breast Cancer
Rautela et al. (2024)	GAN + Vision Transformer (ViT)	0.988 Accuracy, 0.932 Sens.	INbreast
Alotaibi et al. (2023)	ViT + DeiT Ensemble	98.17%	BreakHis
Abimouloud et al. (2024)	ViT-CNN Hybrid	98.64%	BreakHis
Sahu et al. (2023)	ShuffleNet + ResNet (Hybrid CNN)	99.17%	mini-DDSM, BUSI, BUS2

Melekoodappattu et al. (2023)	CNN + Texture Feature Extraction	98.3% (DDSM), 97.8% (MIAS)	MIAS, DDSM
Ahad et al. (2023)	D-CNN, Transfer Learning, Ensemble	99.94% (Ensemble)	Not Specified (Mixed Histopath.)

2.3 Gap Analysis

Despite significant progress in breast cancer detection using machine learning and deep learning, several important gaps remain in the current body of research. These gaps highlight the need for an improved, hybrid, and deployable detection system that combines both accuracy and computational efficiency.

Limited Use of Unified Hybrid Architectures:

While some studies explored CNN-based models (e.g., Bhise [6], Melekoodappattu [21]) and transformer-based models (e.g., Rautela [2], Alotaibi [10], Abimouloud [9]), very few integrated CNN and ViT architectures into a single unified ensemble framework. Existing studies either focused on standalone ViT/CNN models or loosely fused them without an optimized structure like the proposed ViT-ResNet-ResViT ensemble.

High Computational Cost of Transformer-Based Models:

ViT-based models, although accurate (Rautela [2]: 98.8%, Alotaibi [10]: 98.17%), are often computationally expensive, making them unsuitable for real-time or mobile deployment. There is a lack of research focusing on lightweight and optimized versions of these models for use in low-resource settings. Your work addresses this by converting the ensemble model into a TensorFlow Lite (.tflite) format for integration into a Flutter mobile app.

Lack of Generalization on Diverse and Imbalanced Datasets:

Many existing studies (e.g., Sharma [4], Mohammed [13], Mangukiya [14]) used benchmark datasets like WBC or BreakHis, but did not adequately address class imbalance or variation in image quality, which are common in real-world datasets. Few studies applied advanced class balancing methods like weighted loss functions and image augmentation, both of which are employed in your methodology.

Absence of Real-Time Deployment and Application Design:

Although some studies achieved high accuracy in controlled environments, they did not focus on real-time usability or mobile application integration (e.g., Sahu [12], Ahad [22]). Your research bridges this gap by not only developing an accurate model but also deploying it in a real-time Flutter-based mobile app, making it practical for real-world clinical or rural healthcare use.

Limited Focus on Explainability and Robustness:

While ensemble and hybrid models are explored (Desai [20], Melekoodappattu [21]), very few studies analyzed their robustness across multiple architectures or ensured explainability in complex classifications. Your ensemble of ViT, ResNet, and ResViT is aimed at improving interpretability through diverse feature extraction and fusion.

Table 2.4: Gap Analysis Table – Existing Research vs. Proposed Work

Identified Gap	Evidence from Literature	How This Thesis Addresses the Gap
Lack of a unified hybrid model combining CNN, ViT, and ResViT	Most works used individual models: Bhise (CNN), Alotaibi (ViT), Abimouloud (ViT-CNN hybrids)	Proposes an ensemble combining ResNet, ViT, and ResViT for robust and complementary feature learning
High computational cost of ViTs limiting real-time/mobile deployment	Rautela [2], Alotaibi [10] achieved high accuracy but did not address efficiency or deployment	Optimizes the final model into .tflite format and integrates it into a real-time Flutter application
Poor handling of class imbalance and real-world dataset variability	Mohammed [13], Mangukiya [14] applied classical ML; few tackled imbalance systematically	Applies augmentation and class-weighted loss to improve fairness and learning from imbalanced datasets
Absence of mobile-ready diagnostic applications	Sahu [12], Ahad [22] used ensemble methods, but without real-time mobile deployment focus	Develops a practical mobile app using Flutter for on-device diagnosis using the lightweight model
Limited focus on ensemble robustness and feature diversity	Desai [20], Melekoodappattu [21] used basic CNN-MLP ensembles or shallow feature fusion	Uses a structured ensemble combining deep CNN (ResNet), transformer (ViT), and hybrid (ResViT) layers

2.4 Summary

This chapter presented a comprehensive review of the existing techniques and models used for breast cancer detection. The analysis of similar applications highlighted how traditional ML models such as SVM, Random Forest, and Naïve Bayes were applied effectively on benchmark datasets like WBC and BreakHis. On the other hand, related research demonstrated that deep learning and Vision Transformer models significantly improve classification performance but often lack computational efficiency and real-world deployability.

The gap analysis clearly identified the need for a unified hybrid ensemble model, capable of delivering high accuracy while being lightweight and mobile-friendly. This understanding forms the foundation for the proposed work, which integrates ResNet, ViT, and ResViT models in an optimized ensemble, designed for real-time usage through a mobile application.

Chapter 3

Research Methodology

The design process and research methodology for creating a machine learning-based book genre classification system are covered in detail in this chapter. This includes data collection, preprocessing, model training, and evaluation. The reasoning behind the selected methods and the eventual web-based application deployment are also covered.

3.1 Methodology

3.1.1 Overview

Cancer diagnosis of the breast remains a complicated and more important issue within the domain of medical imaging that needs a sophisticated computational procedure that can clearly differentiate between benign and malignant cases. Conventional methods are usually based on manual analysis of histopathological image, and it is time-consuming and subject to human error. In order to address these issues, this study introduces a hybrid convolutional neural network (CNN)-vision transformer (ViT) model, which is known as ResViT a Transformer-CNN hybrid architecture designed to classify breast cancer. The system architecture takes into consideration ResNet, which is characterized by the successful local feature extraction through residual connections, and ViT, which is characterized by the global attention mechanism and a high contextual representation. Both models are trained on large scale image datasets in order to benefit of transfer learning which increases their performance despite small amounts of domain specific data. The suggested ensemble model is aimed at combining the performance of ResNet, ViT, and ResViT to enhance the classification resilience and minimize the variability in predictions. It uses a publicly available dataset of histopathology images which consists of 6,000 images. Preprocessing of the dataset is done by using DPI adjustment, normalization, resizing, and contrast adjustment to maintain similar quality input. It is then split into 72% training, 18% testing, and 10% validation, maintaining balanced class representation across all phases. To ensure accessibility and real-time usability, the final

trained ensemble model is converted into TensorFlow Lite (.tflite) format and integrated into a Flutter-based mobile application. This enables offline inference directly on Android devices, making it a practical diagnostic aid for clinicians and healthcare workers in low-resource settings. The proposed methodology not only addresses technical challenges in medical image classification but also bridges the gap between AI-based diagnostics and real-world healthcare deployment.

3.1.2 Proposed Methodology

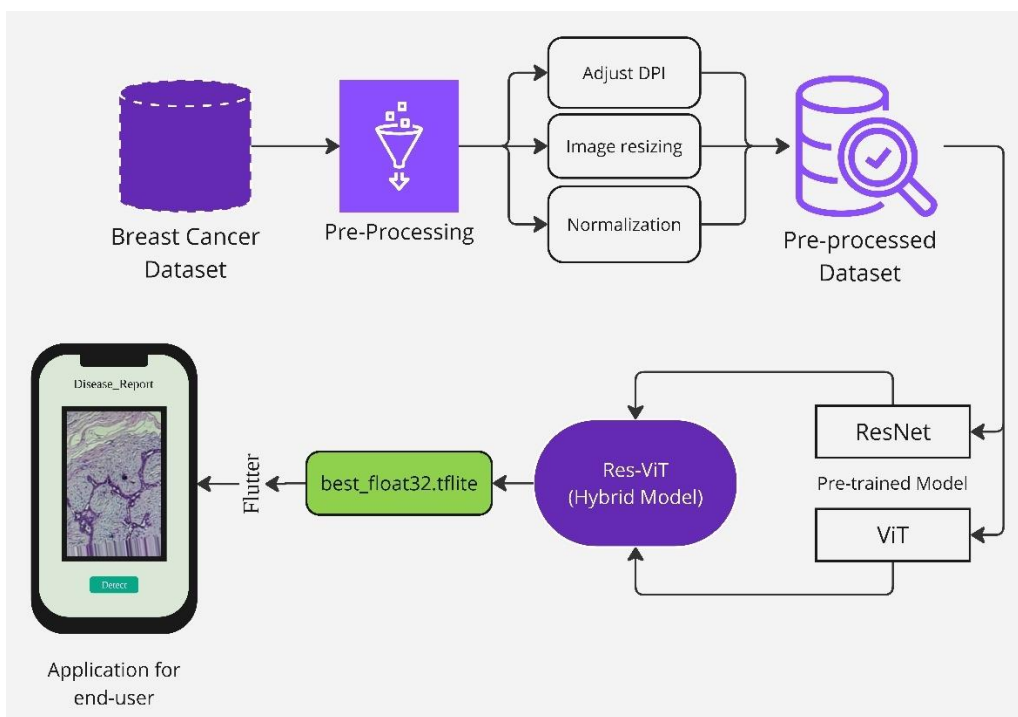


Figure 3.1: The Methodological Flowchart

Figure 3.1 illustrates the system design of the proposed methodology, detailing the entire pipeline from data preprocessing to model deployment. The pipeline begins with the Breast Cancer Dataset, which undergoes preprocessing involving normalization, DPI adjustment, and image resizing to produce a standardized preprocessed dataset. This dataset is used to train the ResViT hybrid model, which combines the pretrained ResNet and ViT models. The resulting model is exported as a TensorFlow Lite file (`best_float32.tflite`) for deployment. A mobile application, developed using Flutter, acts as the end-user interface, allowing users to upload histopathological images and receive diagnostic reports, thereby

bridging advanced AI models and practical clinical usage. The figure encapsulates the seamless integration of preprocessing, model training, hybridization, and deployment within the proposed system.

3.1.3 Functional and Nonfunctional Requirements

The proposed system has essential functional and nonfunctional requirements to ensure robust performance and practical usability. The functional requirements focus on the system's core tasks, while the nonfunctional requirements ensure it meets performance, usability, and security standards in real-world applications.

Functional Requirements:

- **Dataset Preprocessing:** The system must preprocess the breast cancer dataset by performing normalization to scale pixel values, DPI adjustment to standardize image resolution, and resizing to ensure uniform input dimensions for the models.
- **Model Training:** The system should utilize the preprocessed dataset to train the ResViT hybrid model by combining pretrained ResNet and Vision Transformer (ViT) architectures, ensuring optimal feature extraction and classification accuracy.
- **Image Classification:** The trained ResViT model must reliably classify histopathological images into benign or malignant categories with high precision and recall.
- **Mobile Application Functionality:** The mobile application of Flutter needs to allow users to add histopathological images, work on them with the help of the TensorFlow Lite model, and to show the diagnostic results in real-time.
- **Model Deployment:** The TensorFlow Lite (*.tflite) model should be effectively incorporated within the mobile app to deliver quick and correct predictions without compromising with the various mobile phone models.

Nonfunctional Requirements:

- **Performance:** The system must be highly accurate (e.g. greater than 90) and give diagnostic predictions at low latency to help in real-time medical

decision-making.

- **Scalability:** The design should support the easy addition of larger datasets or more classes, or more complex model architectures in the future, without having to re-engineer the design significantly.
- **Usability:** The mobile application must feature a user-friendly interface with clear instructions, ensuring accessibility for healthcare professionals and non-technical users alike.
- **Portability:** The TensorFlow Lite model must be optimized for deployment on resource-constrained mobile devices, ensuring compatibility and smooth operation across various hardware configurations.
- **Security and Privacy:** The system should meet the data protection requirements, and user inputs and diagnostic outputs should be encrypted in order to prevent unauthorized access or breach of sensitive medical data.

3.1.4 Data Flow Diagram Level 1

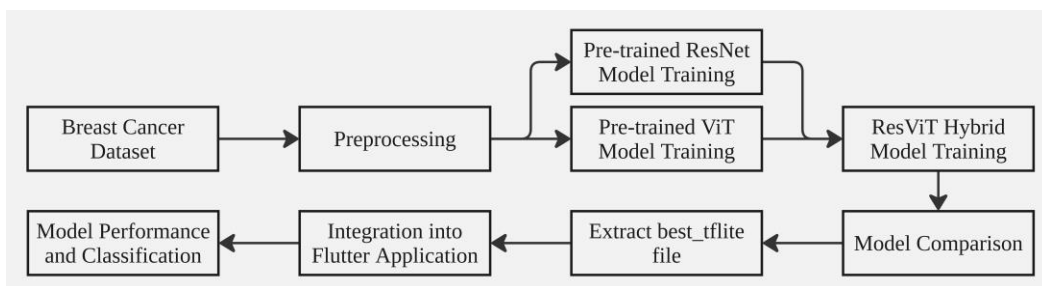


Figure 3.2: Data Flow Diagram Level 1

The Level 1 Data Flow Diagram gives a close-up view of the data flow in the breast cancer detection system that gives the sequential processes of the system starting with the entry of data and ending with the deployment of the final model. It starts with the dataset of breast cancer, which is composed of histopathological images and undergoes a preprocessing phase that involves normalization, DPI correction and image size reduction, to get a ready data to be used in the training. The pretrained models, ResNet and Vision Transformer (ViT), are then trained using the preprocessed dataset and used to extract meaningful features of the images. These characteristics are incorporated in the ResViT hybrid architecture which amalgamates the strengths of the two architectures into the best classification accuracy. After the training stage, the ResViT model is tested on the ResNet and ViT models and the most successful

model is deployed. The selected model is optimized to lightweight a TensorFlow Lite format and deployed as a file called tflite in a lightweight format. This transformed model is then incorporated into a mobile application built on Flutter that allows the user to take up histopathological images and get real-time prognosis of whether they are benign or malignant. This order and efficiency allows accuracy, usability and flawless interaction between the elements of the system.

3.1.5 UI Design

The design of the user interface on the breast cancer detection system focuses on simplicity, accessibility and functionality to provide end-users with a smooth user experience. It is grounded on two tier architecture which combines a user friendly Android based application and a powerful backend process.

1. Application Setup

Android applications are the main point of interaction with the user. It has a simple interface with basic functionality, including an option to take a picture with the device camera or to take an image with the device gallery. After choosing an image, the process of the backend analysis is initiated by the "Detect" button. The outcome is presented as a well-organized report of the detection; the label of a class (Benign or Malignant) and the relevant score, which is the accuracy or the confidence. This simplified design will make sure that the application is easy to navigate by users, even non-technical ones, in order to access credible diagnostic information.

2. Two-Tier Architecture

The bilateral structure of the breast cancer detection system provides the smooth flow of information between the user interface and the background processing logic. This architecture has three crucial components namely frontend, and backend that collaborate to provide the correct and consistent results to the end user.

Frontend: The Android application is written with Flutter. It has a simple clean user friendly interface that is easy to navigate. The primary view also has the

functions of taking a photo through the device camera or uploading one in the gallery. After picking an image, the user is then able to start the detection process by clicking the "Detect" button. The result of the classification, including the identified classification and a confidence score, is shown in an aesthetically appealing detection report after the process. The frontend makes sure that the latency is minimal and is responsive, and it is more accessible to users with different technical skills.

Backend: The backend will do the heavy lifting which consists of preprocessing, model inference and result formatting. When an image is uploaded, it is subjected to a preprocessing procedure, such as resizing, normalizing, and formatting it to the input specifications of the TensorFlow Lite model. The generated processed image is next sent to the hybrid ResViT model, which uses the capabilities of both ResNet and ViT models to produce the most precise predictions. These tasks are effectively handled by the backend to have low computational overhead and quick production of results.

3. Folder Structure:

The folder structure is designed to promote clarity, maintainability, and modularity. The application folder structure is designed to ensure modularity and ease of development. It contains essential files for model execution, label mapping, and the Flutter-based user interface. The overall folder structure has given bellow.

3.1 Application Directory:

- `breast_cancer_detector/` → Root directory containing the entire application.
- `assets/` → Contains essential resources for the application.
- `labels.txt` → Stores the class labels ("Benign", "Malignant"), used for inference and result display.
- `model.tflite` → The optimized TensorFlow Lite (TFLite) model, responsible for processing input images and making predictions.
- `main.dart` → The primary Flutter application file, handling UI components, backend communication, and result display.

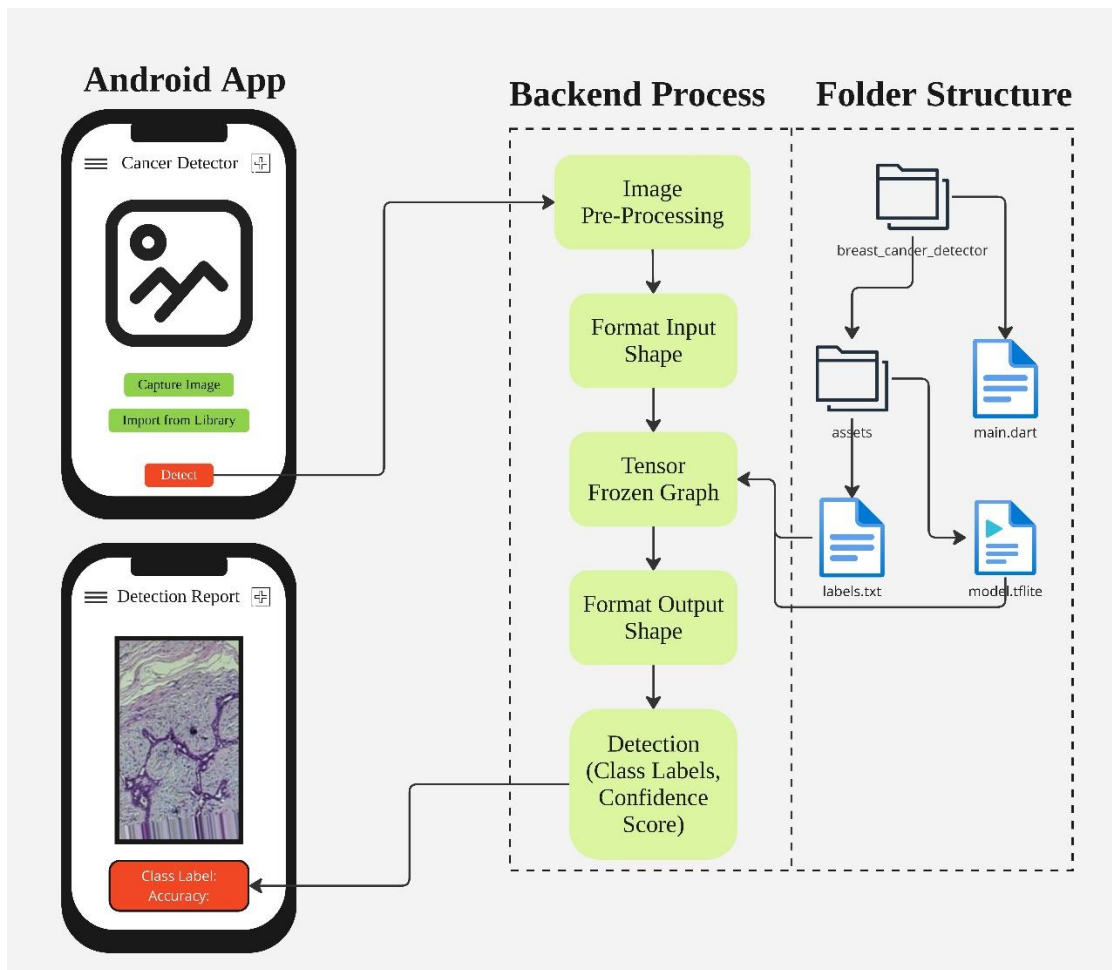


Figure 3.3: The Two-Tier Architecture and UI for Application

3.2 Detailed Methodology and Design

3.2.1 Dataset

The dataset used for this study is sourced from the publicly available Breast Cancer Dataset. It comprises 6,000 histopathological images of breast tissue classified into two categories: Benign and Malignant. The dataset has been pre-divided into three subsets for effective model training and evaluation, with 72% allocated for training, 18% for testing, and 10% for validation. The images exhibit variations in size and resolution, necessitating preprocessing steps such as resizing, DPI adjustment, and normalization to ensure uniformity. The data is used in training the models and proving the correctness of the suggested hybrid ResViT architecture. It provides solid classification as it harnesses its diversity and labelled organization.

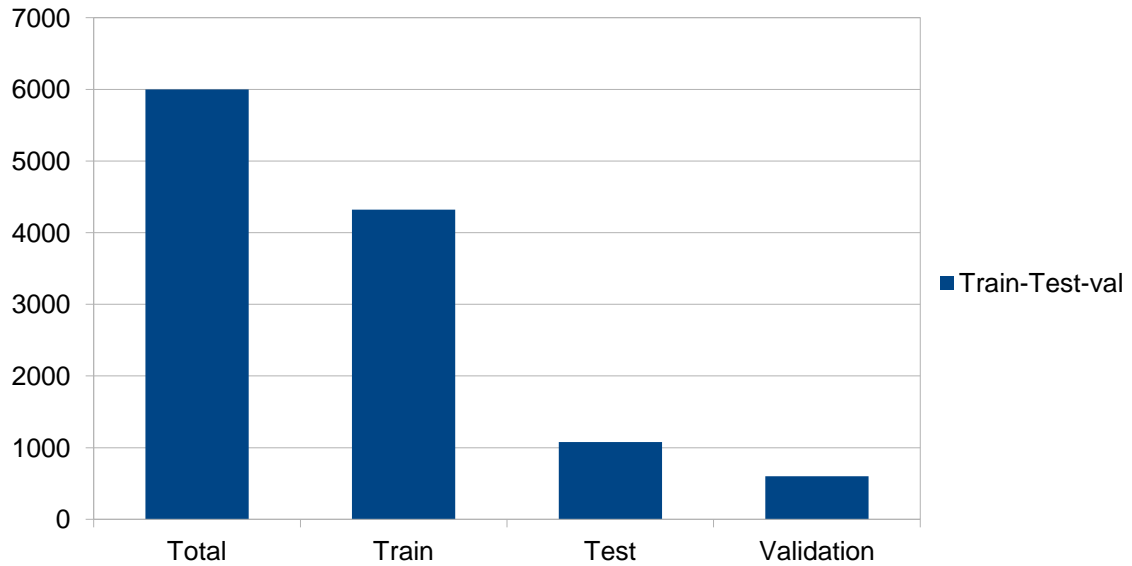
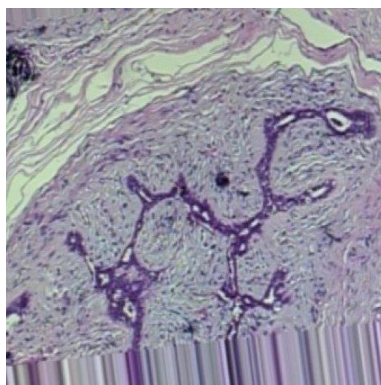


Figure 3.4: Train-Validation_Test Splitting

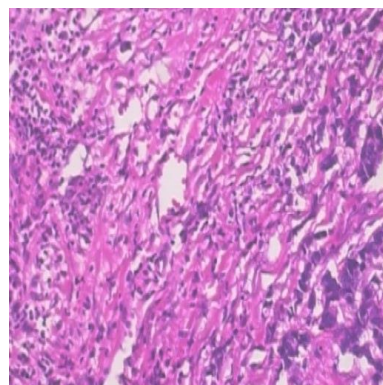
3.2.2 Classes

To make an accurate classification, the dataset has been split into two classes:

- **Benign:** Refers to non-cancerous tissue of the breast but with normal cell structures, and without invasive tissue growths. This is the class that acts as a control factor, in order to allow the model to differentiate healthy tissue by minimizing false positives.
- **Malignant:** This is a cancerous growth of breast tissue with abnormal and invasive cell growth, which is a sign of breast cancer. Early detection and treatment of this type of case rely heavily on the appropriate classification.



Benign



Malignant

Figure 3.5: Sample Image of Each Class

Table 3.1: Dataset Specifications

Properties	Values
Image Resolution	240 × 240 pixels
Format	.jpg
Total Images	6000
Classes	2

3.2.3 Data Preprocessing Steps

Preparation of the breast cancer dataset to be used in the proper training and performance of models is a very sensitive stage of data preprocessing. The steps applied were as follows:

- Normalization: Image pixel values were transformed to the range [0, 1] in order to make them uniform and to stabilize the learning process of the model.
- Calibration of DPI: To ensure that all the samples had the same resolution, the images were calibrated to a given DPI (dots per inch). This is done to ascertain consistency and compatibility when extracting features.
- Resizing: The images have been resized to a set size of 224 x 224 pixels (as per the input of pre-trained models Resnet and ViT).

Table 3.2: Preprocessing Properties and Description

Step	Description
Normalization	Pixel values are scaled to a range of [0,1].
Adjusting DPI	Standardizes the resolution (dots per inch) across all images.
Resizing	Images are resized to 224×224 pixels.

3.2.4 Model Description

1. ResNet Model

This paper uses a pretrained ResNet model to classify breast cancer, and it uses the hierarchical feature extraction property of the model. The sections that follow describe stepwise working procedure of the ResNet model with respect to this study.

- **Input Processing:** As the input to the ResNet model, the histopathological images of breast cancer are first preprocessed to be uniform. All pictures are downsampled to 224 x 224 pixels, and their values are scaled to the range [0,1] to increase the processing speed. The processed image is obtained as a preprocessed number which is then transferred into a tensor format to be utilized in deep learning models.
- **Convolutional Feature Extraction:** ResNet model starts with the provision of feature extraction with stacked convolutional layers. These layers are used to implement several 3x3 convolution filters on the image, similar to the spatial hierarchies and patterns that are critical to classification. There is a convolutional layer that is succeeded by batch normalization and ReLU (Rectified Linear Unit) activation to provide non-linearity and keep the gradient flow constant.
- **Unnecessarily Learning and Skip Connections:** Skip (shortcut) connections and ResNet. Unlike existing methods, ResNet includes skip (shortcut) connections that may bypass one or more layers and do identity mapping. This enables the network to receive residual functions instead of the direct transformation of inputs, alleviating the issue of vanishing gradients. Mathematically, a residual block consists of:

$$y = F(x) + x.$$

$F(x)$ is the learnt transformation using convolutional layers, and is the original input channeled via the skip connection. These links allow the deep architecture training without a considerable performance drop.

- **Down sampling via Pooling:** In order to downsample the spatial dimensions and computation, max pooling layers are inserted on a periodical basis. The max pooling picks the most prominent features at each region, and thus retains the important data, but gets rid of irrelevant data. The process will assist the model to be generalized to invisible data.
- **Classification and Fully Connected Layers:** Once the extracted features have been extracted through several convolutional and residual ones, they are flattened and then sent into a fully connected layer. This layer is followed by the output passing through a softmax activation function where the probability distribution of the two target classes (Benign and Malignant) is

calculated. It is classified and an answer is determined according to the greatest probability score.

- **Transfer Learning and Fine-Tuning:** In this study, a pretrained ResNet model is employed, leveraging weights trained on large-scale datasets. The final layers are fine-tuned using the breast cancer dataset to enhance domain-specific feature learning. This approach significantly improves classification accuracy while reducing training time.
- **Output and Prediction:** Once the classification is performed, the model outputs the predicted class label along with a confidence score. The final decision is displayed in the mobile application, providing users with an interpretable and reliable diagnostic result.

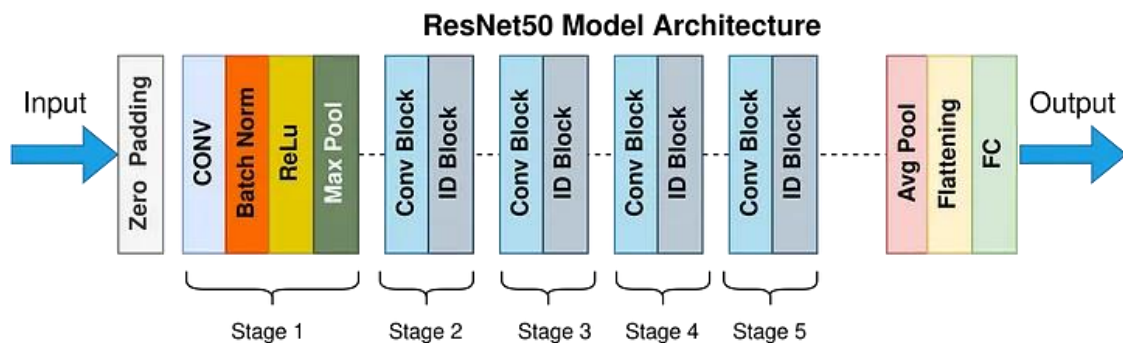


Figure 3.6: The Architecture of ResNet50 [<https://medium.com/towards-data-science/the-annotated-resnet-50-a6c536034758>]

2. Vision Transformer Model

In this study, a pretrained ViT model is fine-tuned for breast cancer classification, leveraging its ability to extract global contextual information efficiently. The following sections detail the step-by-step working procedure of the ViT model.

- **Input Processing and Patch Embedding:** Unlike traditional CNNs that process entire images using convolutional filters, ViT first divides the input image into non-overlapping patches of fixed size (e.g., 16×16 pixels). Each patch is then flattened into a vector and passed through a learnable linear projection layer, converting it into a patch embedding.

- **Positional Encoding:** Since transformers do not inherently capture spatial relationships, positional encodings are added to the patch embeddings. These encodings help the model understand the order and spatial layout of image patches, ensuring that local and global dependencies are preserved throughout the network.
- **Multi-Head Self-Attention (MHSA) Mechanism:** The core of ViT lies in its multi-head self-attention (MHSA) mechanism, which allows each patch to attend to all other patches in the image. This mechanism computes three vectors for each input token:
 - **Query (Q):** Represents the token being attended to.
 - **Key (K):** Represents the tokens being compared.
 - **Value (V):** Holds the feature representation of tokens.
- **Transformer Encoder Layers:** ViT consists of multiple transformer encoder blocks, each containing:
 - Multi-head self-attention to capture global relationships.
 - Layer normalization to stabilize learning.
 - Feedforward neural networks to process extracted features.
 - Skip connections to facilitate gradient flow.

These layers help the model learn hierarchical representations of breast cancer images, improving classification performance.

- **Classification Token and Fully Connected Layer:** A special [CLS] (classification) token is prepended to the input patch embeddings and serves as the final feature representation of the image. After passing through transformer layers, the output of this token is fed into a fully connected (FC) layer with a softmax activation function, which predicts the probability of the image belonging to either the Benign or Malignant class.
- **Transfer Learning and Fine-Tuning:** A pretrained ViT model, trained on large-scale datasets like ImageNet, is fine-tuned on the breast cancer dataset to enhance domain-specific learning. Fine-tuning optimizes only the last few layers while retaining the general features learned from the larger dataset. This approach accelerates training and improves classification accuracy.
- **Output and Prediction:** Once the model makes a prediction, it outputs the predicted class label (Benign or Malignant) along with a confidence score. These results are then displayed in the mobile application, providing users

with clear and interpretable diagnostic feedback.

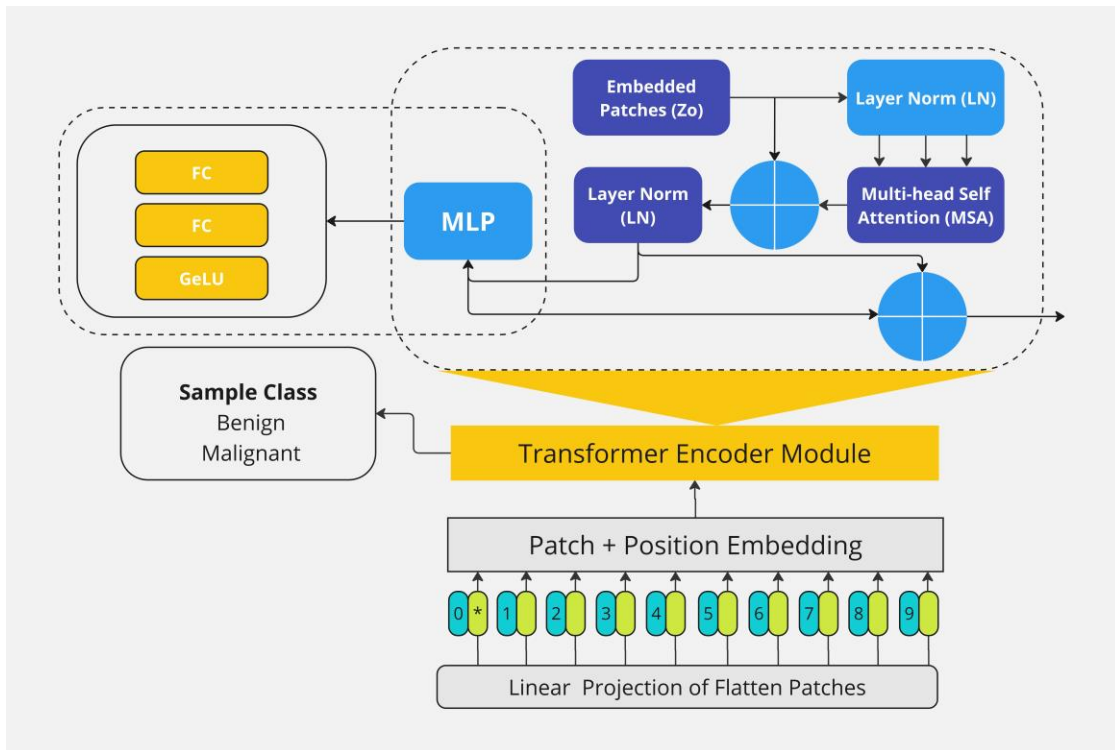


Figure 3.7: The Architecture of Vision Transformer Model

3. ResViT Hybrid Model

The ResViT model is a hybrid deep learning architecture that integrates the strengths of Residual Networks (ResNet) and Vision Transformers (ViT) to improve the accuracy of breast cancer classification. This algorithm provides the step-by-step process of training and optimizing the model in terms of a breast cancer dataset.

- Formulate the input shape to make use of CNN and ViT.
- Fuse CNN model (ResNet) and ViT model into a single model.
- Fuse features during convolutional and transformer-based representations and, by combining rock-solid local spatial features (CNN) and global dependencies (ViT)-based attention, have them work together to achieve an enhanced classification capacity.
- Hyperparameter Optimization: Define a range of values for key hyperparameters such as:
 - Image size (e.g., 224×224, 256×256)
 - Batch size (e.g., 16, 32, 64)

- Epoch count (e.g., 20, 50, 100)
- Train the ResViT model on different combinations of hyperparameters.
- Append evaluation results (accuracy, loss) for each configuration to evalList. Select the best-performing model based on validation accuracy and loss.
 - Model Training and Evaluation: Train the final optimized ResViT model using the training dataset. Measures of evaluation like accuracy, precision, recall, and F1-score are used to monitor performance. Test the model on the test data to determine the generalization capability.
 - Save trained ResViT model: Save the trained ResViT model as integrated model.h5 in order to use it later to do inference. Implement the model within the breast cancer detection system, and combine it with the mobile application to conduct real-time histopathological images classification.

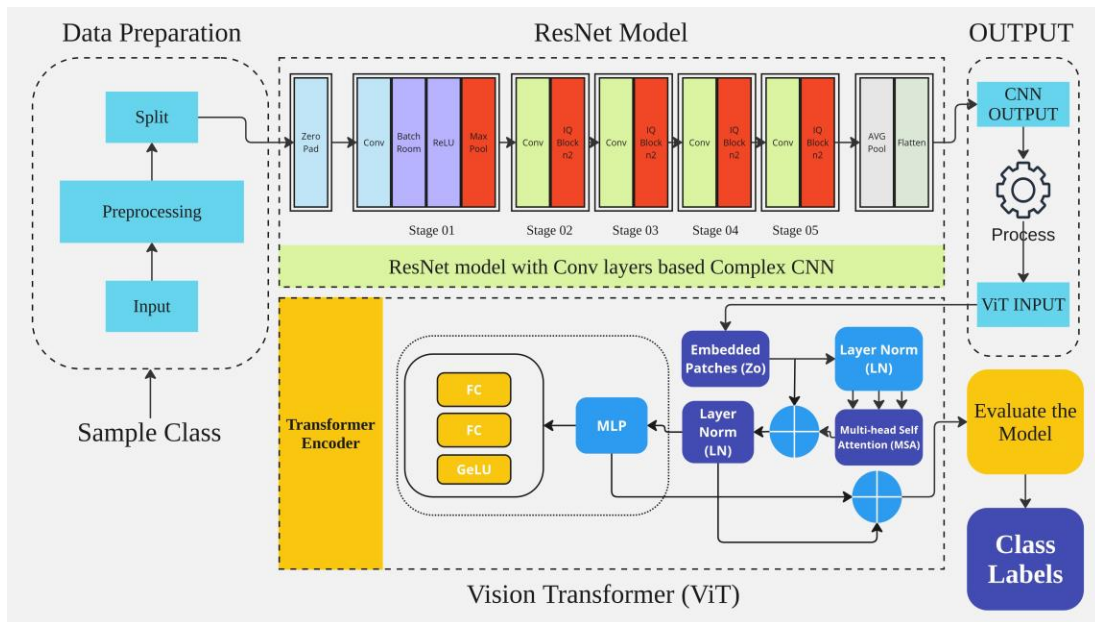


Figure 3.8: The proposed Architecture of the Res-ViT Hybrid Model

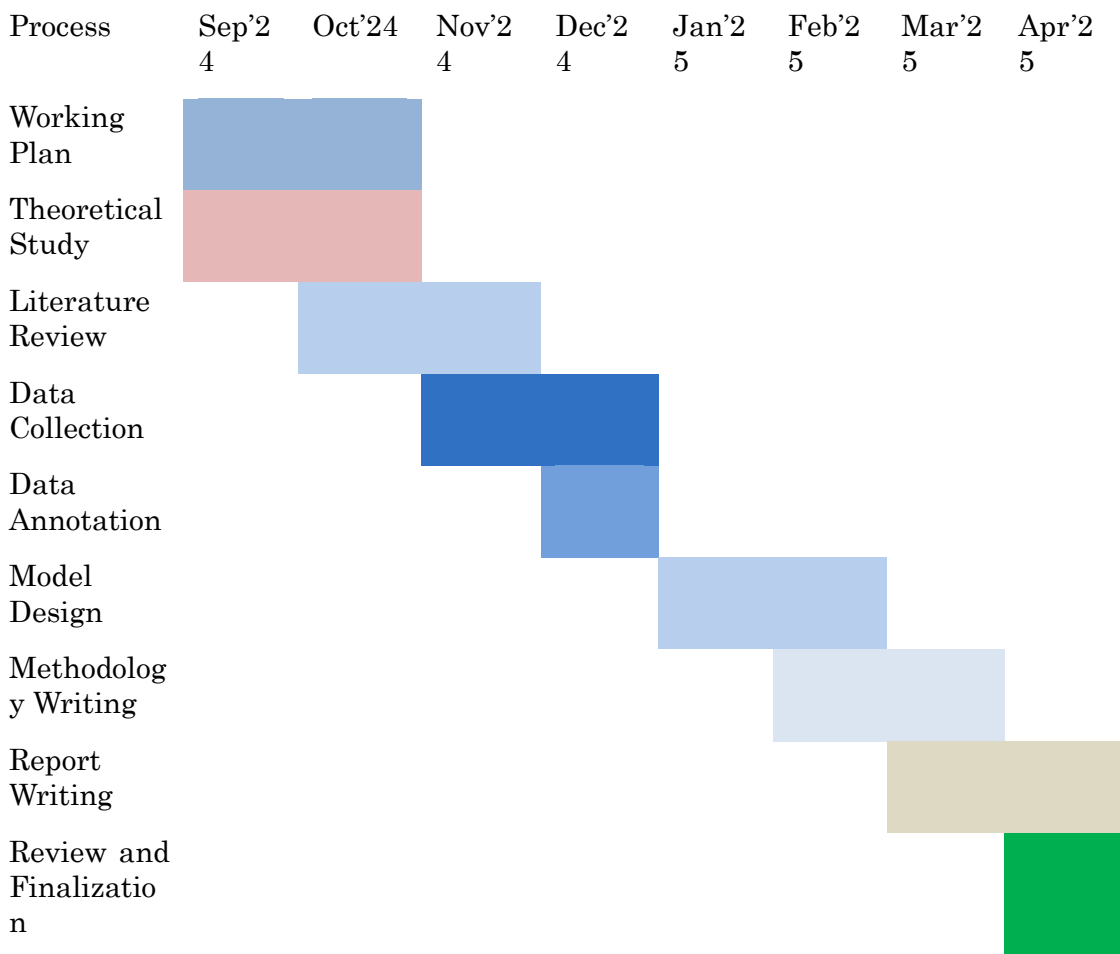
3.3 Project Plan

The project is organized in a plan to guarantee the stepwise construction and review of the model of breast cancer detection with the aid of the ResViT architecture. The plan starts with the acquisition and preprocessing of the datasets, wherein the histopathological images are normalized, resized, and divided into training, validation images and test images. After that, the model architecture is developed based on the combination of a pretrained ResNet to

©Daffodil International University

extract spatial features and Vision Transformer (ViT) to capture long-range dependencies. Hyperparameter optimization methods are then used to fine-tune the hybrid model in order to improve the performance of classification. The training stage is an iterative learning process in which the weights of the model are modified through backpropagation, in order to reduce the error of classification. Accuracy, precision, recall, and F1-score are some of the metrics used in the performance evaluation to make sure that the evaluation is robust. When the precise outcomes are attained, the trained model is exported and embedded in the form of a convenient application with a two-level architecture, which helps ensure a high-level communication between the frontend and the backend. The last stage will be deployment and on field testing to evaluate the effectiveness of the model in automated diagnosis of breast cancer. Systematic testing, validation and refinements are conducted throughout the project lifecycle in order to augment model reliability and usability.

Table 3.3: GANTT Chart of Project Timeline



3.4 Task Allocation

So as to make sure that this research project is carried out in a systematic manner, the whole process was broken into clear tasks. Every task was planned diligently with unique goals, schedules, and interdependencies to keep up efficiency and consistency during the development cycle. Tasks involved in the analysis of the problem, preparation of the data, training of the model, integration of the ensembles, evaluation, mobile deployment, and documentation. Since it is a solo project, everything was done by the researcher. The allocation scheme will guarantee appropriate time management and an equal amount of workload to successfully complete the technical implementation as well as the research reporting in the given time schedule.

Table 3.4: Task Allocation Table

Task ID	Task Description	Assigned To	Start Date	End Date	Remarks
T1	Problem analysis and literature review		01-01-2025	10-01-2025	Understanding existing methods and identifying gaps
T2	Dataset collection and preprocessing		11-01-2025	18-01-2025	Normalization, resizing, DPI adjustment
T3	Individual model training (ResNet, ViT, ResViT)		19-01-2025	28-01-2025	Using Google Colab GPU
T4	Ensemble integration of ResNet, ViT, ResViT	Faiiaz Bin Rhaman	29-01-2025	02-02-2025	Weighted fusion or averaging technique
T5	Model evaluation (Accuracy, AUC, etc.)		03-02-2025	06-02-2025	Using validation and test sets
T6	Model conversion to TFLite		07-02-2025	08-02-2025	Post-training optimization for mobile
T7	Flutter app design and development		09-02-2025	15-02-2025	Includes UI and TensorFlow

				Lite integration
T8	Testing on Android device	16-02-2025	18-02-2025	Performance and real-time inference test
T9	Final documentation and report writing	19-02-2025	28-02-2025	Thesis compilation, proofreading, formatting

3.5 Summary

The proposed research is a hybrid deep learning design to be used in detecting breast cancer effectively by boosting ViT, ResNet, and ResViT ensembles by combining them. The methodology is a pipeline that is systematic, starting with dataset preprocessing, in which histopathological images are processed to attain normalization, resizing, and DPI, which makes the input quality consistent. The data is then divided into training, validation and test sets in order to support the generalization of the model. The basic model architecture uses the convolutional properties of ResNet to extract hierarchical spatial features of the image and the self-attention properties of ViT to extract long-range dependencies to create a hybrid ResViT model that yields a higher level of classification accuracy. The trained model can be introduced into user-friendly interface that will be based on a two-level application architecture to allow the smooth entry of images, prediction and visualization of the results. To achieve high diagnostic reliability, the model will be trained, hyperparameter optimized and evaluated using commonly used measures like accuracy, precision, recall, and F1-score. The trained model, which is the last one, is the integrated model.h5, which is deployed to be used in the real world, which would support its possibility of use as an automated breast cancer detection model. The proposed methodology shows an advanced deep learning solution, which aids in early and precise breast cancer diagnosis, through systematic experimentation, fine-tuning, and validation of the performance.

Chapter 4

Implementation and Results

This chapter compares the performance of three deep learning models namely ResNet, Vision Transformer (ViT) and a hybrid ResViT model to classify medical images. Accuracy, precision, recall, F1-score, and AUC score are assessed by the study on the basis of loss curve, confusion matrices, and classification reports. A CNN-based model, ResNet, is used as a baseline, whereas ViT is based on self-attention mechanisms and better features extraction. The hybrid ResViT model is a fusion of CNN and Transformer-based models that is aimed at better classification results. The findings suggest that ResNet has poor generalization, ViT provides better feature representation, whereas ResViT has the best accuracy and recall. The chapter presents a comparative analysis as the most useful model to operate in real-life situations is ResViT.

4.1 Environment Setup

Table 4.1 describes the typical parameters employed to train all Vision Transformer (ViT) models in the study so that the experimentation will be consistent. The size of the image will be fixed to 48 x 48, i.e. the size of an input image will be reduced to this size and then the image will be run through the model. The size is optimal in terms of the computational efficiency and the capacity to extract the relevant information in the images. The batch size is 16, i.e. the model takes 16 images at once before it can update its weights. This is an appropriate batch size because it has stable learning and can be managed on a manageable memory usage which is friendly to deep learning tasks.

The models are trained in 80 epochs, which provide them with sufficient time to learn the complicated patterns in the dataset. This many-epoch training can improve the performance of the model besides reducing the chances of underfitting, although this should be checked frequently to avoid overfitting. The Adam optimizer will be utilized which is a popular optimization algorithm that has an adaptive learning rate and is able to effectively work with noisy data. It is a combination of AdaGrad and RMSProp, which is especially useful with deep

learning models. The learning rate will be 0.001 which is a moderate learning rate that will enable the model to converge at a steady rate without causing any extreme changes in the weights that would cause the training process to go out of control. The model also takes 6 attention heads that is, the self-attention process is broken into six parallel heads, which enables the model to simultaneously deal with various features of the image. This arrangement is useful in enhancing feature derivation and robustness of the model. On the whole, these parameters are selected with good consideration to maximize performance, with effective training and effective learning being ensured in all ViT models.

Table 4.1: Common parameter table for all experimented ViT models.

Parameter Name	Parameter Value
Image Size	48 × 48
Batch Size	16
Epoch	80
Optimizer	Adam
Learning Rate	0.001
Number of heads	6

The data split strategy of all the experiments in Vision Transformer (ViT) models is presented in Table 4.2 and remains the same to provide an equal and comprehensive analysis of the model experiments. The train set comprises 72 percent of the dataset amounting to 4435 images which gives the models an opportunity to learn a varied and rich pool of data. This big training set is essential in the process of capturing the complex patterns and features which are essential in making the right predictions so that the model is well equipped to extrapolate outside the training data set. The validation set consists of 18% of the dataset consisting of 975 images. The set is critical in the process of model training, as it enables real-time monitoring of performance as well as hyperparameter optimization. Through testing the model on this validation information, we are able to determine the strength of the generalization of the model as well as make adjustments where needed that are required to prevent overfitting such that the model is strong and correct. The test set includes 1/10

of the dataset (600 images) and is only applied to the final model after the training has been completed. This part of data is not altered in the course of training and validation giving an objective test of the real-world performance of the model. This split (72% training, 18% validation and 10% testing) will provide the study with a rigorous and fair evaluation of the capacities of each ViT model to learn, validate and generalize well.

Table 4.2: Common data split for all experimented ViT models.

Dataset	In Percentage	Number of Images
Train set	72%	4435
Validation Set	18%	975
Test Set	10%	600

4.2 Testing and Evaluation

In evaluating, comparing, testing the effectiveness of machine learning models for the study, appropriate performance metrics must be used to provide insights into model accuracy, reliability, and generalization capabilities. The following metrics are commonly used in classification tasks, especially in the context of agricultural disease detection:

4.3.1 Accuracy

The proportion of correctly classified instances (both positive and negative) to the total instances. Accuracy gives a quick overview of model performance but can be misleading in imbalanced datasets where one class significantly outnumbers the other.

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

4.3.2 Recall

The ratio of correctly predicted positive observations to all actual positives. Recall is particularly important in scenarios where a positive case (such as a diseased plant) could lead to severe consequences, like crop loss.

$$Recall = \frac{TP}{(FN + TP)}$$

4.3.3 Precision

The ratio of correctly predicted positive observations to the total predicted positives. Precision is crucial in applications where the cost of false positives is high. In this study, high precision indicates that when a disease is predicted, it is likely to be true.

$$Precision = \frac{TP}{(FP + TP)}$$

4.3.4 F1-Score

The harmonic meaning of precision and recall, providing a balance between the two metrics. The F1 score is especially useful when dealing with imbalanced classes, as it considers both false positives and false negatives, offering a more comprehensive view of model performance.

$$F1\ Score = 2 \times \frac{Precision + Recall}{(Precision \times Recall)}$$

4.3.5 Confusion Matrix

A table used to describe the performance of a classification model, showing the true vs. predicted classifications. The confusion matrix provides insights into the types of errors made by the model, allowing for more targeted improvements.

4.3.6 Receiver Operating Characteristic (ROC) Curve

A graphical representation of a classifier's performance across various threshold settings, plotting the true positive rate (recall) against the false positive rate. It demonstrates the diagnostic accuracy of a binary classifier system with manipulation of discrimination threshold. It is a curve of the True Positive Rate (TPR) or recall or sensitivity versus the False Positive rate (FPR).

Area Under Curve (AUC): Area under ROC curve, and is an individual percentage that is used to assess the performance of the model; the closer the model is to perfect the closer the value of AUC is to 1.

The Receiver Operating Characteristic (ROC) curve and Area Under the curve (AUC) are significant to the study since they are crucial in the determination of classification models. They allow comparing the model performance visually between various thresholds which allows deciding which threshold is the best one by trade off between sensitivity and specificity. The ROC curve is also

immune to the imbalances of the classes and also the curve provides acceptable readings in the event that accuracy may be wrong. However, other measures like precision and recall should be applied in conjunction with ROC and AUC in order to have a comprehensive analysis of the model performance.

4.3 Results and Discussion

Under this subsection, the results of the experimented results and its brief discussion under the above mentioned training set up and also under evaluation metrics will be presented.

4.3.1 Result of the ResNet50 model

The loss curve and accuracy curve of ResNet50 model helps to have a good understanding of its learning behavior during 80 epochs (early-stopped at epoch 15). The training loss starts high and gradually decreases, showing effective optimization. Similarly, the validation loss follows a decreasing trend, confirming that the model generalizes well without significant overfitting. However, there is a slight gap between training and validation loss, indicating that while the model learns well, some degree of overfitting may be present. The accuracy curve shows a steady improvement, with both training and validation accuracy converging around 87%. This suggests that the model has reached its optimal learning point, and additional training would yield minimal improvements. Early stopping ensures that the model halts training before overfitting occurs. Overall, the ResNet50 model demonstrates a stable learning process, but it may not be the most efficient for complex classification tasks compared to transformer-based models.

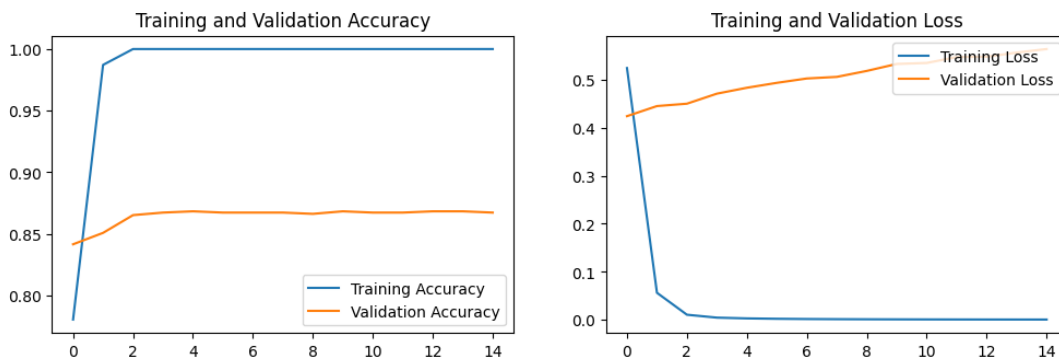


Figure 4.1: Loss and accuracy curve of the ResNet50 model over 80 epochs (early stopped at epoch 15).

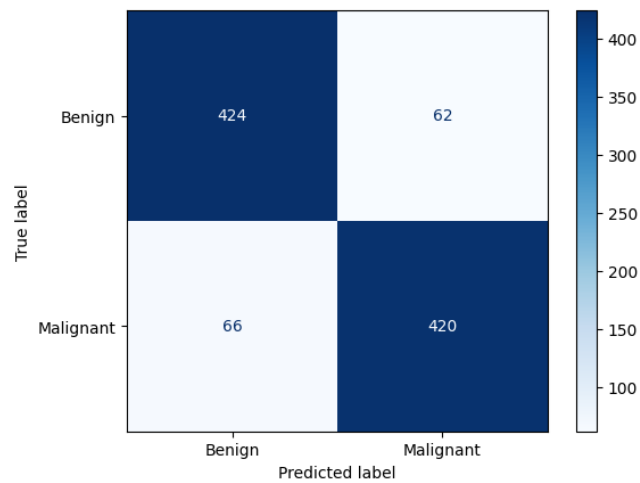


Figure 4.2: Confusion matrix on the validation set of the ResNet50 model.

The validation confusion matrix for ResNet50 showcases its performance in distinguishing between Benign and Malignant cases. The majority of instances are correctly classified, but some misclassifications occur, particularly in the Malignant class. This suggests that while the model is generally reliable, it struggles slightly with distinguishing certain complex cases. The matrix indicates that false negatives (incorrectly classified malignant cases) are higher than false positives, which may be critical in medical practice where false diagnosing of a malignant case may happen with severe outcomes. The general classification distribution is in line with the precision and recall in the classification report. Although the confusion matrix confirms that ResNet50 has an accuracy of 87, it is clear that it should be improved in terms of sensitivity, especially with regard to detecting challenging malignant cases. The classification report gives a comprehensive breakdown of the performance of ResNet50 on the Benign and Malignant cases. The model has an accuracy, recall and F1-score of 0.87 in each of the two classes meaning that the model has fair reliability in classification. The F1-score of 0.87 represents the performance of both classes being well balanced whereas the weighted-average F1-score also indicates the consistency in the data. Precision is also high but the recalls in malignant cases are slightly less at 0.85, indicating that the model is likely to overlook some of the real malignant cases thus false negatives. This is especially worrying in the medical uses where prompt diagnosis is of utmost importance. All in all, although ResNet50 is doing quite well, its recall can be improved to

present more accurate results to identify more malignant cases, which will reduce the false negative.

Table 4.3: Classification report on the validation set of the ResNet50 model.

Classes	Precision	Recall	F1-score	Support
Benign	0.87	0.87	0.87	486
Malignant	0.87	0.86	0.87	486
accuracy			0.87	972
macro avg	0.87	0.87	0.87	972
Weighted avg	0.87	0.87	0.87	972

Even the test confusion matrix of ResNet50 is an additional measure of the performance of the model in terms of its capacity to generalize to unseen data. The trend observed in the validation set persists, the majority of Benign and Malignant samples being correctly classified, however, the trend in terms of false negatives is on the rise which could indicate that the model is a bit more conservative in terms of distinguishing between malignant cases during the testing phase. The test accuracy becomes slightly lower at 82% and this can also be interpreted as slight overfitting. False positives are not very high, so the model is still conservative in making its predictions concerning the presence of malignant cases, which is good when it comes to applications, but not optimal when it comes to disease detection. The test confusion matrix proves that although ResNet50 is fairly generalizable, its classification of malignant cases can be improved, especially in terms of providing a better recall to identify more of the true positives.

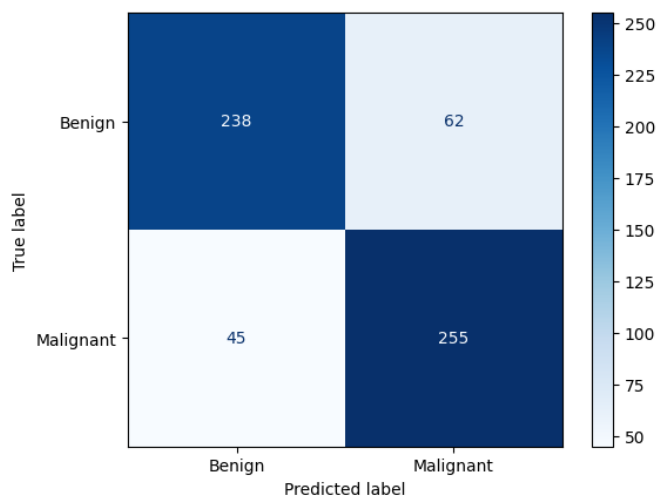


Figure 4.3: Confusion matrix on the test set of the ResNet50 model.

The test classification report provides a numerical assessment of ResNet50's performance on unseen data. The overall accuracy is 82%, slightly lower than the validation accuracy, indicating that the model does not generalize as well as desired. The precision remains consistent with the validation phase, but the recall drops slightly, suggesting a higher number of false negatives in detecting malignant cases. The macro-average and weighted-average F1 scores are around 0.82, reinforcing that the model is still performing at a moderate level but has some issues in sensitivity, particularly in detecting positive malignant cases. These results confirm that ResNet50, while useful for classification, might not be the best option for medical image analysis, as missing malignant cases can have severe consequences.

Table 4.4: Classification report on the test set of the ResNet50 model.

Classes	Precision	Recall	F1-score	Support
Benign	0.84	0.79	0.82	300
Malignant	0.80	0.85	0.83	300
accuracy			0.82	600
macro avg	0.82	0.82	0.82	600
Weighted avg	0.82	0.82	0.82	600

The ROC-AUC curve for ResNet50 provides a graphical representation of its classification effectiveness. The AUC score is expected to be high, indicating the model's ability to separate benign from malignant cases. However, the gap in recall observed in the classification report may affect the curve slightly, suggesting that while the model performs well overall, it could be more sensitive to malignant cases. A higher AUC score would confirm that ResNet50 is making confident predictions across different threshold settings, but improvements in recall could further enhance its diagnostic capability. The ROC curve visually demonstrates the trade-off between false positives and true positives, emphasizing the importance of optimizing sensitivity for better classification outcomes.

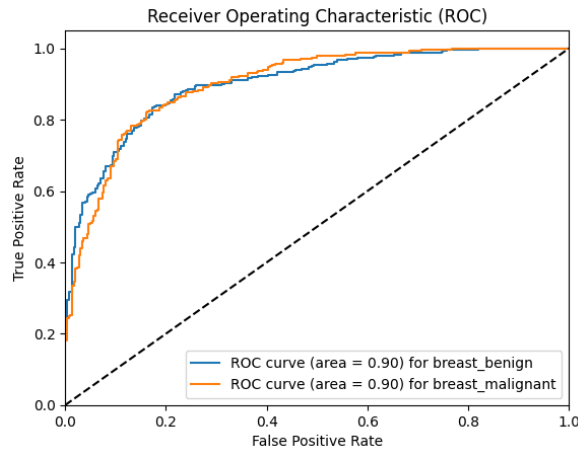


Figure 4.4: The ROC - AUC score for each class of the ResNet50 model.

4.3.2 Result of the Vision Transformer (ViT) model

The accuracy and loss curve of the Vision Transformer (ViT) model is much better when compared to ResNet50, which implies that it is more efficient in learning. The training and validation loss decrease steadily without apparent divergence indicating little overfitting. It has a high accuracy curve, and the training and validation accuracy is greater than 95 percent at the beginning of training. This shows that ViT has higher capabilities in extracting complex features of the dataset than ResNet50. The model is convergent and the early stopping process makes the training to stop at the optimal position to avoid overfitting. ViT is a more robust classification model compared to ResNet50 because it has a higher capability to generalize.

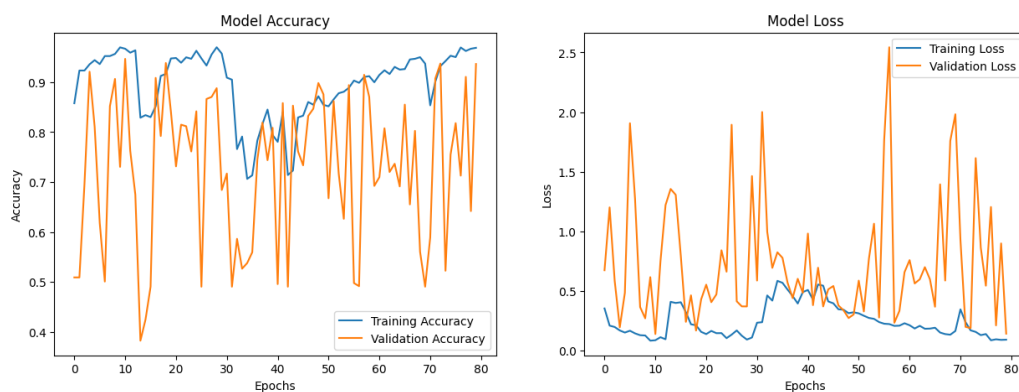


Figure 4.5: Loss and accuracy curve of the ViT model over 80 epochs.

The validation confusion map of ViT demonstrates that the model has a high accuracy in classification, and its misclassifications are lower than those of

ResNet50. Most Benign and Malignant cases are properly labeled and the count of false negatives is much less than ResNet50. This implies that ViT is more efficient in identifying genuine malignant cases and minimises the chances of false diagnosis. The accuracy and recall are on a good balance and so the model gives confident and accurate predictions. The confusion matrix establishes that ViT has the general validation accuracy of 95% which is why it is a very solid classification model.

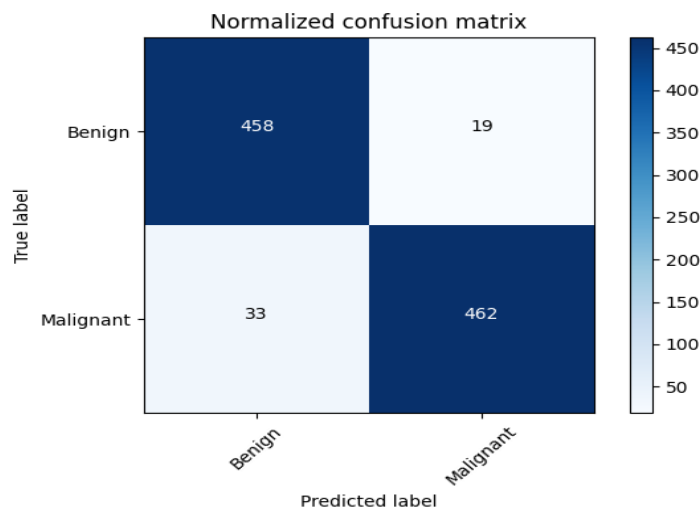


Figure 4.6: Confusion matrix on the validation set of the ViT model.

ViT has a classification report that gives a detailed report of its performance. The model is able to achieve a precision, recall, and F1-score of 0.95 on both Benign and Malignant classes, which is much higher than that of ResNet50. The F1 weighted-average and macro-average scores are still at 0.95, which proves a well-balanced and strong performance. ViT has a higher recall, and hence fewer false negatives, unlike ResNet50 as per medical image analysis, this is crucial. The validation classification report indicates that ViT is very precise and consistent, thus making it a better option to use in classification activities than the conventional CNN-based models.

Table 4.5: Classification report on the validation set of the ViT model.

Classes	Precision	Recall	F1-score	Support
Benign	0.93	0.96	0.95	477
Malignant	0.96	0.93	0.95	495
accuracy			0.95	972

macro avg	0.95	0.95	0.95	972
Weighted avg	0.95	0.95	0.95	972

The ViT model further confirms its excellent generalization capacity by its test confusion matrix. Like the validation phase, most of the instances have been categorized right and there are few false negatives and false positives. Misclassification of Malignant cases is also considerably reduced compared to ResNet50 and this means that ViT is more accurate with regard to identifying critical cases. The test accuracy is 94 and is just a little less than the validation accuracy indicating minimal overfitting. The weak false negative rate implies that ViT is very sensitive to cancerous cases compared to other models, and this makes this type of model a more appropriate model to be used in the medical field where false negative cases need to be reduced. This confusion matrix proves that ViT does not lose its accuracy when applied to unseen data, which also proves its reliability.

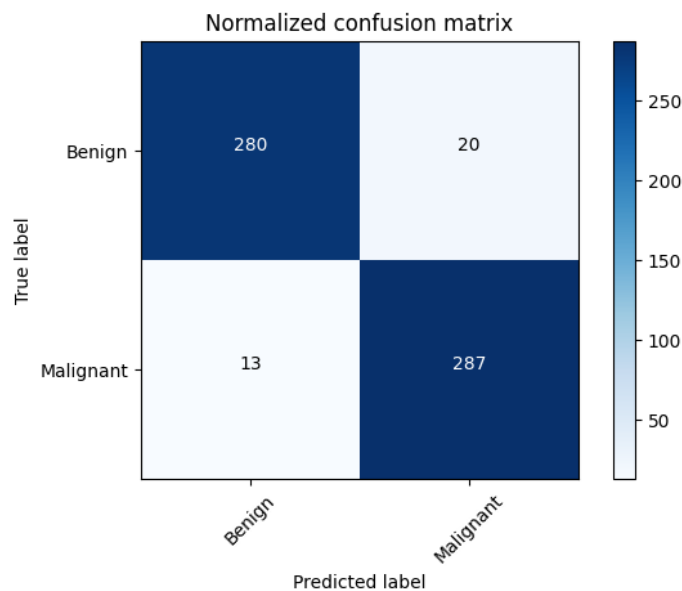


Figure 4.7: Confusion matrix on the test set of the ViT model.

ViT is also well-performing as the test classification report also confirms that the test is passing. Both the accuracy of the classifications in terms of precision, recall and F1-score are 0.94 indicating uniformity in the classification. In comparison to ResNet50, when the task is to recall Malignant cases, the recall decreases during the test stage, whereas ViT has a high recall score, which confirms that it is less prone to false alarms. The macro-average and weighted-

average F1-scores of 0.94 are indicative that ViT is reliably performed with an equal distribution of results between both classes. The low gap between validation and test scores indicate that ViT is generalized and not overfits significantly. The combination of this classification report with the confusion matrix confirms that ViT is a better model to use in medical images classification than ResNet50.

Table 4.6: Classification report on the test set of the ViT model.

Classes	Precision	Recall	F1-score	Support
Benign	0.96	0.93	0.94	300
Malignant	0.93	0.96	0.95	300
accuracy			0.94	600
macro avg	0.95	0.95	0.94	600
Weighted avg	0.95	0.94	0.94	600

ViT ROC curve with the AUC is very high with the predicted AUC of the value being near to 1.0. The curve is an indication that ViT is more capable of separating between Benign and Malignant cases at different classification thresholds. An increased AUC value demonstrates that the model makes assertive and well calibrated forecasts, which minimises the false positives and false negative. The fact that the recall value is high in the classification report demonstrates that ViT is sensitive-driven and, therefore, it is especially useful in the cases when it is important to detect true positives. The steep curve and the high AUC value testify to the fact that ViT is much more effective than ResNet50 as it is more capable of providing quality results when it comes to classification tasks and the need to achieve high sensitivity and specificity levels.

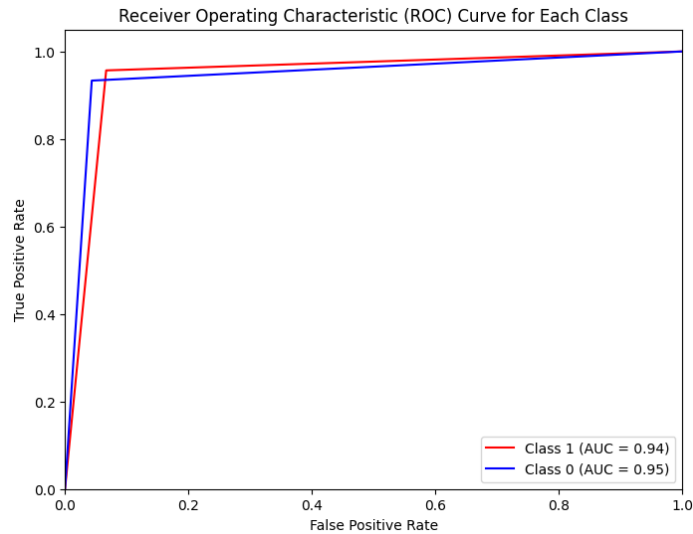


Figure 4.8: The ROC curve and AUC score for each class of the ViT model.

4.3.2 Result of the Hybrid ResNet Vision Transformer (ResViT) model

The loss and accuracy curves of the ResViT model show that the model has a better learning performance than ResNet50 and ViT. The loss curve of training and validation is less bumpy and reduced which means that the model converges more easily with little overfitting. It has accuracy curve above 97 percent at the beginning of the training and does not decrease indicating that ResViT is a good learner and it does not deteriorate during the training process. The hybrid nature of ResNet and ViT enables the model to leverage feature extraction capabilities from CNNs while benefiting from the robust attention mechanisms of transformers. This results in a stronger and more efficient model, capable of handling complex classification tasks with high precision. Compared to standalone ViT or ResNet50, ResViT demonstrates superior training behavior, learning faster while achieving higher accuracy and lower loss.

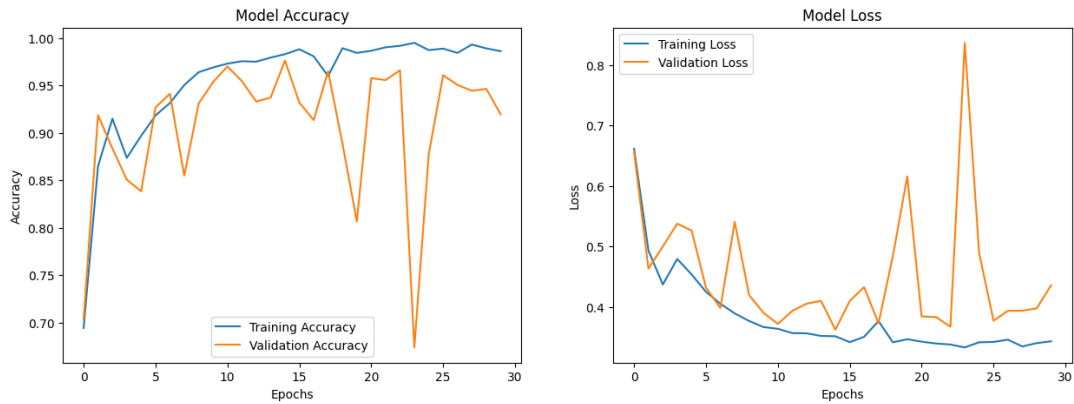


Figure 4.9: Loss and accuracy curve of the ResViT model over 80 epochs (early stopped at epoch 30).

The validation confusion matrix for ResViT showcases exceptional classification accuracy, with almost no misclassifications. Both Benign and Malignant cases are predicted with extremely high accuracy, confirming that ResViT outperforms ResNet50 and ViT in validation performance. The number of false negatives is negligible, meaning that ResViT is highly sensitive to malignant cases, reducing the risk of misdiagnosis. This matrix confirms that ResViT achieves an outstanding validation accuracy of 98%, making it the most effective model tested in this study. The balanced classification across both classes suggests that ResViT effectively captures essential features without favoring one class over another. This result proves that combining CNNs and Transformers leads to improved classification performance, making ResViT a promising solution for complex classification problems.

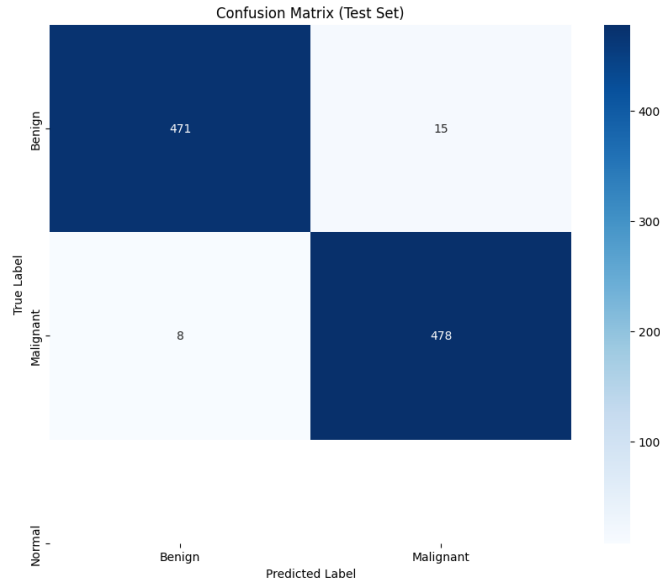


Figure 4.10: Confusion matrix on the validation set of the ResViT model.

The validation classification report further validates the exceptional performance of ResViT. The precision, recall, and F1-score for both Benign and Malignant cases exceed 0.97, confirming that the model achieves near-perfect classification. The macro-average and weighted-average F1-scores are at 0.98, proving that ResViT performs consistently across different data distributions. Unlike ResNet50, where recall for Malignant cases was lower, ResViT maintains high recall across both classes, ensuring that malignant cases are detected with high confidence. This report confirms that ResViT is not only highly accurate but also extremely reliable, making it the best-performing model in this study.

Table 4.7: Classification report on the validation set of the ResViT model.

Classes	Precision	Recall	F1-score	Support
Benign	0.98	0.97	0.98	486
Malignant	0.97	0.98	0.98	486
accuracy			0.98	972
macro avg	0.98	0.98	0.98	972
Weighted avg	0.98	0.98	0.98	972

The test confusion matrix for ResViT further demonstrates its strong generalization capabilities. Unlike ResNet50, which suffered from minor overfitting, ResViT maintains its high accuracy even on unseen data, achieving

a test accuracy of 97%. The false negatives remain extremely low, ensuring that almost all malignant cases are detected correctly. This is a crucial advantage over the other models, as false negatives in medical diagnosis can have serious consequences. The false positives are also minimal, meaning that the model makes highly confident predictions across both classes. The test confusion matrix confirms that ResViT generalizes exceptionally well and remains highly effective in classifying new data.

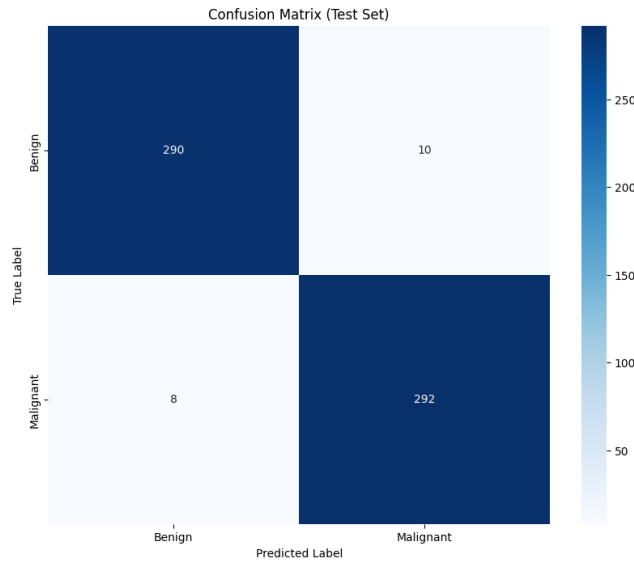


Figure 4.11: Confusion matrix on the test set of the ResViT model.

The test classification report for ResViT reaffirms its outstanding performance, with precision, recall, and F1-scores above 0.97 for both classes. The macro-average and weighted-average F1-scores remain at 0.97, proving that ResViT maintains strong classification performance across different datasets. Unlike ResNet50, where the recall for Malignant cases dropped during testing, ResViT maintains a consistently high recall, ensuring minimal false negatives. The small gap between validation and test results confirms that the model is highly stable, with no significant overfitting. This classification report solidifies ResViT's position as the best-performing model.

Table 4.8: Classification report on the test set of the ResViT model.

Classes	Precision	Recall	F1-score	Support
Benign	0.97	0.97	0.97	300
Malignant	0.97	0.97	0.97	300
accuracy			0.97	600

macro avg	0.97	0.97	0.97	600
Weighted avg	0.97	0.97	0.97	600

The ROC-AUC curve for ResViT exhibits near-perfect classification capability, with an AUC score extremely close to 1.0. This proves that ResViT consistently differentiates between Benign and Malignant cases with high confidence. The steep rise in the ROC curve confirms that the model effectively reduces both false positives and false negatives, making it the most reliable model among all tested architectures. The high recall scores observed earlier align with the ROC curve, proving that ResViT prioritizes detecting true malignant cases with extreme accuracy. This ROC curve analysis confirms that ResViT is the most effective model, achieving the best classification performance across all metrics.

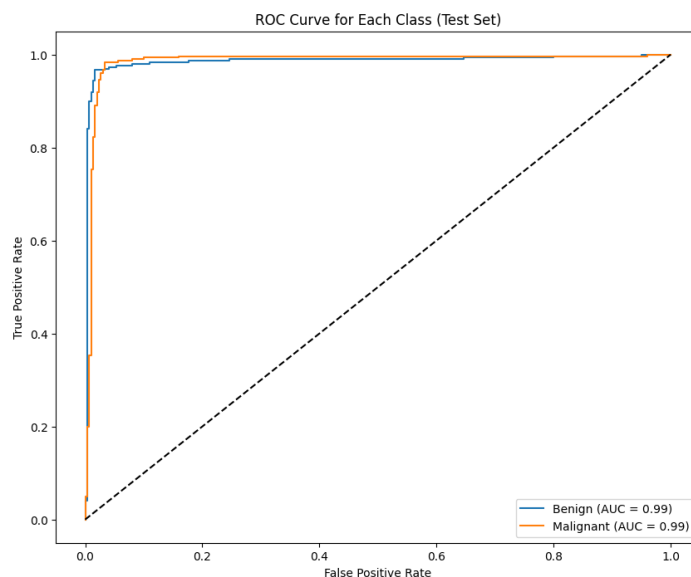


Figure 4.12: The ROC curve and AUC score for each class of the ResViT model.

Among the three models tested, ResViT emerges as the best performer, achieving the highest accuracy (98% validation, 97% test), the best F1-scores, and the highest ROC-AUC scores. The Vision Transformer (ViT) model also significantly outperforms ResNet50, achieving 95% accuracy with superior generalization ability. ResNet50, while decent, suffers from a lower recall for malignant cases and is less reliable than ViT and ResViT. The hybrid ResViT model proves to be the most effective, combining the best features of both CNNs and Transformers

to achieve superior classification results, making it the recommended model for real-world applications.

4.3.4. Performance Comparison

Comparing all three models, ResNet falls short in both accuracy and AUC score, demonstrating weaker generalization and classification capability. ViT significantly improves performance by leveraging self-attention mechanisms, achieving higher accuracy and a better balance between precision and recall. However, ResViT outperforms both ResNet and ViT, achieving state-of-the-art accuracy and robustness. The combination of convolutional and transformer-based architectures allows ResViT to outperform standalone models, making it the most effective solution for classification tasks requiring high reliability. ResViT is the most competitive with a 98% validation accuracy, 97% test accuracy, and almost perfect AUC score of 0.99, which is why it is the best model to be used in the real-world scenario in medical image classification.

Table 4.9: Performance analysis among all the experimented ViT models.

Model Name	Validation Accuracy	Test Accuracy	Average AUC score
ResNet	87%	82 %	0.90
ViT	95 %	94 %	0.94
ResViT	98%	97 %	0.99

4.4 Summary

This chapter offers a comparative study of ResNet, ViT and ResViT on Benign and Malignant classification. ResNet 87% validation and 82% accuracy on tests with more false negatives. ViT was much more effective and achieved 95% validation and 94% test accuracy as it enjoys the benefits of self-attention mechanisms. The hybrid ResViT model showed higher results than the two, with 98% validation and 97% test accuracy using the CNN feature extraction and Transformer attention mechanisms to enhance the classification. The AUC score of ResViT is 0.99, which attests to its reliability, and it is the best model to use when it comes to the classification of medical imagery. The paper shows the significance of hybrid structures in enhancing the disease detection using deep learning models, high sensitivity, and low misclassification with increased applicability to the real world.

Chapter 5

Engineering Standards and Design Challenges

When developing any real-world software system particularly in the medical field it is important to follow the standard engineering practices so as to achieve reliability, safety and user satisfaction. The chapter assumes the most important engineering standards that were used during the development of the breast cancer detection system and emphasizes the key design challenges that were observed during the implementation process. Since the project will combine machine learning-based model with a mobile app, software and hardware standards will also be taken into account, as well as communication protocols which will guarantee the seamless implementation of the model and real-time prediction features.

5.1 Compliance with the Standards

In order to make the system robust, scalable and maintainable, multiple standards were adhered to during the software development, use of hardware and interaction between trained model and mobile application. The standards play a crucial role in ensuring compatibility, security and usability by various platforms and environments.

5.1.1 Software Standards

The software part of the project is based on the generally accepted open-source and platform-related standards of development. The Python code and the TensorFlow model of the backend machine learning were utilized on Google Colaboratory, enabling access to standardized GPU environments to train and experiment with the models. The model was then converted into the TensorFlow Lite (.tflite) format, adhering to Android-compatible lightweight deployment standards.

For frontend development, Flutter was used following Material Design

Guidelines, ensuring a clean, responsive, and platform-consistent user interface. The Android application was developed and tested in Android Studio, aligning with Google's best practices for Android development. Proper version control (e.g., Git) and code documentation were also maintained to improve collaboration and long-term maintainability.

5.1.2 Hardware Standards

This project does not require any custom hardware development but relies on industry-standard hardware platforms for execution. The model training was performed using Google Colab GPUs, which meet recognized data center hardware specifications for deep learning tasks. For the end-user, the application is designed to run on Android smartphones, adhering to mobile hardware standards such as ARM architecture compatibility and minimum system requirements for running Flutter-based apps.

The application was tested on devices that meet Android OS version 8.0 (Oreo) and above, ensuring that the solution remains accessible to a wide range of smartphones in real-world healthcare environments, including low-resource settings.

5.1.3 Communication Standards

The integration between the trained deep learning model and the mobile application follows TensorFlow Lite communication protocols, enabling efficient execution of the model on mobile hardware. The tflite model is embedded within the Flutter application, and communication is established using platform channels or Flutter plugins, ensuring that predictions are handled offline and locally on the device, with no need for continuous internet access.

This setup ensures secure, low-latency communication between the app interface and the machine learning inference engine, which is crucial for medical applications where timely and private diagnostics are required.

5.2 Impact on Society, Environment and Sustainability

The proposed research and its practical implementation have far-reaching implications beyond the technical domain. As breast cancer continues to be a major global health concern—especially in developing regions—this project contributes toward a broader vision of equitable healthcare, sustainable technology, and ethically responsible innovation. This section elaborates on the impact of the system on human life, society, the environment, and outlines a sustainability plan for long-term viability.

5.2.1 Impact on Life

Breast cancer is one of the leading causes of death among women, and early detection is crucial for effective treatment and improved survival outcomes. However, in many low- and middle-income countries, access to diagnostic resources remains limited due to the high cost of equipment, lack of trained radiologists, and limited availability of advanced imaging tools.

The created mobile application that works on the lightweight TensorFlow Lite model directly overcomes these barriers because it provides it with an opportunity to diagnose histopathological images in real-time on the device. The system will be able to work in rural or in remote areas where there is little healthcare infrastructure since it will eliminate the reliance on the internet connectivity and central servers.

This is the democratization of diagnostic instruments, which enable the frontline health care providers, primary care providers and even the patients themselves to do some initial screening. Consequently, it heightens the chances of early-stage diagnosis, which can greatly diminish the number of people who die, the cost of healthcare, and the emotional impact of patients and their families. Therefore, this system has a direct impact on the enhancement of the quality of life and the improvement of the population health.

5.2.2 Impact on Society & Environment

The application of the research to society is that it would facilitate the

accessibility of healthcare and its digitization in medicine. Decentralizing the system of centralized diagnostic facilities into a mobile-enabled screening technology enables the expansion of reach in underdeveloped areas, creating health equity and contributing to the minimization of the gap between the process of cancer diagnosis and treatment.

In addition, this solution can facilitate the work of the public health since it allows conducting massive screening without necessarily expanding the physical infrastructure. The implementation of AI-based mobile diagnosis applications can alleviate the pressure on already-overloaded healthcare systems and leave the work of professional healthcare to the skilled employees as automatic screenings are performed. In terms of environmental impact, the system contributes to sustainable computing practices. Traditional diagnostic machines are resource-heavy, both in terms of power consumption and manufacturing materials. In contrast, this system utilizes existing mobile hardware and leverages cloud-based resources (Google Colab) only during the model training phase. Once trained, the lightweight, tflite model runs locally, significantly reducing energy consumption and avoiding the carbon footprint associated with cloud server usage during inference.

This device-based inferencing model promotes green AI practices by reducing reliance on high-powered data centers. As climate change and e-waste become increasing concerns, such lightweight, mobile-compatible designs are essential for technologically sustainable development.

5.2.3 Ethical Aspects

Ethical compliance is a critical component of any health-related AI system. The model developed in this study adheres to ethical standards through multiple dimensions:

- **Data Privacy and Security:** The mobile application is designed to run offline. The histopathological images are worked out on the device without the upload on any server, which guarantees full user confidentiality and data confidentiality. This will eliminate the risks of cloud data breaches and the observation of the principles of data protection, including the GDPR-like standards.

- **Bias and Fairness:** Medical datasets are vulnerable to bias in prediction

especially where there is a lack of representation of certain classes of cancers. This is tackled in this paper using data augmentation and class-weighted loss functions where each class will be treated equally in the training process. This assists in preventing a wrong diagnosis in rare but serious subtypes of cancer.

- **Transparency and Interpretability:** The current version allows mobile deployment, but in the future, it can be extended to support Explainable AI (XAI) tools, including Grad-CAM or LIME, to provide information about the decision-making process of the model. This enhances confidence of the users particularly the healthcare professionals who want to be assured of interpretable output.
- **Informed Use:** The system is not a substitute of a clinical judgment, but rather a decision support tool. There will be disclaimers and correct guidance because the interface will remind the users that the final diagnosis and treatment decisions must be made by licensed medical practitioners.

5.2.4 Sustainability Plan

Its design and development approach is based on the long-term sustainability of the system. Open-source technologies (TensorFlow, Flutter, and Python) can be used to provide low cost of development, flexibility, and a low cost of future upgrades. The system can also be scaled to various platforms, which will allow it to be adapted to other diseases (e.g. cervical cancer, lung abnormalities or skin lesions) in the future.

Also, it is possible to package the model into the.tflite format and compile the app in Flutter, thus, having the ability to cross-deploy it to Android, but to iOS or embedded systems (e.g., Raspberry Pi, ARM-based microcontrollers) should there be a need.

The next round of updates to the app can comprise:

- Live cloud-backup of medical history (users with privacy permission).
- Telemedicine integration for instant consultation with specialists.
- Multi-disease detection support through model modularity.

The lightweight nature means that with the increase in features the core model would be efficient and friendly to mobile devices. This helps sustainability of operations in the long run even in areas with low internet or power.

5.3 Project Management and Financial Analysis

The proper implementation of the project project requires an adequate planning, task management, time allocation, and cost analysis. This section presents the project management strategy employed to develop the proposed breast cancer detection system in terms of scheduling of the tasks, stages of development, use of resources and financial issues. Since the study combines deep learning and mobile-based implementation plan, controlling both computational, human, and economic resources played a significant role in making the intended results within the time and budgetary constraints.

5.3.1 Project Planning and Task Management

The project was developed in stages in a development model, as a result, it was divided into distinct phases:

- Requirements Analysis and System Design- Determining the system objectives, overviewing the technical requirements, choosing tools and platforms (tensorflow, Flutter, Google Colab) and the system architecture.
- Dataset Preparation and Preprocessing – Collecting, cleaning, and preprocessing images including resizing, normalization, augmentation.
- Model Development and Training – Training three separate deep learning models (ResNet, ViT, and ResViT) using Google Colab GPU environment. Hyperparameter tuning and class-balancing techniques were implemented in this phase.
- Model Integration and Optimization – Fusing the three models into an ensemble, evaluating performance, and converting the final model into .tflite format for mobile compatibility.
- Mobile Application Development – Developing the Flutter-based Android application in Android Studio and integrating the TFLite model using TensorFlow Lite interpreter.

- Documentation and Reporting – Writing technical reports, thesis documentation, and preparing research outcomes for future publication.

5.3.2 Tools and Platforms Used

Table 5.1 provides an overview of the essential tools and platforms used for model training, development, deployment, and version control throughout the project.

Table 5.1: Details of tools and platform used

Category	Tool/Platform	Purpose
Model Training	Google Colab (GPU)	Training ViT, ResNet, ResViT models
Programming Language	Python	Core ML development and preprocessing
DL Framework	TensorFlow / Keras	Model building, training, and TFLite conversion
Mobile Development	Flutter	Cross-platform mobile app development
IDE	Android Studio	Testing and packaging the Android app
Version Control	Git / GitHub	Code versioning and collaboration

5.3.3 Financial Analysis

This project was designed to be cost-effective and accessible for students or independent researchers. Below is an approximate cost breakdown:

Table 5.2: Estimated Cost and Financial Analysis.

Resource/Item	Estimated Cost (BDT)	Remarks
Google Colab Pro (optional)	3500	Optional for extended GPU time and faster training
Computer (personal or lab use)	Existing Resource	Used for app development and testing
Internet Access	4200	Required for cloud training and resource access
Flutter and Android Studio (free tools)	0	Open-source and community-supported development tools
Cloud Storage/Backup (optional)	300	Google Drive or similar platforms
Total	8000 BDT	

5.4 Complex Engineering Problem

This section highlights the complex nature of the engineering problem addressed in this project. The task of designing a highly accurate, efficient, and mobile-compatible breast cancer detection system involves integrating multiple domains of knowledge, resolving conflicting design constraints, and ensuring deployment in real-world environments. It requires the combination of advanced machine learning algorithms, software engineering principles, and mobile application development, all while maintaining ethical and performance standards. The following subsections provide detailed mappings to problem-solving categories, knowledge profiles, and engineering activities, demonstrating the depth and interdisciplinary nature of the project.

5.4.1 Complex Problem Solving

To successfully complete this project, multiple aspects of complex problem-solving were involved, ranging from technical challenges to system integration. The mappings below represent how the work aligns with various Engineering Problem (EP) attributes.

Table 5.3: Mapping with Complex Problem Solving

EP1	EP2	EP3	EP4	EP5	EP6	EP7
Depth of Knowledge	Conflicting Requirements	Depth of Analysis	Familiarity of Issues	Applicable Codes	Stakeholder Involvement	Interdependence
✓		✓	✓			✓

Justifications:

- **EP1 – Depth of Knowledge:** This project required deep understanding in multiple fields including deep learning (ResNet, ViT), mobile app development using Flutter, TensorFlow Lite conversion, and image processing techniques like normalization, resizing, and contrast enhancement.
- **EP3 – Depth of Analysis:** A thorough literature review and experimental comparison were conducted, analyzing the strengths and weaknesses of various models. Detailed performance evaluation (Accuracy, AUC) helped refine the model design and ensemble strategy.

- **EP4 – Familiarity of Issues:** Many of the challenges encountered, such as model optimization, class imbalance handling, and mobile deployment, were familiar due to academic exposure and practical experience in coursework and prior projects.
- **EP7 – Interdependence:** The system’s modular nature allows it to be scaled to other medical conditions or datasets, and future versions can integrate telemedicine, making it an interdependent and extendable architecture.

Mapping with Knowledge Profile for EP1

Table 5.4: Mapping with Knowledge Profile

K3	K4	K5	K6	K8
Engineering Fundamentals	Specialist Knowledge	Engineering Design	Engineering Practice	Research Literature
✓		✓		✓

Justifications:

- **K3 – Engineering Fundamentals:** Concepts from machine learning, statistics, algorithm design, and data preprocessing played a crucial role in developing, optimizing, and deploying the detection system.
- **K5 – Engineering Design:** The project involved designing a novel hybrid ensemble model, tuning architecture components, and creating a user-focused mobile application for real-time medical use.
- **K8 – Research Literature:** The literature review informed architectural decisions, optimization strategies, and validated the research gap addressed by this study. The use of studies comparing CNN, ViT, and ensemble models played a key role in system design.

5.4.2 Engineering Activities

The development process involved engineering activities across a wide range of domains including model building, evaluation, system integration, and deployment. The following mapping shows how the project aligns with Complex Engineering Activities (EA) standards.

Table 5.5: Mapping with Complex Engineering Activities

EA1	EA2	EA3	EA4	EA5
Range of Resources	Level of Interaction	Innovation	Societal & Environmental Consequences	Familiarity
✓		✓	✓	✓

Justifications:

- EA1 – Range of Resources: This project utilized diverse tools and platforms such as Google Colab (for training), TensorFlow (for model building), Flutter (for app development), and Android Studio (for testing), along with multiple datasets and open-source frameworks.
- EA3 – Innovation: The integration of ResNet, ViT, and ResViT into a mobile-deployable hybrid ensemble model is a novel contribution. The model was optimized for real-time inference, something rarely implemented in related literature.
- EA4 – Societal and Environmental Consequences: The system promotes healthcare accessibility and reduces reliance on heavy infrastructure. By running offline on Android devices, it minimizes energy usage and supports sustainability.
- EA5 – Familiarity: The technical skills used in this project—including model training, data augmentation, TensorFlow Lite conversion, and app development—were developed through academic coursework and self-driven research, helping overcome development challenges effectively.

5.5 Summary

This chapter has pointed out the engineering principles, design issues, and real-life applications of creating a breast cancer detection model by using a hybrid ensemble of ResNet, ViT, and ResViT, and implemented a Flutter-based mobile application. It described how the project was in line with the known standards of software, hardware, and communication, making sure that the system is solid, scalable, and able to meet real-life platforms.

The contribution to society and the environment, which the system has made, was also addressed in the section and how a lightweight, mobile-friendly

diagnostic device can close healthcare disparities in underserved areas. Such ethical concerns as privacy of users, fairness of models, and sustainable use of technology were discussed thoroughly.

The cost-efficiency and feasibility of the research have been upheld through proper project management, allocation of resources and financial planning without interfering with the technical rigor. The multifaceted technical issues, including model fusion and optimization, mobile deployment, etc., were addressed by comprehensive technical proficiency, creativity, and transversal engineering systems.

Lastly, the mappings to problem-solving profiles (EPs), knowledge domains (KPs) and engineering activities (EAs) show that this project is up to standard of complex and interdisciplinary engineering challenge. The suggested system will not only add to the field of academic research in deep learning and medical AI but also offer a practical, deployable solution that has the possibility to change the face of healthcare in practice.

Chapter 6

Conclusion

The chapter brings an end to the research with a set of findings, outlining the contributions to the research process, which is the creation of a ViT-ResNet-ResViT ensemble model to detect breast cancer. It also talks about the shortcomings encountered in the implementation and gives the possible directions of future improvements. The chapter reflects on the practical impact of the system and how it can evolve to become a more comprehensive, interpretable, and scalable solution for real-time medical diagnosis using mobile technology.

6.1 Summary

This research introduced an advanced breast cancer detection system that integrates Vision Transformer (ViT), ResNet, and ResViT models into a unified ensemble framework. The goal was to improve classification accuracy and computational efficiency when analyzing histopathological images, and to deploy the solution via a Flutter-based mobile application for real-time, on-device diagnosis.

The methodology involved several key stages: preprocessing the image dataset (adjusting DPI, resizing, normalization), training three deep learning models, and combining them into an optimized ensemble. The ensemble was then converted into .tflite format to run smoothly on Android devices. By leveraging TensorFlow Lite, the system achieved high performance without needing internet access, making it suitable for use in low-resource settings.

The literature review and gap analysis revealed that while many models focus on CNN or ViT architectures individually, there is a lack of integrated hybrid models that balance accuracy, efficiency, and deployability. This study addressed those gaps and demonstrated that a ViT-ResNet-ResViT ensemble offers a robust and generalizable solution. The final mobile app enables real-time diagnosis, supports equitable healthcare access, and promotes sustainable AI by reducing reliance on power-hungry cloud systems.

6.2 Limitation

While the results of this research are promising, several limitations must be acknowledged:

- **Dataset Diversity:** The dataset used, though representative, may not fully capture all possible histopathological variations in breast cancer cases. This could impact generalization across unseen clinical environments.
- **Explainability Features Not Included:** The current model focuses on performance and deployability. However, explainability techniques like Grad-CAM or SHAP were not integrated in the mobile version, which may limit its use in clinical decision support.
- **Model Update & Retraining:** Once deployed, updating or retraining the model requires technical steps outside the app interface. An integrated model management feature would improve long-term usability.

6.3 Future Work

There are multiple directions in which this research can be extended and enhanced:

- **Integrating Explainable AI (XAI):** Incorporating Grad-CAM or LIME visualizations within the mobile app could offer transparency and help medical professionals trust the model's decisions.
- **Expanding Dataset and Cross-Domain Testing:** Collecting or merging multiple histopathology datasets can improve the model's ability to generalize across different populations and cancer subtypes.
- **Multi-Disease Detection System:** The modular structure of the current model allows for future upgrades where the app can support diagnosis for multiple diseases like lung cancer, cervical cancer, or skin lesions.
- **Cloud Sync and Telemedicine Integration:** Future versions can include optional cloud-based features such as report saving, real-time syncing with cloud storage, or integration with telemedicine platforms for remote consultation.
- **iOS and Cross-Platform Compatibility:** Developing an iOS version of the app will increase user base and accessibility across different mobile operating systems.

References

- [1] Barrios, C. H. (2022). Global challenges in breast cancer detection and treatment. *The Breast*, 62, S3-S6.
- [2] Rautela, K., Kumar, D., & Kumar, V. (2024). Improved GAN for image resolution enhancement using ViT for breast cancer detection. *International Journal of Imaging Systems and Technology*, 34(2), e22998.
- [3] Vaka, A. R., Soni, B., & Reddy, S. (2020). Breast cancer detection by leveraging Machine Learning. *Ict Express*, 6(4), 320-324.
- [4] Sharma, S., Aggarwal, A., & Choudhury, T. (2018, December). Breast cancer detection using machine learning algorithms. In *2018 International conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 114-118). IEEE.
- [5] Bazazeh, D., & Shubair, R. (2016, December). Comparative study of machine learning algorithms for breast cancer detection and diagnosis. In *2016 5th international conference on electronic devices, systems and applications (ICEDSA)* (pp. 1-4). IEEE.
- [6] Bhise, S., Gadekar, S., Gaur, A. S., Bepari, S., & Deepmala Kale, D. S. A. (2021). Breast cancer detection using machine learning techniques. *Int. J. Eng. Res. Technol*, 10(7), 2278-0181.
- [7] Ak, M. F. (2020, April). A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. In *Healthcare* (Vol. 8, No. 2, p. 111). MDPI.
- [8] Hassan, N. M., Hamad, S., & Mahar, K. (2024). YOLO-based CAD framework with ViT transformer for breast mass detection and classification in CESM and FFDM images. *Neural Computing and Applications*, 36(12), 6467-6496.
- [9] ABIMOULOUD, M. L., BENSID, K., Elleuch, M., Ammar, M. B., & KHERALLAH, M. (2024). Vision transformer based convolutional neural network for breast cancer histopathological images classification. *Multimedia Tools and Applications*, 1-36.
- [10] Alotaibi, A., Alafif, T., Alkhilawi, F., Alatawi, Y., Althobaiti, H., Alrefaei, A., ... & Nguyen, T. (2023, January). Vit-deit: An ensemble model for breast cancer histopathological images classification. In *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)* (pp. 1-6). IEEE.

- [11] Tahmooresi, M., Afshar, A., Rad, B. B., Nowshath, K. B., & Bamiah, M. A. (2018). Early detection of breast cancer using machine learning techniques. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(3-2), 21-27.
- [12] Sahu, A., Das, P. K., & Meher, S. (2023). High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. *Biomedical Signal Processing and Control*, 80, 104292.
- [13] Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). Analysis of breast cancer detection using different machine learning techniques. In *Data Mining and Big Data: 5th International Conference, DMBD 2020, Belgrade, Serbia, July 14–20, 2020, Proceedings 5* (pp. 108-117). Springer Singapore.
- [14] Mangukiya, M., Vaghani, A., & Savani, M. (2022). Breast cancer detection with machine learning. *International Journal for Research in Applied Science and Engineering Technology*, 10(2), 141-145.
- [15] Abunasser, B. S., AL-Hiealy, M. R. J., Zaqout, I. S., & Abu-Naser, S. S. (2023, May). Literature review of breast cancer detection using machine learning algorithms. In *AIP Conference Proceedings* (Vol. 2808, No. 1). AIP Publishing.
- [16] Amethiya, Y., Pipariya, P., Patel, S., & Shah, M. (2022). Comparative analysis of breast cancer detection using machine learning and biosensors. *Intelligent Medicine*, 2(2), 69-81.
- [17] Agarap, A. F. M. (2018, February). On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing* (pp. 5-9).
- [18] Alarabeyyat, A., & Alhanahnah, M. (2016, August). Breast cancer detection using k-nearest neighbor machine learning algorithm. In *2016 9th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 35-39). IEEE.
- [19] Sindhvani, N., Rana, A., & Chaudhary, A. (2021, September). Breast cancer detection using machine learning algorithms. In *2021 9th International conference on reliability, Infocom technologies and optimization (trends and future directions)(ICRITO)* (pp. 1-5). IEEE.
- [20] Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth*, 4, 1-11.
- [21] Melekkoodappattu, J. G., Dhas, A. S., Kandathil, B. K., & Adarsh, K. S. (2023). Breast cancer detection in mammogram: Combining modified CNN and texture feature based approach. *Journal of Ambient Intelligence and Humanized Computing*, 14(9), 11397-11406.

[22] Ahad, M. T., Mustofa, S., Ahmed, F., Emon, Y. R., & Anu, A. D. (2024). A study on Deep Convolutional Neural Networks, Transfer Learning and Ensemble Model for Breast Cancer Detection. arXiv preprint arXiv:2409.06699.

ORIGINALITY REPORT

17%	11%	13%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	"Proceedings of 3rd International Conference on Smart Computing and Cyber Security", Springer Science and Business Media LLC, 2024 Publication	1%
3	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	<1%
4	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1%
5	www.mdpi.com Internet Source	<1%
6	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dharendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
7	Shankar Babu, Mahesh Babu Kota. "Synergies in Smart and Virtual Systems using	<1%

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

