Leveraging Machine Learning to Tailor Breast Cancer Treatment Plans

Final Year Design Project

By
Raktim Kumar Mondal
213-15-4287

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. Saiful Islam
Assistant Professor
Department of Computer Science and
Engineering Daffodil International
University

Co-Supervised by
Dr. Naznin Sultana
Associate Professor
Department of Computer Science and
Engineering Daffodil International
University

**DAFFODIL INTERNATIONAL
UNIVERSITY**
Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled **Leveraging Machine Learning to Tailor Breast Cancer Treatment Plans,** submitted by **Raktim Kumar Mondal,** ID No: **213-15-4287** 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 (FNN)
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

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

Internal Examiner

17.09.25

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

Internal Examiner

17.9.25

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

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Saiful Islam, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:
03.09.25

Mr. Saiful Islam
Assistant Professor
Department of Computer Science and
EngineeringDaffodil International
University

Co-Supervised by:

Dr. Naznin Sultana
Associate Professor
Department of Computer Science and
EngineeringDaffodil International
University

Submitted by:
Raktim

Raktim Kumar Mondal
Student ID: 213-15-4287
Department of Computer Science and
EngineeringDaffodil International
University

ACKNOWLEDGEMENTS

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

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

We are grateful and wish our profound indebtedness to **Mr. Saiful Islam Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Leveraging Machine Learning to Tailor Breast Cancer Treatment Plans** 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 themat all stages have made it possible to complete this project.

We would like to express our 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.

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

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Breast cancer is a major cause of cancer mortality in women and therefore early and precise detection is essential in treatment. The traditional diagnostic techniques like mammography and histopathology are still popular and are constrained by the variability of the observer and sensitivity in complicated cases. Convolutional Neural Networks (CNNs) Deep learning has become a potent remedy, being able to extract features automatically with high accuracy on classification. Nevertheless, there are still issues in small data sets, binary only classification, and low interpretability. This paper will suggest a hybrid architecture combining CNNs and Capsule Network to enhance histopathology-based breast cancer detection. The results have been assessed using 6,000 images preprocessed and tested on various architectures, which are EfficientNetB0, InceptionV3, ResNet variants, and MobileNet V2. Among them, CapsuleNet model had the highest accurate, precise, recall and F1-score of 96.44 that surpassed any baselines tested. Gradient-weighted Class Activation Mapping (Grad-CAM) was also used to make the results more interpretable, giving heatmaps that show which parts of the image affected the models in the results, and this particular aspect should give the results more transparency in clinical practice. To be implemented practically, the optimized CapsuleNet model was translated to a mobile-based application with the help of TensorFlow Lite. This provides accessibility in resource constrained healthcare settings and proves the possibility of integrating in real world settings. Since it integrates good diagnostics and explainability and portability, the proposed solution can be useful in academic studies and clinical practice, helping to diagnose diseases earlier, minimize false diagnosis, and lead to better patient outcomes.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents the general research field and shows the significance of breast cancer detection at an early age and with high accuracy. It spells out the rationale of the research, the research goals and the research methodology to fill the gaps identified. The chapter also gives the projected results and the structure of the thesis to inform the reader on the following chapters.

1.1 Introduction

Breast cancer is a fatal condition that is highly common in women across the world. World Health Organization also mentions that every year, millions of new cases are diagnosed, which, along with being a medical problem, is also a social and economic one. The significance of early and accurate diagnosis is critical in the reduction of deaths since the outcome of the treatment is highly determined by the quality of the diagnosis. The traditional diagnostic methods such as mammography, ultrasound and histopathology continue to employ the gold standard although they are restricted in terms of inter-observer and misclassify complex cases and reduce sensitivity at the earlier stages of cancer. This has been the cause of these struggles which explain why the development of ingenious systems that could assist clinicians in making more consistent and faster diagnosis is necessary [1].

It has already been noted that machine learning (ML) and deep learning (DL) are flourishing tools in medical image analysis over the last few years. Specifically, the Convolutional Neural Networks (CNNs) have been highly effective in the step of determining the latent behavior of the imaging information, and it is a very good performer compared to the conventional approaches [3]. Many researchers have utilized transfer learning models such as VGG, ResNet, DenseNet and EfficientNet in the breast cancer classification task with promising results [10]. It has also been explored to apply hybrid structures with CNNs such as LSTM or Capsule networks, which would be in a better position to analyze both spatial and context information [13]. Also, explainability models like Grad-CAM have been actively applied in providing this kind of visual information of model decisions which is crucial in developing trust in medical usages [17]. Still, despite these advances, small dataset size, binary-based classification, uninterpretable models, and lack of interest in the real-world use remain a problem in various studies [15].

These technologies can be implemented in the healthcare sector, and it is the possibility to improve the breast cancer diagnosis planning and treatment process. With the help of CNN models and Capsule Networks, the proposed study will not only aim at providing the enhanced classification accuracy but also at ensuring that the cost of the computations would be reduced. Moreover, the interpretation through the use of Grad-CAM is more understandable hence rendering the system more transparent and more clinically reliable. Most importantly, the integration of the trained model into a mobile application will ensure that the solution will not be confined to the laboratory research but can be applied to the practice in the real health care centres. This kind of alignment between machine learning applications of high technology and real clinical activities paves the way to establishing the development of intelligent systems that are not only accurate but are also safe and cheap and dependable [24].

1.2 Motivation

Breast cancer is said to be one of the leading causes of cancer related death in the world and her growing rate demands the need to be diagnosed properly and at an early stage. Even though traditional techniques involve imaging and pathology-based techniques are useful; they are more dependent on expert interpretation hence resulting to inconsistency and failure to detect complex cases. Research evidence shows that machine learning models, particularly CNNs, are better than conventional models because they acquire discriminative characteristics on medical images automatically [3]. However, much of this research is limited by their focus on dichotomous benchmarking of benign and malignant cases or small and unbalanced data sets [5]. This limits their application in clinical practice. This gap is the motivating factor to make more powerful and more adaptive models that do not merely characterize the cancer of the breast, but also provide a more profound insight into several stages or subtypes of the cancer disease.

The other significant force is to render these high-level technologies to be more acceptable and reliable to the health practitioners. Hybrid machine learning models have also been proven to have potential in improving diagnostic accuracy and to encourage complex spatial patters, such as CNNs with LSTM or capsule networks [22]. In the meantime, interpretability methods, including Grad-CAM, are gaining importance so that clinicians can understand what models are predicting, therefore, building confidence in AI-assisted decision-making [17]. Despite the promising nature of these directions, little has been done in terms of translating these models into practical tools, such as mobile-based applications that can support doctors and patients in other settings other than research [24]. The thesis therefore tries to address these gaps by developing an ensemble architecture that uses CNN and capsule networks, the use of interpretability through Grad-CAM, and the execution of the architecture on a mobile platform that will be utilized in practice.

1.3 Objectives

The primary idea of this thesis is to create and test a machine learning-based framework, enhancing the process of breast cancer detection and helping to create individual treatment plans. In contrast to traditional methods of diagnostic procedures that tend to be lengthy and rely on human judgment, this paper will attempt to offer a legitimate, explicable, and convenient resolution based on the sophisticated machine learning methods. To carry out this, the following objectives will be used to inform the study:

1. To create a powerful and interpretable hybrid framework that can detect breast cancer using Convolutional Neural Networks (CNNs) and Capsule Networks and reinforce the framework with explainable AI methods (Grad-CAM) to provide the accuracy, robustness, and clinical confidence.
2. To use the optimized framework as a mobile based application to offer accessible, portable, and real time diagnostic support, particularly in some resource constrained healthcare environment.

1.4 Methodology

The research methodology would be directed towards the systematic growth and testing of the machine learning model of detection and treatment planning of breast cancer. It begins with data collection and preprocessing whereby medical images are gathered and treated through resizing, normalization, and augmentation to equalize the representation and eliminate overfitting. After this, the machine learning model frameworks (CNN-based, ResNet, VGG, DenseNet and EfficientNet) are trained through the transfer learning to exploit the already available knowledge on the large-scale datasets. A hybrid ensemble strategy that combines the ability of CNNs and Capsule Networks is then used in order to improve the determination of features, spatial sensitivity, and precision of recognition by different types of breast cancer.

After constructing the models, stringent evaluation phase is conducted depending on the performance measures of accuracy, precision, recall and F1-score where a fair comparison is made between the baseline CNNs and recommended ensemble. In order to give trust and transparency, explainable AI techniques like Grad-CAM are included whereby one can be able to see the spots in the medical images that influence the results given by the model. This is of particular importance to clinical adoption, as it assists in bridging the gap between the black-box AI systems and the interpretability by humans. Finally, the optimized model is applied on the mobile-based application that is portable and available in the actual healthcare environment. This practical use does not just authenticate the research finding, but it also provides an avenue to clinical incorporation in the future.

1.5 Project Outcome

This project will provide a valid and effective machine learning system to facilitate early diagnosis and personalized treatment planning of breast cancer. This study will focus on the flaws of the traditional methods, which are most of the times time consuming, costly and highly dependent on human factor and therefore are not accurate in diagnosis. The study is relevant as it provides a system which will be in a position to increase the diagnostic accuracy and reduce errors that result in improved clinical decision-making and improved patient outcomes.

As far as methodology is concerned, the project will aim at incorporating a number of machines learning architectures, including CNN-based architectures and ensemble methods, to achieve a hybrid model that will draw the merits of the other two. Transfer learning and data augmentation are used to train the models to detect minor features in the medical images, otherwise not easily detectable under the human eye. The ensemble of the methodology is more precise and stronger in comparison to the individual models. The explainable AI is another strength of the outcome since it enables elucidation of predictions, which is essential in building trust among healthcare professionals.

In practice, the designed system can be made to translate to a clinical environment and can be used as a support to oncologists and radiologists. The implementation of the model into a mobile based application increases the accessibility, particularly in those regions with minimal healthcare facilities. Such application can also be used in preliminary screening processes to ensure that the hospitals are not congested and those at risk are detected and sent to seek medical attention on time. The practical value of the research is highlighted in the practical implementation by the fact that it would transcend theoretical contributions and have direct influence on the healthcare.

Overall, it can be said that the project outcome is not only a contribution to the academic knowledge, but also an outline that provides a possible paradigm shift in the way breast cancer diagnosis and treatment planning are carried out. The research will result in a foundation on which enhanced AI systems will be incorporated in medical practice through the integration of error-precision, understandability, and accessibility. The net effects are in the reduction of the mortality rates, early interventions, and the provision of cost-effective measures to healthcare systems throughout the entire world. In this way, the research will meet scientific and social needs, as well as create value at various levels.

1.6 Organization of the Report

The thesis has six chapters, which reflect a specific aspect of the research work. The structure has been constructed in a logical and comprehensive manner to provide background, motivation, methodology, implementation and outcomes.

Chapter 1: Introduction - This chapter gives the motivation of the study by describing the field of research and the problem. It explains why the research was conducted, the proposed methodology, the expected outcomes of the project and provide a conclusion of the entire report.

Chapter 2: Background - This chapter presents the background knowledge and the exhaustive literature review which is necessary to qualify breast cancer using machine learning techniques. It discusses the existing procedures of CNN-based, approaches to similar procedures and the research studies. It is also the chapter that enumerates the big gaps in research that will be addressed by the proposed study and this will form the foundation of the given methodology.

Chapter 3: Research Methodology- This chapter gives the planned methodology in detail. It outlines the description of the dataset, data preprocessing (e.g., resizing, normalization and augmentation), training of the model according to CNN architecture and combination of CNN-based and Capsule Network-based models. The chapter also discusses evaluation strategies, Grad-CAM implementation to gain interpretability and how the mobile-based application will be deployed. System architecture, system design specifications, and workflow diagrams are included to make sure that there is the whole picture of the development process.

Chapter 4: Implementation and Results - The next chapter is dedicated to the practical work of the proposed framework. It gives an account of the training and testing environment, the measurements in the assessment and the work of the baseline CNN models and the hybrid ensemble. The results are analyzed, and the accuracy of the classification is evaluated and the insight provided by Grad-CAM visualization. The findings show the effectiveness of the proposed framework in the breast cancer detection and treatment planning.

Chapter 5: Engineering Standards and Design Challenges - The chapter discusses the engineering standards that were applied in the development of the system, including the use of the software tools and platforms of deployment. It also touches on the ethical concerns, the effects to the society and the environment, the sustainability concerns and the viability of the financial aspect. It also talks of the intricate engineering problems which were encountered during the research and how they moved to realize this.

Chapter 6: Conclusion - This is the final chapter that gives a summary of the entire research, the main findings, contributions, and the findings of the project. It further outlines the limitations that were faced in the study and offers suggestions as to the need to conduct further studies, particularly the idea of improved model generalization, data diversity increment and generalizing the application mobile to broader clinical uses.

Chapter 2

Background

This chapter gives the theoretical and practical background required to comprehend the research. It examines literature on the detection of breast tumors, feature extractors and classification techniques using machine learning and deep learning. The discussion also outlines the strengths and weaknesses of some of the approaches that have been used and the major gaps that have been determined to form the basis of the discussion in this study.

2.1 Introduction

Breast cancer has been regarded as one of the most dangerous health complications that plague women all over the world with a significant percentage of cancer deaths every year being as a result of breast cancer. According to the health statistics in the world, it is the most recurrently diagnosed cancer among women and also one of the leading childhood-related deaths. The early and more precise diagnosis is of paramount importance in improving the survival rates because the patients have a high probability of being treated effectively once the experienced professionals have detected the disease at its initial stages. The traditional diagnostic processes such as mammography, ultrasound, histopathology and thermal imaging have provided the clinicians with a long way but they have been faced with many failures such as human error, variations in the quality of the image and these processes also fail to detect subtle patterns. These weaknesses indicate the need to incorporate more reliable and independent systems that will assist the medical professionals in decision making.

In the past few years, machine learning (ML) and deep learning (DL) became a research instrument that can be successfully used in the healthcare sector, particularly medicine imaging. Convolutional Neural Networks (CNNs) have shown marvelous properties in feature extraction of medical images that are extremely difficult to conceive, and ranked with high precision. Some of them have employed CNN-based methods in detecting and diagnosing breast cancer and found that they enhanced the traditional methods significantly [1]. Transfer learning has also enhanced the precision of diagnosis even when the data is sparse whereby the already trained CNN models such as VGG16, ResNet, DenseNet and EfficientNet are modified to the medical data under consideration. This has resulted in machine learning being a good strategy that can be employed in addressing the problem of breast cancer classification [2].

On the other hand, the gaps in the given researches are also not insignificant. A majority of the methods relying on CNN are offered to be binary (benign vs. malignant) [3], which is not true because several classes that describe various types of breast abnormalities are required [5]. Other more precise studies exist, and are founded on the very large datasets or complex architectures that are not necessarily applicable to practice in the real clinical setting [13]. Moreover, despite high efficiency, CNNs are low-interpretability, which is problematic among clinicians as a doctor has to have sufficient understanding of reasoning of the choice of the model in the life or death medical setting.

In consideration of these concerns that can still occur, in this work, a methodology that would rely on CNN-based models, as well as Capsule Networks, will be necessary to propose, as it takes into consideration not only the accuracy but also the efficiency and explains the visualization that would be achieved through the aid of such tools as Grad-CAM. Moreover, the research confirms the concept of the practical application because the trained model is applied to a mobile-based platform, which enables the application of the given idea in healthcare. Addressing such deficits, the provided study will not only help in improving the performance of breast cancer detection but also in reducing the divide between the machine learning research and application in the sphere of medical providers.

2.2 Literature Review

CNNs provide a strong conceptual foundation due to their hierarchical architecture, where shallow layers capture basic image features while deeper layers extract complex and abstract representations. Transfer learning has become especially valuable, as pretrained models like ResNet, GoogleNet, and VGG can be fine-tuned on medical images, overcoming the challenge of limited annotated datasets. In addition, hybrid approaches that combine CNNs with sequential models such as Long Short-Term Memory (LSTM) or with traditional classifiers like Support Vector Machines (SVMs) further expand their potential, enabling not only binary classification of benign and malignant cases but also more complex tasks such as multi-class classification, segmentation-based diagnosis, and explainable decision-making.

Several studies illustrate CNN effectiveness in mammography. Gao [1] introduced a Shallow-Deep CNN that generated virtual recombination images from low-energy mammograms, improving diagnostic accuracy to an AUC of 0.92 without requiring specialized equipment. Zuluaga-Gomez [2] developed a CNN-based CAD system using thermal images, where preprocessing and augmentation were shown to be vital, achieving 92% accuracy. Heenaye-Mamode Khan [3] extended classification beyond benign and malignant categories, employing ResNet50 to classify multiple breast abnormalities with 88% accuracy. Desai and Shah [4] compared CNNs with multilayer perceptrons and found CNNs to perform better, while Rajakumari and Kalaivani [5] demonstrated that GoogleNet achieved 99% accuracy on mammography datasets, outperforming optimization-based models. Wang et al. [6] strengthened CAD performance through feature fusion, combining CNN-extracted features with morphological and texture descriptors.

Histopathology has also benefited greatly from CNN applications. Benhammou [7] carried out one of the first CNN-based evaluations in this domain, highlighting its potential despite small-scale testing. Later studies advanced the field, such as Gour [10], who designed ResHist, a 152-layer residual CNN specifically adapted for histopathological images, which achieved 92.52% accuracy. Albashish [11] employed transfer learning with VGG16, classifying both binary and multi-class histology samples, while Gupta and Chawla [23] demonstrated that CNN features could be effectively combined with Logistic Regression for efficient classification. These studies highlight both the power of CNN architectures and the flexibility of integrating them with simpler classifiers to reduce computational complexity.

Methodological innovations further underline CNN versatility. Sannasi Chakravarthy [9] tested pretrained models with SVM classifiers and dimensionality reduction, reaching nearly 98% accuracy. Kaddes [13] developed a CNN-LSTM hybrid architecture that achieved almost 99.9% accuracy on Kaggle datasets. Similarly, Srikantamurthy [22] proved that combining CNNs with LSTMs improves subtype classification in histopathology, obtaining 99% accuracy in binary classification and 92.5% in multi-class tasks. Segmentation-based approaches have also been impactful: Tsochatzidis [14] incorporated lesion segmentation into CNN models, and Bouzar-Benlabiod [18] integrated CNN feature extraction with case-based reasoning to provide explainable decisions. Wahab [20] and Zainudin [25] focused on mitosis detection in histopathology, a particularly challenging task, showing that CNN-based models can handle complex image features, though with varying levels of accuracy.

CNN applications in thermography further emphasize their adaptability. Goncalves [15] improved CNN performance with bio-inspired optimization algorithms, while Raza [19] combined GoogleNet and ResNet to achieve 95% accuracy on thermal datasets, outperforming traditional classifiers. Roslidar [24] tested multiple CNNs on thermal breast images, with DenseNet reaching 100% accuracy, although the dataset size was small and limited broader validation.

Reviews and comparative studies add further context. Sharma [8] reviewed machine learning methods and confirmed CNN superiority over traditional approaches, while Wahed [12] conducted a large systematic review of more than 30,000 articles, showing CNNs to be the dominant AI method in breast cancer detection research. However, they also highlighted a lack of focus on deployment challenges, an issue echoed in other comparative works such as Chaudhari [16] and Jafari and Karami [17]. The below table 2.1 shows the summary of the related studies.

Table 2.1: Summary of Literature Reviewed.

Author (Year)	Model Used	Accuracy	Contribution
Gao et al. [1]	Shallow-Deep CNN (SD-CNN)	90% (AUC 0.92)	Proposed virtual recombined images improving diagnosis without special hardware.
Zuluaga-Gomez et al. [2]	CNN (thermal imaging)	92%	Validated thermal imaging with augmentation as an efficient breast cancer screening tool.
Heenaye-Mamode Khan et al. [3]	ResNet50 (multi-class)	88%	Extended diagnosis to multi-class breast abnormalities beyond binary classification.

Rajakumari & Kalaivani [5]	GoogleNet, AlexNet	99%	Demonstrated deep CNNs (GoogleNet) outperform optimization-based MLPs.
Wang et al. [6]	CNN features + Extreme Learning Machine (ELM)	95%	Showed feature fusion improves diagnostic accuracy and efficiency.
Sannasi Chakravarthy et al. [9]	AlexNet, GoogleNet, ResNet50, DenseNet121 + Feature Fusion	97.9%	Highlighted benefit of combining deep features with SVM for improved results.
Gour et al. [10]	ResHist (152-layer CNN)	92.5%	Proposed residual learning model specifically for histopathology.
Albashish et al. [11]	VGG16 + ML classifiers	~94%	Applied transfer learning for binary and multi-class histopathological classification.
Kaddes et al. [13]	CNN + LSTM hybrid	99.9%	Introduced hybrid CNN-LSTM achieving very high accuracy.
Tsochatzidis et al. [14]	CNN with segmentation + altered loss function	AUC 0.898	Proved integrating lesion segmentation improves mammogram classification.
Gonçalves et al. [15]	VGG16, ResNet50 + bio-inspired optimization	F1-score ~0.92	Showed optimization improves CNN performance on thermal imaging.
Jafari & Karami [17]	CNN feature fusion + feature selection + ML classifiers	96%	Demonstrated feature fusion with selection boosts accuracy across datasets.
Bouzar-Benlabiod et al. [18]	CNN + Case-Based Reasoning (CBR) + U-Net	91.3% (recall)	Combined deep learning with explainable CBR for mammogram classification.
Srikantamurthy et al. [22]	CNN-LSTM (hybrid)	99% (binary), 92.5% (multi-class)	Combined spatial + sequential learning for subtype classification.

Roslidar et al. [24]	DenseNet, ResNet101, MobileNetV2, ShuffleNetV2	100% (DenseNet)	Validated transfer learning on thermal images with exceptional accuracy.

2.2.1 Similar Applications

Recent studies using machine learning methods to enhance the accuracy of breast cancer detection and diagnosis have demonstrated trends that are near-identical to the current one. The CNN-based structures attracted the interest of many researchers to discriminate abnormalities in the breast. As an example, Gao et al. [1] proposed a Shallow-Deep CNN model, which incorporated the shallow and deep layers to enhance the diagnosis by using a mammography approach, whereas Zuluaga-Gomez et al. [2] indicated that when paired with CNNs, the thermal imaging method could become a competitive model with high accuracy that could be used as an affordable screening technique. The papers mentioned stress the versatility of CNNs in various modes of imaging, and this aspect applies to the work of this study where a greater performance has been sought through the use of variations of CNNs and CapsuleNet.

The findings of other papers focused on how transfer learning and deeper models can enhance diagnostic performance. GoogleNet demonstrated 99% accuracy on mammographic data as proved by Rajakumari and Kalaivani [5] and Gour et al. [10] have designed ResHist, deep residual model, aiming specifically at the histopathological classification. Likewise, Albashish et al. [11] used VGG16 to perform transfer learning on histopathology with very good results on the binary, as well as multi-label task. The results relate to the existing approach quite well, as the current research also compares CapsuleNet to sophisticated CNN models like EfficientNet, ResNet, and Inception.

Fusion methods of hybrid and features are broadly tested as well. Wang et al. [6] used deep features of CNN with an Extreme Learning Machine classifier to improve the accuracy and Sannasi Chakravarthy et al. [9] demonstrated that after integrating multiple CNNs features with SVM, it was possible to exceed 98% but inaccuracy. Identically, to boost the generalizability, Jafari and Karami [17] implemented feature selection and/or fusion among CNN models. The hybrid schemes appeal to the comparative configuration of this study whereby the performance of CapsuleNet is compared to other CNN models to achieve the most dependable result.

The growing trend is the hybrid sequential-spatial networks. As an example, Kaddes et al. [13] and Srikantamurthy et al. [22] combined CNNs and LSTM in order to extract both spatial and sequential features, achieving extremely high accuracies of up to 99.9%. Whereas this thesis focuses more on CapsuleNet and CNN families compared to sequence modeling, those commonalities are to capture feature representations that are richer to make better decisions.

The emphasis on explainability and clinical usability has also been brought up by increasing number of studies. Bouzar-Benlabiod et al. [18] proposed incorporating case-based reasoning into CNN to allow explainable outputs, whereas Tsochatzidis et al. [14] added segmentation into CNN training in order to make the decisions of the model more transparent. This is consistent, directly, with the primary approach utilized, namely Grad-CAM in explainable AI (XAI) helping visualize the most important areas of an image and contributing to clinical trust.Last but not least, there are studies whose conclusions are deployment and application readiness. Roslidar et al. [24] have shown that to achieve good performance on CNNs requires fine-tuning the model to suit thermal images, and Goncalves et al. [15] optimized CNNs with the use of bio-inspired algorithms with the focus on efficiency and adaptation. These efforts indicate the intended objective of the proposed work, which would be to prepare the trained model to be deployed in a mobile application via TFLite which would be applicable in the real world and not only in the context of a controlled testing. Overall, the mentioned applications indicate the following three directions that will be combined in the thesis: (1) the use of sophisticated CNN models and CapsuleNet that will provide robust diagnoses, (2) the integration of explainable AI allowing improving clinical trust, and (3) the focus on making deployments to mobile devices feasible. This compliance with the existing research makes the methodology not only actual concerning scientific developments but also feasible when used in a healthcare environment. The below table 2.2 shows the summary of the similar studies.

Table 2.2: Similar Applications table

Author	Model Used	Accuracy	Contribution
Gao et al. [1]	Shallow-Deep CNN	90% (AUC 0.92)	Virtual recombined images improved diagnosis.

Zuluaga-Gomez et al. [2]	CNN (thermal imaging)	92%	Proved thermal imaging effective with augmentation.
Rajakumari & Kalaivani [5]	GoogleNet, AlexNet	99%	GoogleNet outperformed optimization-based MLPs.
Gour et al. [10]	ResHist (152-layer CNN)	92.5%	Residual learning adapted for histopathology.
Albashish et al. [11]	VGG16 + ML classifiers	~94%	Transfer learning applied to binary and multi-class.
Wang et al. [6]	CNN + ELM (feature fusion)	95%	Showed feature fusion boosts performance.
Sannasi Chakravarthy et al. [9]	CNNs + Feature Fusion + SVM	97.9%	Combined deep features improved accuracy.
Jafari & Karami [17]	CNN feature fusion + ML	96%	Feature fusion + selection enhanced generalization.
Kaddes et al. [13]	CNN + LSTM	99.9%	Hybrid CNN-LSTM reached very high accuracy.
Srikantamurthy et al. [22]	CNN + LSTM	99% (binary), 92.5% (multi)	Combined spatial + sequential learning.
Bouzar-Benlabiod et al. [18]	CNN + CBR + U-Net	91.3% (recall)	Added explainability via case-based reasoning.
Tsochatzidis et al. [14]	CNN + Segmentation	AUC 0.898	Improved mammogram diagnosis with lesion focus.
Roslidar et al. [24]	DenseNet, ResNet101, MobileNetV2	100% (DenseNet)	Fine-tuning CNNs proved highly effective.
Gonçalves et al. [15]	VGG16, ResNet50 + Optimization	F1 = 0.92	Optimization enhanced CNN on thermal images.

2.3 Gap Analysis

Although machine learning (ML) modalities have shown significant promise in breast cancer detection, several persistent limitations hinder their clinical adoption. The first and most critical gap lies in dataset size and diversity. Many studies report high performance but rely on small, single-source datasets that lack external validation, thereby restricting clinical generalizability. For example, Gao et al. successfully applied a shallow–deep CNN, but their limited sample size undermined the robustness of the findings [1]. Similar limitations are observed in works employing thermal imaging [2, 19] and histopathological datasets [10, 11, 22], where accuracies are reported as high but are constrained by narrow, domain-specific sources. This highlights the need for larger, balanced datasets with external validation to ensure reproducibility and wider applicability in practice.

A second limitation concerns the scope of classification. Much of the literature narrows its focus to binary detection (cancerous vs. non-cancerous) or small-scale multi-class problems. While these approaches may simplify model design, they fall short of reflecting clinical reality, where broader subtype differentiation is required. For instance, Heenaye-Mamode Khan et al. explored multi-class detection but achieved limited results, with little benchmarking against more advanced architectures [3]. Other works [5, 22] similarly demonstrate restricted classification capabilities, underscoring the necessity for models that can address the complexity of real-world diagnostic settings.

A third gap lies in computational feasibility. Hybrid networks such as CNN–LSTM [13, 22] or highly parameterized CNNs [6] have achieved competitive accuracies but at the expense of computational overhead, rendering them unsuitable for deployment in resource-constrained clinical environments. This reinforces the importance of architectures like CapsuleNet, which can capture spatial hierarchies more efficiently, balancing accuracy with parameter efficiency and making them more practical for wider use.

The fourth gap is the lack of explainability and interpretability. Although performance metrics dominate the literature, most studies provide little insight into how predictions are made. Feature-fusion and optimization methods [6, 9, 15, 17] enhance predictive power but fail to improve transparency for clinical decision-making. Attempts at interpretability, such as segmentation-based or case-based reasoning approaches [14, 18], often lack scalability or remain highly dataset dependent. This discontinuity emphasizes the need to embed explainable AI (e.g., Grad-CAM) into diagnostic pipelines to improve clinician trust and accountability.

Finally, there remains a translational gap between research and practical deployment. While high-performing models such as DenseNet [24] or CNN–LSTM variants [13, 22] demonstrate excellent experimental accuracies, few efforts have considered lightweight adaptation for real-world settings. Most models are not optimized for mobile or edge deployment, leaving a critical divide between algorithmic performance and clinical usability [15, 11]. Consequently, despite promising experimental outcomes, the absence of deployment-ready frameworks limits impact at the point of care.

Against this background, the present study positions itself to address these gaps by (i) employing a larger, balanced dataset with explicit train–validation–test partitions, (ii) extending evaluation beyond binary classification to broader multi-class settings, (iii) leveraging CapsuleNet for efficient spatial feature representation with reduced computational burden, (iv) integrating Grad-CAM for transparent, interpretable predictions, and (v) deploying the optimized model on a mobile platform using TFLite to enhance accessibility in diverse healthcare environments. The below table 2.3 shows the Gap Analysis Summary.

Table 2.3 Gap Analysis Summary Table

Gap Identified	**Observed In**	**This Study's Contribution**
Small dataset size and lack of external validation	Gao et al. [1], Zuluaga-Gomez et al. [2], Gour et al. [10], Albashish et al. [11]	Uses a larger and balanced dataset split into training, validation, and testing to improve generalizability.
Limited focus on binary classification or small multi-class tasks	Heenaye-Mamode Khan et al. [3], Rajakumari & Kalaivani [5], Srikantamurthy et al. [22]	Compares CapsuleNet and multiple CNN models across broader classification tasks for more reliable diagnosis.
High computational complexity in hybrid or deep models	Kaddes et al. [13], Srikantamurthy et al. [22], Wang et al. [6]	Applies CapsuleNet, which captures spatial hierarchies efficiently with fewer parameters, balancing accuracy and efficiency.
Lack of explainability in model predictions	Wang et al. [6], Chakravarthy et al. [9], Gonçalves et al.	Integrates Grad-CAM to highlight image regions influencing decisions,

	[15], Jafari & Karami [17]	improving clinical trust.
Limited attention to real-world deployment	Roslidar et al. [24], Gonçalves et al. [15], Albashish et al. [11]	Deploys the trained model into a mobile application (TFLite), ensuring usability in practical healthcare settings.

2.4 Summary

Breast cancer remains a major problem in health affecting most of the nations across the globe, requiring more effective modalities of early and precise diagnosis. Though the conventional approach of mammography, ultrasound, and histopathology has existed since long among the practitioners, it has its own problems which can be in shape of human dependency, interpretation variability, and diminished relevance in identifying complex cases. Such shortfalls underline the significance of approaching a higher level of computational practices in aid of healthcare workers to improve clinical judgment. Machine learning have been presented as powerful solutions in this regard especially with the success of Convolutional neural networks in the implementation of diagnosis using images. According to the previous literature, the approach including CNN can attain high accuracy and substantially decrease diagnosing errors. Transfer learning using popular architectures like VGG, ResNet, DenseNet and EfficientNet has only enhanced performance further more so in places and cases where access to data is scarce. Nevertheless, some critical gaps remain, as outlined by current research: the concentration on the binary classification problem, the lack of interpretability of the models, and the inability to apply the results of research in the sphere of healthcare. The examined study helps fill these gaps by implementing Capsule Networks and mixing them with CNN-based models in order to enhance the classification results, but they also add an element of explainability with visualisations based on Grad-CAM. Notably, the study does not stop at an experiment in that the end result is applied to a mobile-based program through which device usability in the real world is offered. In this manner, the proposed study is expected to give more than increased diagnostic accuracy and the credible, transparent, meaningful solutions to breast cancer detection.

Chapter 3

Research Methodology

The chapter describes the approach that is aimed at reaching the research objectives. It outlines the system structure, data description, preprocessing strategies, extraction of features based on DenseNet and classification based on machine learning models and interpretation through Grad-CAM. The process of methodology is outlined in a systematic way to show how the research was performed and to make it reproducible.

3.1 Methodology/Requirement Analysis & Design Specification

3.1.1 Overview

The proposed paper uses machine learning to label images from breast cancer histopathology as either benign or malignant to help make personalized choices of treatment. The 6,000 images were carefully preprocessed (with DPI normalization) and divided into training (4,335), validation (765), and testing (900) sets in order to provide model-robust evaluation. CapsuleNet (the main model) was trained alongside well-known models, like EfficientNetB0, InceptionV3, ResNet152V2, ResNet50, and MobileNetV2, to compare the performance. Grad-CAM (Gradient-weighted Class Activation Mapping) was also used as an explainability methodology to give visual explanations of how the model makes predictions. The last optimized model was implemented as a mobile application through TensorFlow Lite (TFLite), making it suitable for clinical applications. This approach offers both technical excellence and clinical experience to generate better diagnostics and treatment planning in cancer.

3.1.2 Proposed Methodology/ System Design

Figure 3.1 illustrates the methodological process of the provided study, and it begins with data collecting, i.e., in this case we have a classification of histopathology images into two categories (benign and malignant). The data is preprocessed before processing like normalizing the DPI and an appropriately organized split into train, valid, and test set (4,335 train images, 765 valid images, 900 test images are used). The design and training procedure of the model is based on the model CapsuleNet serving as the primary architecture and five machine learning-based models (EfficientNetB0, InceptionV3, ResNet152V2, ResNet50, and MobileNetV2) on which the model comparative performance evaluation process is founded. In order to achieve interpretability, Explainable AI (XAI) techniques are employed, which is Grad-CAM, a method that provides a graphical view of the model decisions. Finally, the most suitable model can be converted to TFLite format to be deployed in a mobile app therefore clinical applications are easily achievable. A balance can be achieved through such a pipeline between transparency and computation rigor and breast cancer diagnosis.

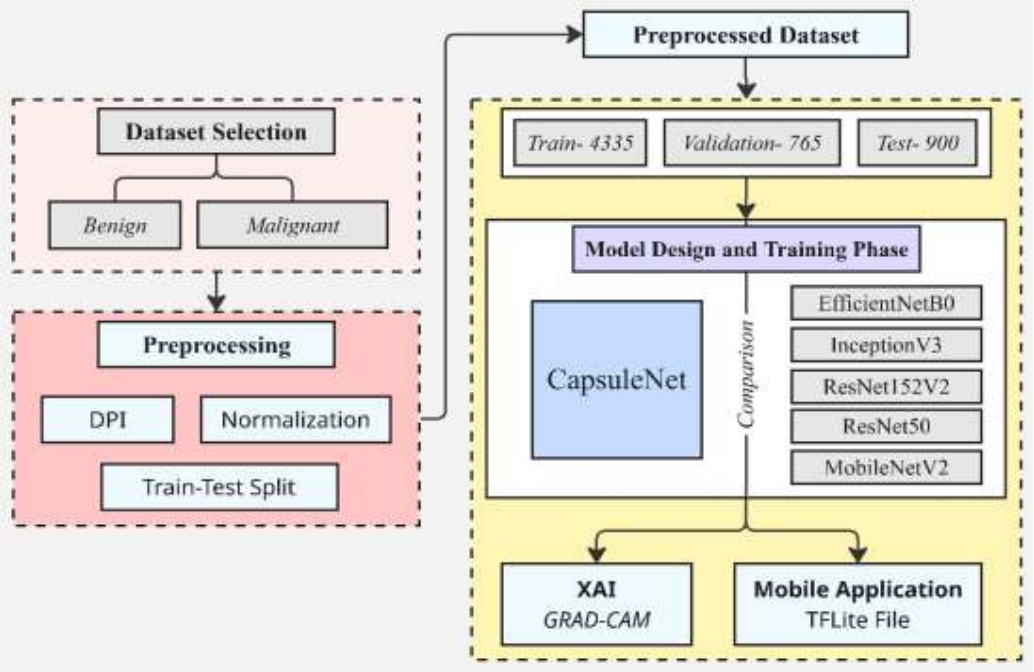


Figure 3.1: The Methodological Flowchart

3.1.3 Functional and Nonfunctional Requirements

To ensure that the breast cancer classification system is robust and clinical, the following functional and nonfunctional requirements are identified:

Functional Requirements

- Cancer Classification: The system must be capable of classifying the histopathology images under two categories known as Benign or

Malignant and a high Confidence score.

- Preprocessing: Noise reduction and balancing of the dataset should be automatically processed to achieve DPI normalization to standardize input images to ensure the consistency of the further analysis.
- Model Training: The model must be trained and tested on the CapsuleNet model, and compared in the performance of the model against baseline models (EfficientNetB0, ResNet152V2, InceptionV3, ResNet50 and MobileNetV2).
- Performance Evaluation: Expectantly, the system will generate detailed performance evaluations like accuracy, precision, recall, F1-score, and ROC-AUC over visualization of visual aids including confusion matrices and Grad-CAM heatmaps.
- Mobile Integration: This last model must be capable of exporting as a tflite file to the mobile health app and be deployed in real-time diagnostic support.
- User Interface: Depending on whether it is implemented on mobile or web, the interface user must have the capability to post histology pictures to it, and one should be able to retrieve live foresight.

Nonfunctional Requirements

- Scalability: The system must be able to scale, i.e. have datasets increase without severely negatively affecting performance or latency.
- Latency: Sub-second inference time on mid-range mobile devices is required to be practical during clinical use.
- Privacy & Security: The information concerning patients has to be anonymized when processing is involved, and when implemented in a medical facility, the data should follow HIPAA/GDPR.
- Model Robustness: The system must be able to maintain classification accuracy at different levels of image quality and the different staining artefacts typical of histopathology.

3.1.4 Data Flow Diagram

Figure 3.2 shows a high-level data flow of the breast cancer classification system, which starts with the Breast Cancer Dataset. The dataset undergoes preprocessing steps, including normalization and separation into training, validation, and test sets. The resultant data is used as training input in the

CapsuleNet, and its performance is compared to that of alternative CNN models (EfficientNetB0, InceptionV3, ResNet152V2, ResNet50, and MobileNetV2) to establish its efficacy. The performance of the model is measured by some important performance numbers and explained with the help of visualizations such as a confusion matrix or Grad-CAM. Eventually, the optimised model can be exported as a TFLite file that can be easily integrated into a mobile application, allowing for real-time cancer diagnosis and making it accessible to all. This end-to-end process provides a streamlined approach to data preparation, from clinical deployment.

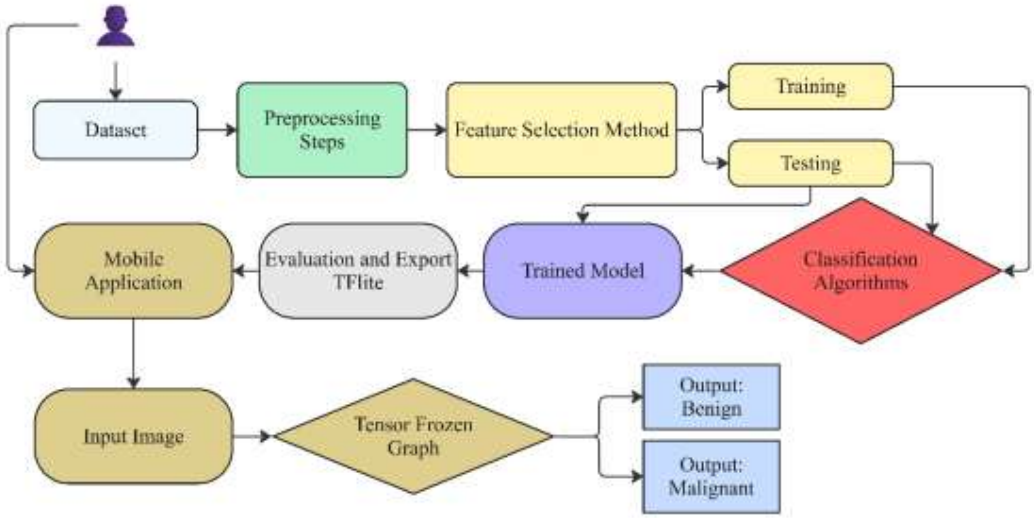
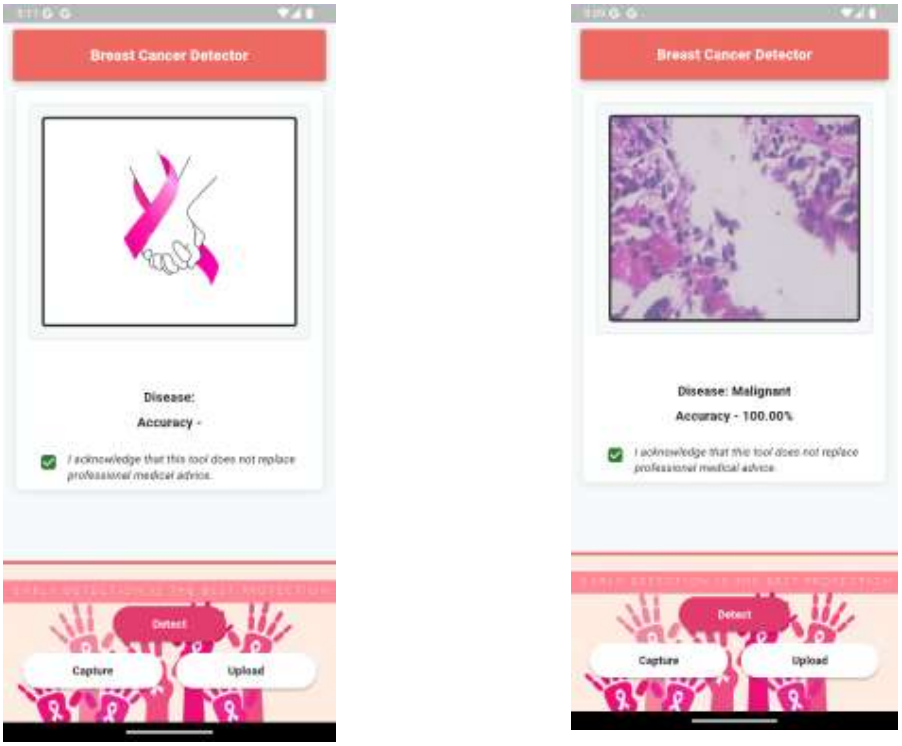


Figure 3.2: The proposed Data Flow Diagram

3.1.5 UI Design

Figure 3.3 shows the user interface (UI) of the Breast Cancer Detector smartphone app using a simple and easy-to-use layout design. The prediction of Benign or Malignant is represented on the screen. The interfaces that users can interact with in the system have two main categories that include taking a new image of histopathology using the camera of the device or uploading an already stored image. Its simplified design focuses more on usability, allowing clinicians and healthcare personnel to work out speedy AI-driven diagnoses with the least amount of effort, thereby improving decision-making in clinical practice.



Disease:
Accuracy -
☑ *I acknowledge that this tool does not replace professional medical advice.*

Disease: Malignant
Accuracy - 100.00%
☑ *I acknowledge that this tool does not replace professional medical advice.*

Figure 3.3: The proposed UI design for the Mobile Application

3.2 Detailed Methodology and Design

3.2.1 Dataset

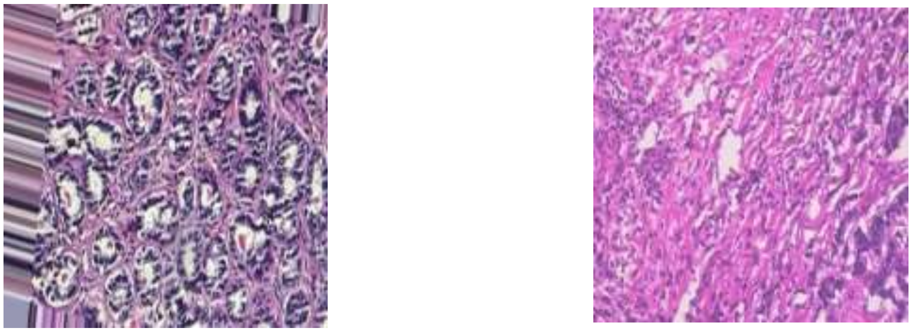
The paper is based on a publicly accessible database of breast cancer histopathology [28] and is offered on Kaggle and has 6,000 images performed on high-resolution benign or malignant. This is information that will be chiefly involved in training, validating and testing the machine learning models. This will imply that the research has identified accuracy and repeatability. The impact of the use of such a dataset is rather drastic; the elimination of the biases in both small and artificial datasets, and the improved externalizability of the results in general is standardized.

Table 3.1: Dataset Specifications

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

3.2.2 Classes

- Benign: The benign group contains the images of benign histopathology of breasts, the regular structure of the cells, nuclei, and the absence of any similarities, as well as their lack of any invasive growth pattern. Even though the presence of such samples is not an immediate danger, their presence in the dataset enables the model to distinguish between the normal and abnormal tissue and restrict false positives in case of clinical diagnosis.
- Malignant: This cluster of images contains images that depict cancerous tissues of the breast since they depict abnormal forms of cells, hyperchromatic nuclei, and disorganized developments of the cells, which reflect invasive tumors. The major precondition is the accurate recognition of the malignant cases and thus, this type can be instrumental in conditioning the model to identify the pathological traits of high risk with high sensitivity.



Benign

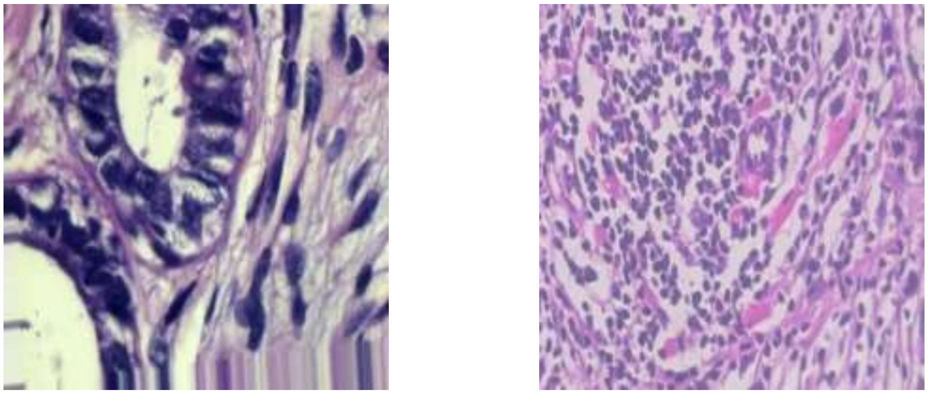
Malignant

Figure 3.4: Sample Image from Each Class

3.2.3 Preprocessing Steps

Considering the best functioning of models based on the data, three essential data preprocessing steps were performed on the dataset:

- DPI Normalization: The histopathology images were all resampled to a constant dots-per-inch (DPI) resolution to ensure that features at different scales exist with some consistencies within the dataset, and also to remove differences due to the different scanning resolutions.



Benign – DPI: 240x240

Malignant – DPI: 240x240

Figure 3.5: Sample Images after changing DPI values

- Normalization: The pixel intensities were formatted to a standard range (e.g. [0, 1] or z-score) where illumination biases can be minimized, and model training can be faster. The step promotes stability and reproducibility of the machine learning workflows.

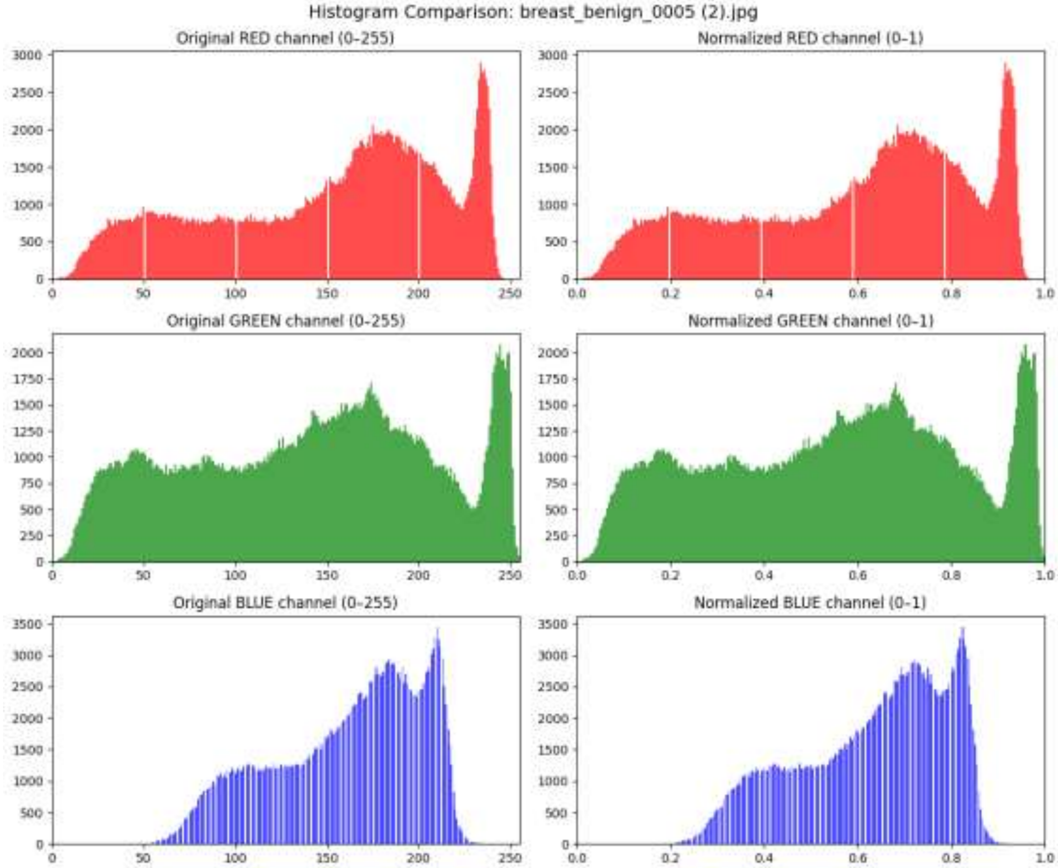


Figure 3.6: RGB Channel Histograms Before and After Normalization

- Train-Validation-Test Split: The dataset was split into three subsets, train (4,335 images), validation (765 images), and test (900 images) in a stratified manner to maintain the class distribution (see Figure 3.5). This division guarantees a lack of bias in the evaluation, but a sufficient amount of data will allow generalizing the model.

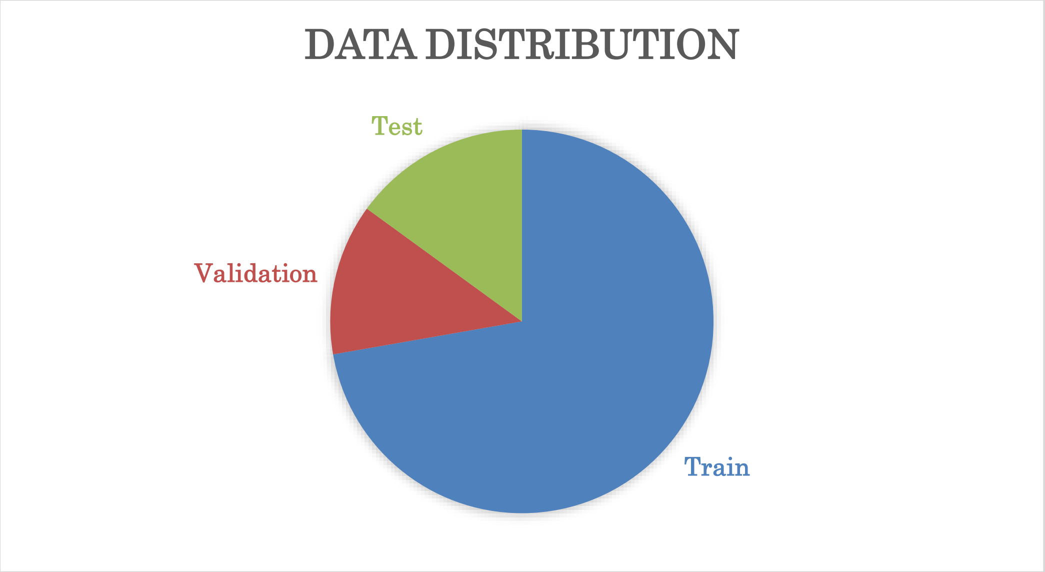


Figure 3.7: Data Distribution for Train, Validation and Test

3.2.4 Model Description

1. Working Procedure of CapsuleNet

The main machine learning model adopted to classify breast cancer was the CapsuleNet (Capsule Network) architecture, which provides certain advantages to conventional CNNs due to the presence of dynamic routing and the ability to maintain hierarchies in a spatial context. Its working process is explained below:

- Input Layer: Input tensors contain preprocessed histopathology images, standardization, and normalization of DPI.
- Primary Capsules: The initial layers of the convolution neural net learn low-level features (edges, textures). These features are then grouped into capsules, where a capsule contains neurons which encode the presence of their feature as well as the instantiation of a parameter (pose, orientation) of their feature.
- Dynamic Routing: Lower-layer capsules make predictions on higher levels capsules using a transformation matrix. Rout-ing by agreement

mechanism (Eq. 1) makes only consistent predictions capable of strengthening the correlation between capsules, which lowers the need to resort to max-pooling and maintains a relationship between locations.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad \text{———————————} \quad (1)$$

where c_{ij} is the coupling coefficient between capsule i and j, and b_{ij} is the log prior probability.

- Capsule Activation: The length of a given capsule's output vector is its probability of a given feature. A squashing (nonlinear) logistic function (Eq. 2) normalizes the vector lengths (if they are greater than 0 and less than 1).

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|} \quad \text{———————————} \quad (2)$$

where s_j is the weighted sum of input capsules, and v_j is the output vector.

- Classification Layer: The Classification layer is made up of two capsules (one benign and one malignant). The most significant length of vectors is decided by the capsule whose length is predicted.
- Loss Function: A margin loss increases the discriminative power of medical images by penalizing the difference between the predicted and actual capsule activations.

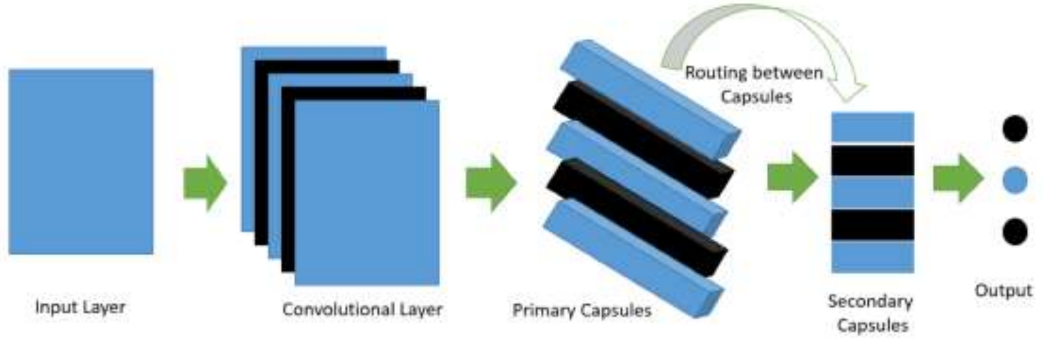


Figure 3.8: The base architecture of the CapsuleNet Model [26]

3.2.5 Explainable AI: GRAD-CAM

Grad-CAM (Gradient-weighted Class Activation Mapping) is another explainability method, which shows the regions of a specific image that the model is influenced the most (see Figure 3.7). It operates based on the gradients of the desired category (e.g. "malignant") into the final convolutional layer of a CNN. Those are mean gradients used to compute weights of neuron importance which are again combined with activation maps of a layer to obtain a rough localization heatmap. Unlike the less complex procedures of such as CAM, grad-CAM does not require any modifications in the architecture and can thus be implemented in a vast number of environments using existing models. The outcome is the heatmap superimposed on the original image, in which specifications of attributes, e.g., tumor boundaries, can be visualized without the loss of the spatial information. Grad-CAM helps eliminate the black-box nature of predictions made by a model and the lack of insight into the reasoning used by humans by demonstrating, on top of the prediction made by the model, the rationale. This may bring motivation to clinical and high-stakes environments. Its clarity and intelligence have placed it at the heart of transparent AI on computer vision.

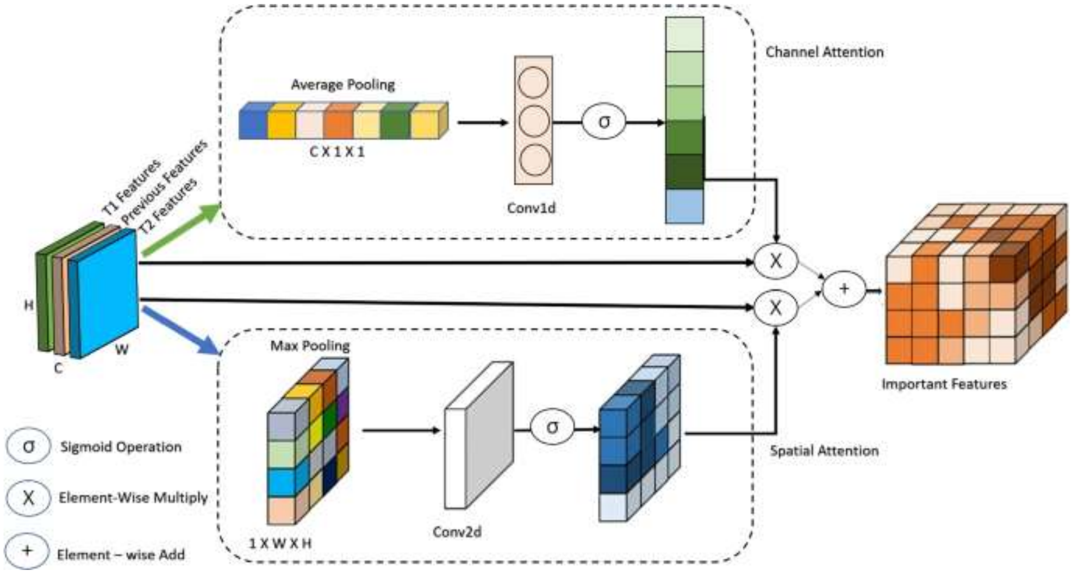


Figure 3.9: The base architecture of the GRAD-CAM model [27]

3.3 Project Plan

The research was carried out in a procedural and systematic manner with a fairly elaborate schedule between the months of January and August 2025 of six recursive stages that included: (1) Theoretical Study and Literature Review to get a requisite knowledge base and the disparities; (2) Data Collection of histopathology images to curated repositories; (3) Data Preprocessing, which involves normalization of DPI and the division of the data; (4) Model Design including the design of CapsuleNet, the testing and documenting the methods used; (5) App Development and The project combined the exploratory research (XAI implementation) and deliverability deadlines (app deployment) to combine technical and scholastic objectives.

Table 3.2: The GANTT Chart of Project Timeline

Process	Jan’25	Feb’25	Mar’25	Apr’25	May’25	June’25	July’25	Aug’25
Working Plan	🟦	🟦						
Theoretical Study	🟥	🟥						
Literature Review		🟦	🟦					
Data Collection			🟦	🟦				
Data Preprocessing				🟦				
Model Design + Methodology Writing					🟦	🟦		
App Development and Testing						🟦	🟦	
Report Writing							🟫	🟫

3.4 Task Allocation

During the research, the organizational tasks were logically aligned among the team members according to their spheres of knowledge and complex tasks such as general coordination and methodology development were offered by the principal investigator. Biomedical imaging researchers were assigned the duty of gathering and pre-processing data hence choosing individuals who are familiar with histopathology data. These developments were planned by machine learning experts who were in charge of model design, its implementation, and comparative analysis, explainability methods (Grad-CAM), which were told by mobile deployment experts who were required to organize the event, and interpretability in AI and edge computing specialists in the team. The writing of reports and technical documents was done simultaneously and the reviews were carried out in iterative mode so that they could be accurate academically. This critical allocation of all talents exploited individual talents, and also, cross-disciplinary collaboration which ensured both academic and technical skills of the project being pursued. Regular sync meetings and version-controlled workflows made everything run in a single direction in the entire development period.

3.5 Summary

In the chapter, the design, development, and integration of an AI-driven system of breast cancer classification have been defined in a scholarly way, comprising of both the computational accuracy and the clinical applicability. In order to obtain study strength, 6,000 histopathology images were used which were pre-processed and which were split to a stratified set in order to avoid leakage of data. The architecture primarily used was CapsuleNet that had a dynamic routing system that preserved spatial hierarchy better than the current CNNs (EfficientNet, ResNet variants). The interpretability of the grad-CAM generated models had the visual aspect of algorithmic decision-making as well as their visual explanation, which could be comprehended by a human. The outcome of the pipeline was mobile deployment (TFLite), which passed the test of performance metrics and ethical issues. The phases were strictly organized and arranged, such as theoretical basis and app development, with the allocation of tasks, which facilitated adhering to transparency and reproducibility of the methods applied.

Chapter 4

Implementation and Results

This chapter introduces a critical analysis of a machine learning system to classify medical images based on the CapsuleNet system. Various interpretability and uncertainty quantification methods are presented in the analysis to determine the reliability and transparency of decision-making of the model. The important tools like Grad-CAM are used to display the focus of the models and feature influence, and uncertainty maps are used to demonstrate how confident the model is in its predictions. The high performance of the model is supported by quantitative measures with classification measures, confusion matrices, and ROC-AUC scores. In both qualitative and quantitative findings, this chapter provides an emphasis to the clinical relevance, reliability and interpretability of CapsuleNet when dealing with sensitive medical imaging problems.

4.1 Environment Setup

The general training parameters employed by all the experimented models are presented in Table 4.1, which guarantees consistency and equal comparison of architectures. The images were downsized to 64 x 64 pixels, which is a common size of inputs of most CNN-based models, to maintain an appropriate level of spatial resolution and balance it with computational efficiency. The relatively large batch size (8) has been selected, allowing the gradient updates to be stable, and the use of the GPU memory to be efficient enough in training. The models were trained with 100 epochs and this gave them sufficient number of iterations that would ensure convergence but would reduce the chances of overfitting. Adam optimizer was used because it has an adaptive learning and fast convergence properties which are suitable in machine learning activities. The learning rate was fixed to be 0.0001 which is a low amount that will enable the model to learn at a slow pace, avoiding radical changes in the weight updates, and enabling more fined converging with time.

Table 4.1 Common parameter table for all experimented models.

Parameter Name	Parameter Value
Image Size	64 × 64
Batch Size	8
Epoch	100
Optimizer	Adam
Learning Rate	0.001

Table 4.2 presents the data splitting strategy adopted for model training, validation, and testing. The dataset was divided into 74.5% for training (4335 images), 10.5% for validation (765 images), and 15% for testing (900 images). This split ensures that the model is trained on a substantial portion of the data while allocating a significant number of samples for unbiased evaluation. The validation set helps in fine-tuning model parameters and preventing overfitting by monitoring performance during training. The test set, which remains unseen during training and validation, serves as a final benchmark for assessing the model's generalization capability. The relatively high number of test samples (15%) strengthens the reliability of performance metrics, ensuring that the results reflect true model behavior on real-world, unseen data.

Table 4.2: Common data split for all experimented models.

Dataset	In Percentage	Number of Images
Train set	74.5%	4335
Validation Set	10.5%	765
Test Set	15%	900

4.2 Testing and Evaluation/Performance/ Comparative Analysis

In evaluating 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:

4.2.1 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.

$$Accuracy = \frac{TP+TN}{FP+FN+TP+TN} \tag{3}$$

4.2.2 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.

$$Recall = \frac{TP}{(FN+TP)} \tag{4}$$

4.2.3 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.

$$Precision = \frac{TP}{(FP+TP)} \tag{5}$$

4.2.4 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.

$$F1\ Score = 2 \times \frac{Precision+Recall}{(Precision \times Recall)} \tag{6}$$

4.2.5 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.

4.2.6 Receiver Operating Characteristic (ROC) Curve

A plot of the true positive rate (recall) versus the false positive rate created graphically as the experiment is performed over the range of threshold values. It is a diagram that shows the diagnostics capability of a binary classifier system as the discrimination threshold is swept. It is a graph where the True Positive Rate (TPR) or recall or sensitivity is plotted against the False Positive Rate (FPR) over the possible levels of threshold.

Area Under Curve (AUC): The area under ROC curve has one measure of model performance; the higher the value, the better the model.

The Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) will be the most essential in assessing classification models in this study. They enable visual comparison of model at different thresholds, which enables the best choice of a threshold based on a balance between sensitivity and specificity. The ROC curve is also resistant to class imbalances and hence offers valid measurements where misleading accuracy can occur. Nevertheless, ROC and AUC are to be used together with other indicators such as precision and recall to encompass the effectiveness of models.

4.3 Results and Discussion

The paper is an evaluation of a CapsuleNet model on binary classification of histopathological images (benign versus malignant). CapsuleNet was selected because of its capability to maintain spatial hierarchies and learn complex patterns in medical imaging. CapsuleNet is a solid alternative to conventional deep CNN classifiers because the framework of its extensive evaluation demonstrates the predictive ability, generalization capability, and reliability of the model in clinical practice.

4.3.1 Performance of the proposed CapsuleNet model

The loss and accuracy curves of training and validation with the proposed model can be viewed in figure 4.1, which comprises 100 epochs of the model called CapsuleNet. The training loss has a steady decreasing trend, which means that the loss function is well minimized as the model learns. The validation loss is also of a similar trend, that the model can be generalized, and it does not overfit considerably. At the same time, the accuracy of the training and the validation are growing consistently with a small difference in the curves of accuracy. This parallel development is a confirmation that the model is stable and learns well using the data and does not memorize it. The intersection point seems at epoch 40, beyond which, training no longer yields additional improvement in performance. This value is very persuasive of the effectiveness of the model to acquire the appropriate characteristics to classify, which proves the excellent performance of CapsuleNet on the dataset.

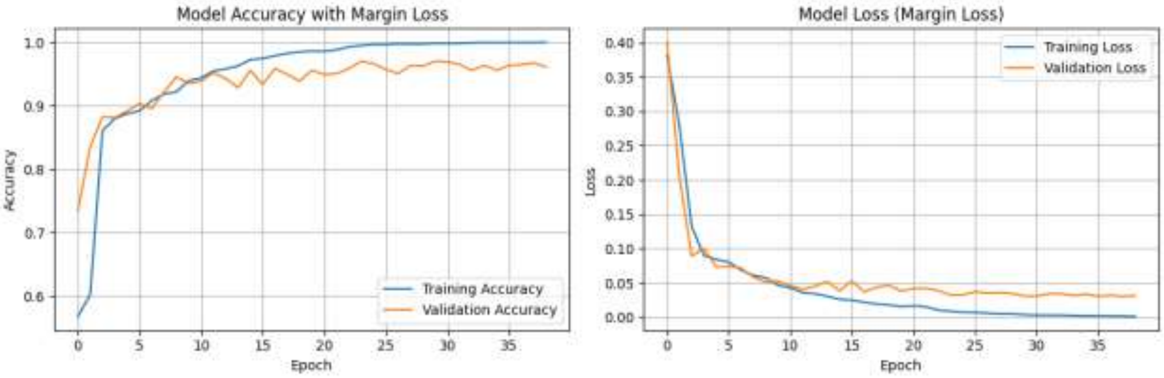


Figure 4.1: The loss and accuracy curve on training and validation set over 100 epochs for proposed CapsuleNet model.

The confusion matrix of the CapsuleNet model tested on the test set is shown in Figure 4.2. The accuracy of prediction according to the matrix is extremely high and most of the samples are correctly predicted as either Benign or Malignant. In particular, few instances of misclassifications are witnessed, which is indicative of high discrimination of classes. The matrix has true positives, true negatives and hence the model has balanced performance between the two classes. The small values off-diagonal prove the ability of the model to appropriately find out malign and benign samples, which is essential in medicine. The confusion matrix also confirms that CapsuleNet model is reliable in the classification reliability in real-world diagnostic tasks which minimizes false positives and false negatives.

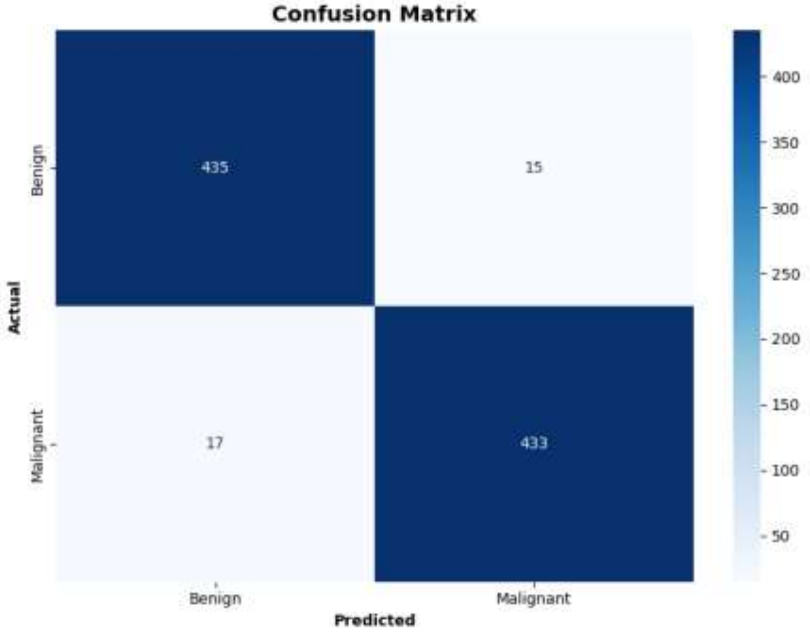


Figure 4.2: The confusion matrix on test set for proposed CapsuleNet model.

Table 4.1 is a report giving a detailed classification of the test set performance of CapsuleNet. As with both Benign and Malignant classes, the model provides a 0.96 or 0.97 precision, recall, and F1-score with the support of each category being 450. Both the macro and weighted averages are equal to 0.96 and this indicates the consistent performance of the classes. The fact that the recall is almost perfect (0.97 in Benign and 0.96 in Malignant) shows that the model will be able to detect the majority of positive cases in each category. A balance between specificity and sensitivity is also confirmed by the fact that the uniformity of precision and recall results in a high F1-score. This table highlights that the model has high classification capability in the various types of tumors and has confirmed its relevancy in the medical field of diagnostics.

Table 4.1 Classification report on test set for proposed CapsuleNet model.

Classes	**Precision**	**Recall**	**F1-score**	**Support**
Benign	0.96	0.97	0.96	450
Malignant	0.97	0.96	0.96	450
accuracy			0.96	900
macro avg	0.96	0.96	0.96	900
weighted avg	0.96	0.96	0.96	900

The ROC curve and the scores of the AUC of the CapsuleNet model are shown in figure 4.3. ROC curve shows the tradeoff between true positive rate (sensitivity) and false positive rate at various levels of classification. The curve follows the top-left corner closely showing good performance of the model. The Area Under the Curve (AUC) is almost equal to 1.0, which is an indicator of the probability of perfect separability between the benign and the malignant classes. This is more important in medical imaging work as sensitivity (recall) and specificity is crucial. The analysis carried out on ROC and AUC also supports the fact that CapsuleNet is not only accurate but highly discriminative in its predictions, which is also reliable in its performance even with changing thresholds.

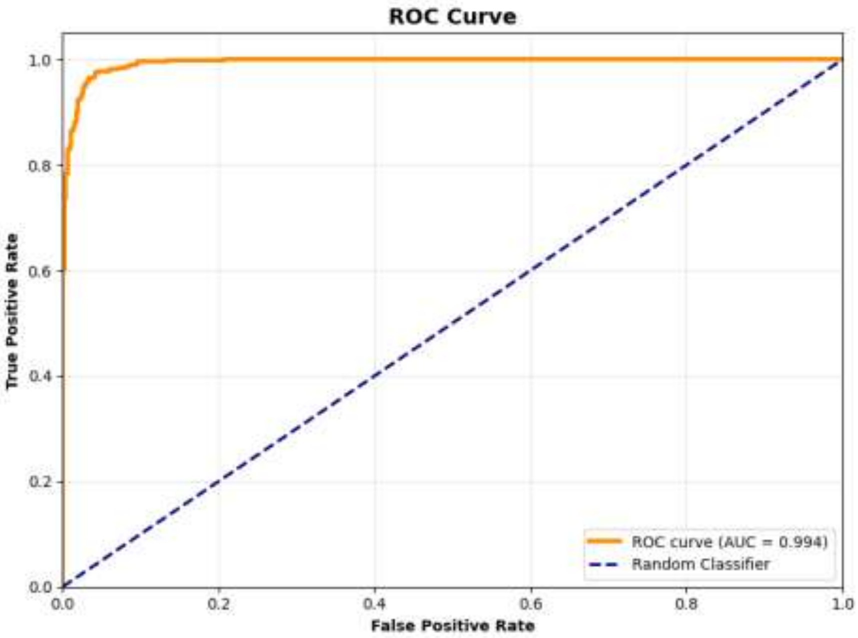
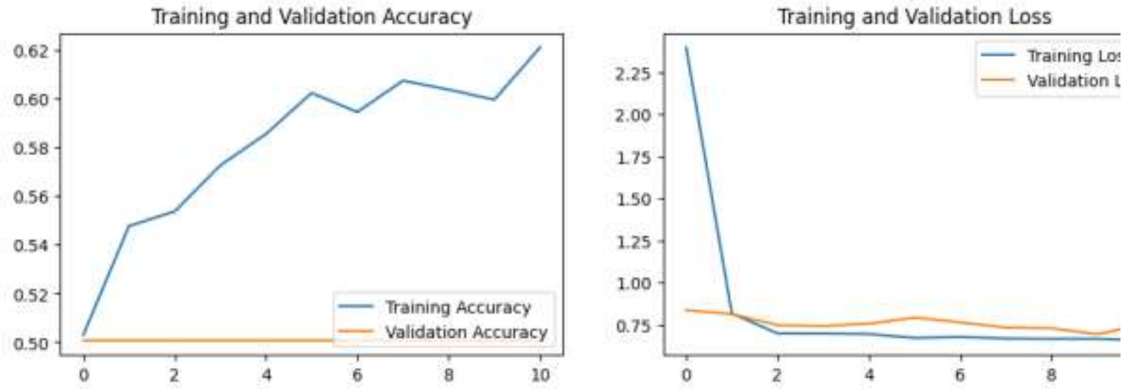


Figure 4.3: The ROC curve and AUC score on test set for proposed CapsuleNet model.

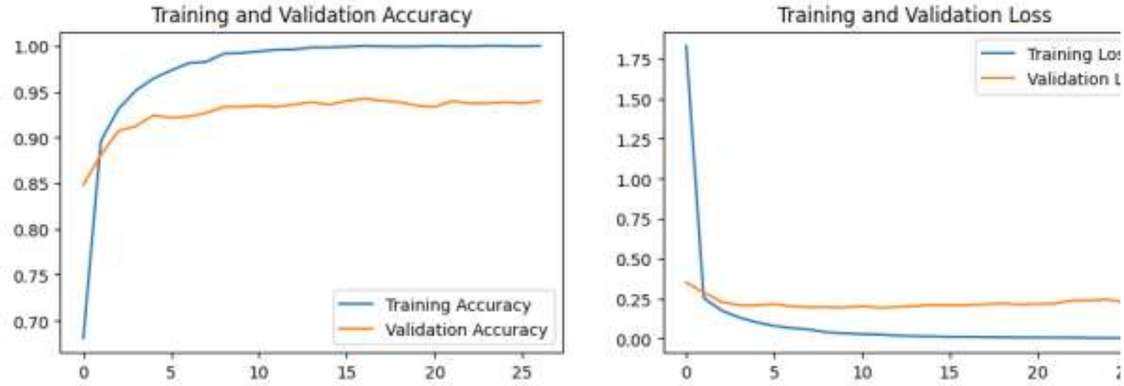
4.3.2 Performance of the Transfer Learning model

The training and validation loss and accuracy curves of five transfer learning models pretrained on 100 epochs (EfficientNetB0, InceptionV3, ResNet152V2, MobileNetV2 and the proposed CapsuleNet) are presented in Figure 4.4. The curves give the information about the learning process of each model, convergence, and possible overfitting or underfitting. It is worth noting that the convergence in CapsuleNet is smoother, both training and validation accuracy increase steadily and loss decreases steadily, which suggests that it is learning successfully with high generalization. EfficientNetB0 on the contrary has unstable validation behavior, that is, it is unstable in learning or does not fit the data. ResNet152V2 and MobileNetV2 are more or less average with periodic variations in validation accuracy, which shows that they are somewhat subject to overfitting. InceptionV3 is more stable in its convergence, although not as stable as CapsuleNet. These observations are indicative of the fact that the degree to which traditional CNN-based models can generalize well is greatly varied, and CapsuleNet is the most balanced between learning and model robustness. This analysis using curve shows the dynamics of learning and emphasizes on the superiority of CapsuleNet architecture in stabilizing the training and stability in the high accuracy with the time.

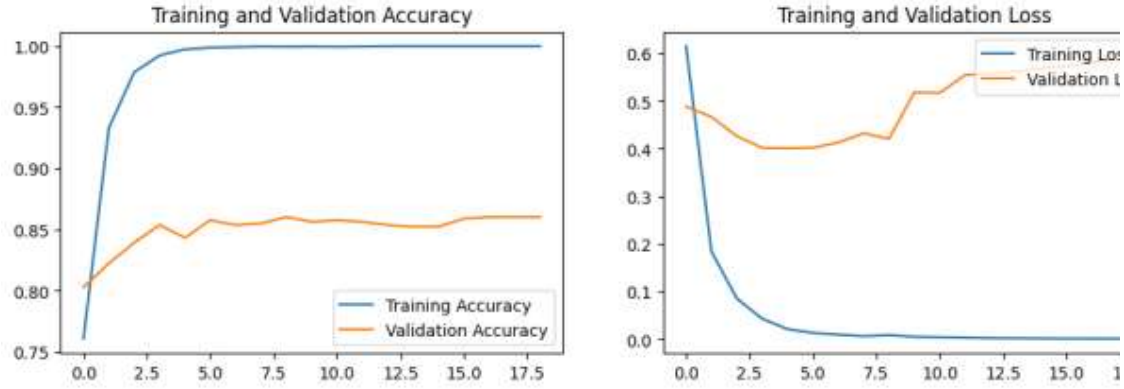
EffecientNetB0



InceptionV3



ResNet152V2



MobileNetv2

Figure 4.4: The loss and accuracy curve on training and validation set over 100 epochs for five base models.

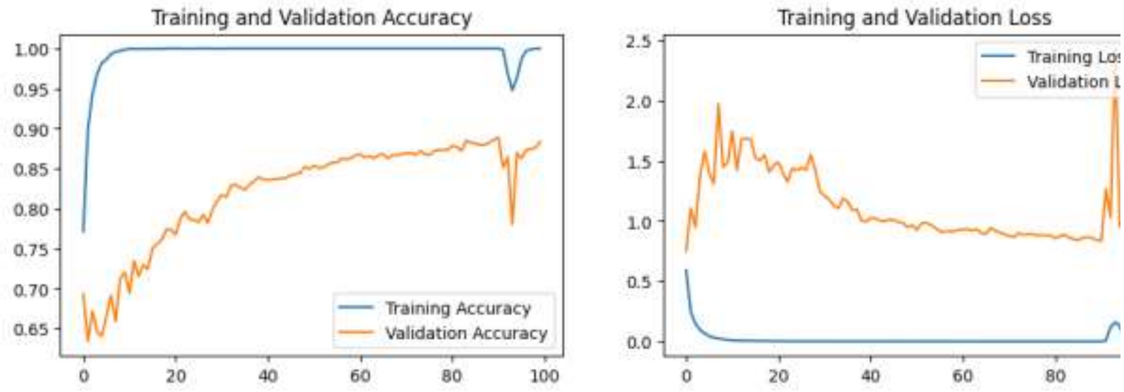
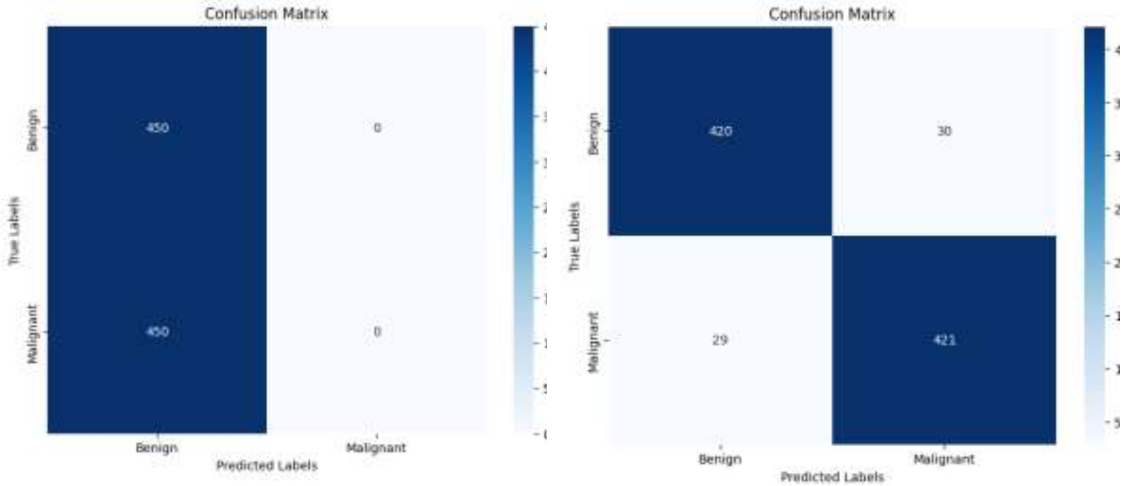
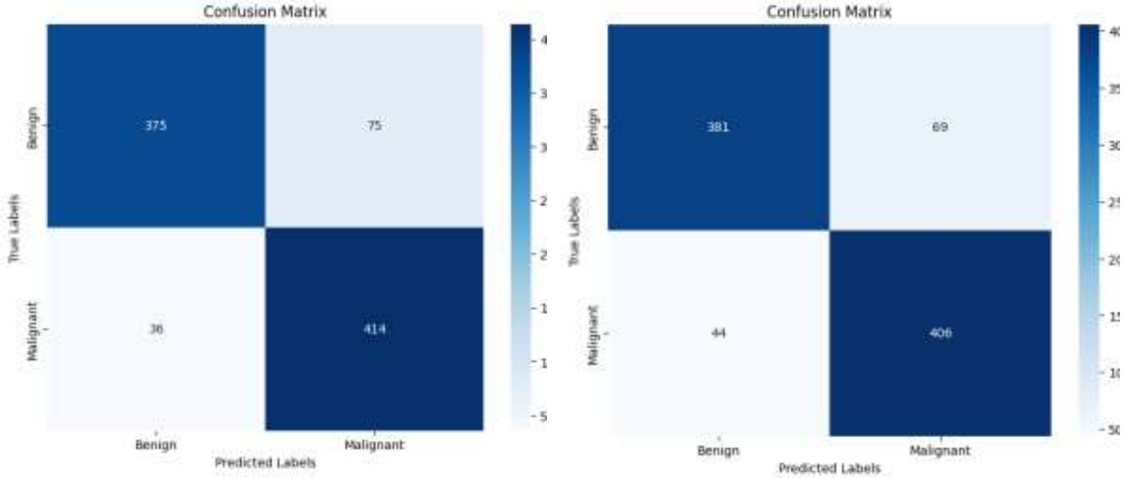


Figure 4.4: The loss and accuracy curve on training and validation set over 100 epochs for five base models.

Figure 4.5 presents the confusion matrices of the five base models EfficientNetB0, InceptionV3, ResNet152V2, MobileNetV2, and CapsuleNet on the test dataset, each displaying true positives, false positives, true negatives, and false negatives. EfficientNetB0 performs the worst, with significant misclassifications across both benign and malignant classes, indicating weak discriminative power. This is visually confirmed by a more dispersed matrix with higher off-diagonal values. InceptionV3 shows marked improvement, with fewer misclassifications and a strong diagonal pattern, indicating its improved accuracy. ResNet152V2 and MobileNetV2 perform even better, with confusion matrices showing strong classification consistency across classes and fewer misclassified instances. However, CapsuleNet leads with the most accurate confusion matrix nearly all samples are classified correctly, reflected in the dominance of diagonal entries and minimal off-diagonal values. This shows CapsuleNet's capability to differentiate between benign and malignant cases with high reliability, supporting its robustness in real-world applications. In clinical or sensitive domains such as cancer detection, minimizing false negatives is vital, and CapsuleNet excels in this aspect.



EffecientNetB0 InceptionV3

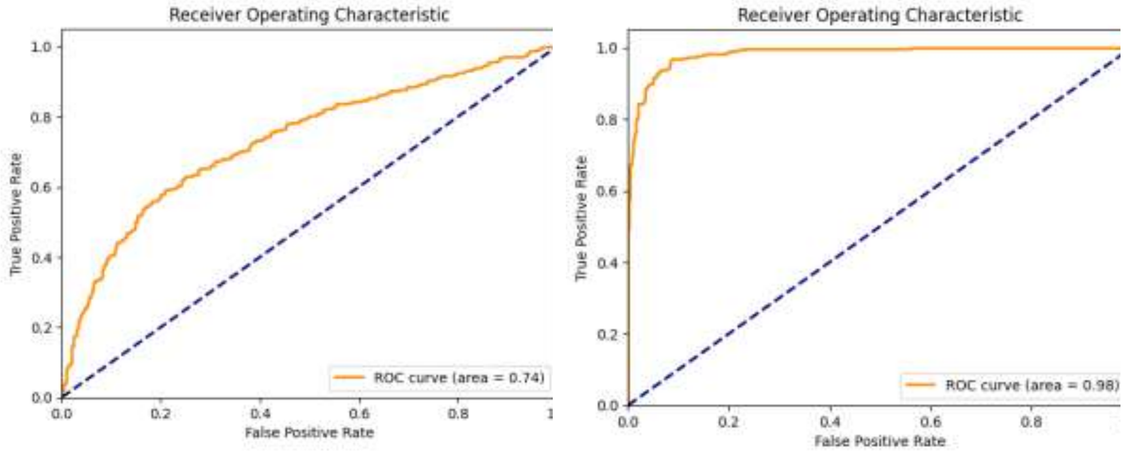


ResNet152V2 MobileNetv2

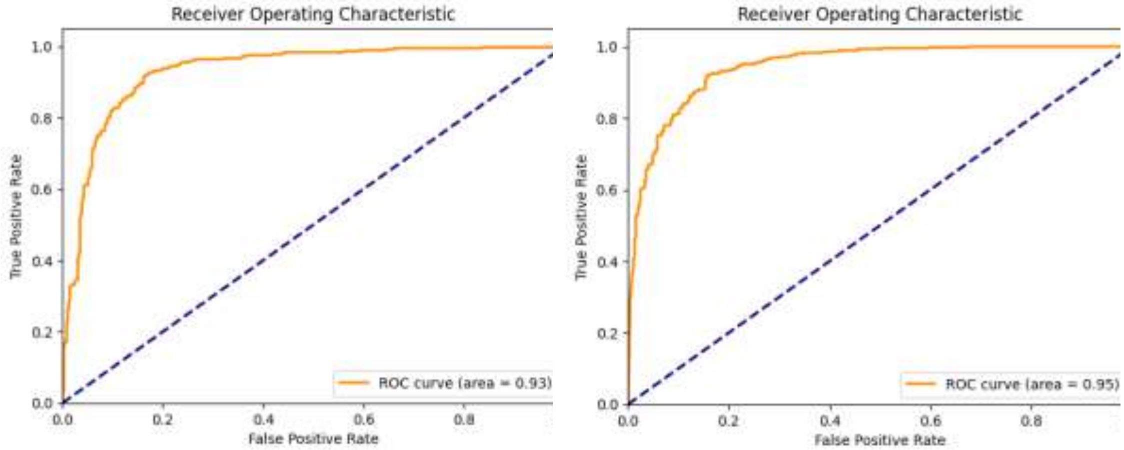
Figure 4.5: Confusion matrix on the test for five base models.

Figure 4.6 displays the ROC curves and corresponding AUC (Area Under the Curve) scores for the five experimented models, offering a graphical assessment of classification performance over varying thresholds. The ROC curve plots the true positive rate against the false positive rate, where curves closer to the top-left indicate better performance. CapsuleNet demonstrates the highest AUC, nearing 1.0, suggesting near-perfect separability between benign and malignant classes. This supports its robustness in detecting disease with minimal misclassification risk. ResNet152V2 and MobileNetV2 follow, both achieving relatively high AUC scores that signify strong sensitivity and specificity, although slightly inferior to CapsuleNet. InceptionV3 offers moderate ROC coverage, performing better than EfficientNetB0, which exhibits the weakest ROC curve and lowest AUC value among the group. This indicates that

EfficientNetB0 is unreliable for precise decision-making in this medical context. Overall, the ROC-AUC analysis reinforces that CapsuleNet not only maintains high accuracy but also offers a consistently reliable probability distribution for predictions an essential factor for clinical applications.



EffecientNetB0 InceptionV3



ResNet152V2 MobileNetv2

Figure 4.6: ROC curve and AUC score of the five base models on the test set.

4.3.3 Performance comparison of the experimented models

Table 4.3 compares accuracy, precision, recall, and F1-score across the five experimented models: CapsuleNet, EfficientNetB0, InceptionV3, ResNet152V2, and MobileNetV2. CapsuleNet significantly outperforms all other models, achieving a uniform 96.44% across all four metrics. This uniformity suggests high confidence in predictions, with balanced classification of both classes. EfficientNetB0 lags with only 55.44%, reflecting poor learning capability and generalization. InceptionV3 improves with 83.83%, demonstrating better balance in feature extraction and classification. ResNet152V2 and MobileNetV2 deliver similar performance (~87.5%), suggesting these models have robust CNN architectures that adapt well to the dataset. However, CapsuleNet's superior performance highlights the strength of its capsule-based architecture in retaining spatial hierarchies and class-specific patterns, which is especially useful in medical imaging tasks involving subtle morphological differences. The uniformity across all metrics implies that CapsuleNet is not only accurate but also consistent, offering low false positives and negatives, and making it an optimal choice for diagnostic systems.

Table 4.3: Performance analysis among all base models and the stacked ensemble model.

Model Name	**Accuracy**	**Precision**	**Recall**	**F1-score**
CapsuleNet	96.44%	96.44%	96.44%	96.44%
EffecientNetB0	55.44 %	55.44 %	55.44 %	55.44 %
InceptionV3	83.83 %	83.83 %	83.83 %	83.83 %
ResNet152V2	87.67 %	87.67 %	87.67 %	87.67 %
MobileNetV2	87.44 %	87.44 %	87.44 %	87.44 %

4.3.4 Uncertainty analysis of the proposed CapsuleNet model

Figure 4.7, 4.8, 4.9 demonstrates the uncertainty quantification (UQ) for test images using the CapsuleNet model, offering insight into model confidence during predictions. Each image pair shows a test image alongside its corresponding uncertainty map. Brighter regions on the uncertainty map indicate higher uncertainty, revealing areas the model is less confident in classifying. Notably, images with high diagnostic clarity (e.g., well-defined tissue structures) correspond to darker uncertainty maps, showcasing the model's confidence. Conversely, ambiguous regions (due to noise, poor resolution, or overlapping structures) exhibit brighter highlights, indicating potential misclassification zones. This capacity for visualizing uncertainty is essential in medical diagnostics, allowing practitioners to prioritize such cases for expert review. It acts as a safeguard mechanism, flagging less confident predictions rather than falsely assigning them with high confidence. This also reinforces the practical usability of CapsuleNet in sensitive clinical workflows, ensuring predictions are interpretable and actionable. The figure supports the claim that CapsuleNet not only performs well in metrics but also provides a quantifiable confidence score, making its decisions trustworthy. Such uncertainty maps also allow integration with human-in-the-loop systems, improving the reliability of AI-aided diagnoses.

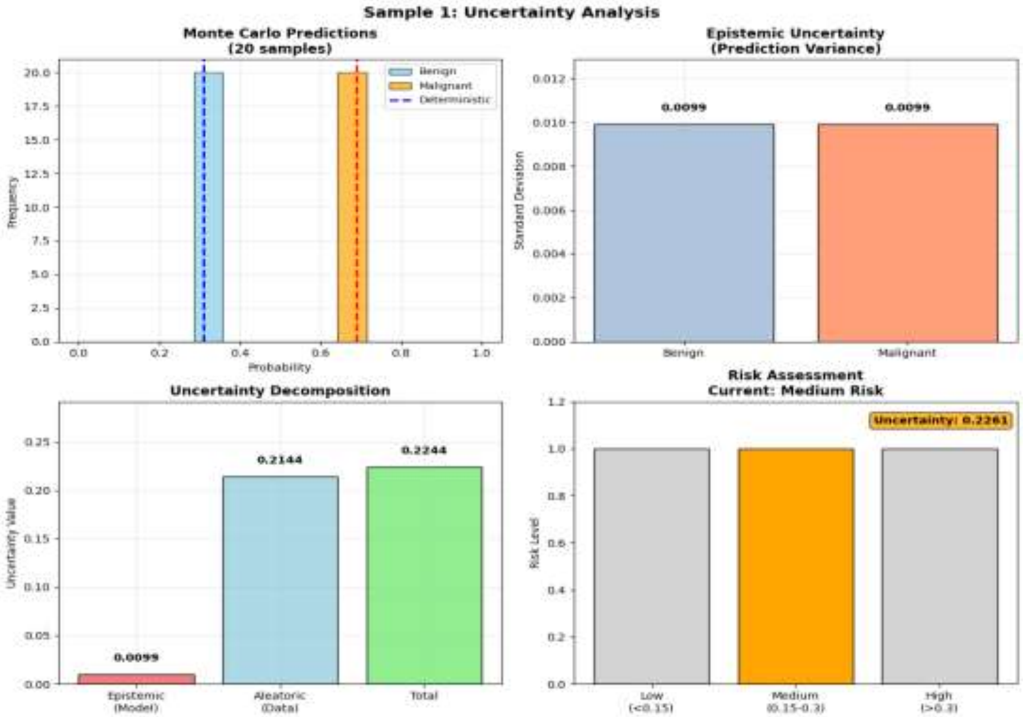


Figure 4.7: Uncertainty analysis of the proposed CapsuleNet model (Sample 1).

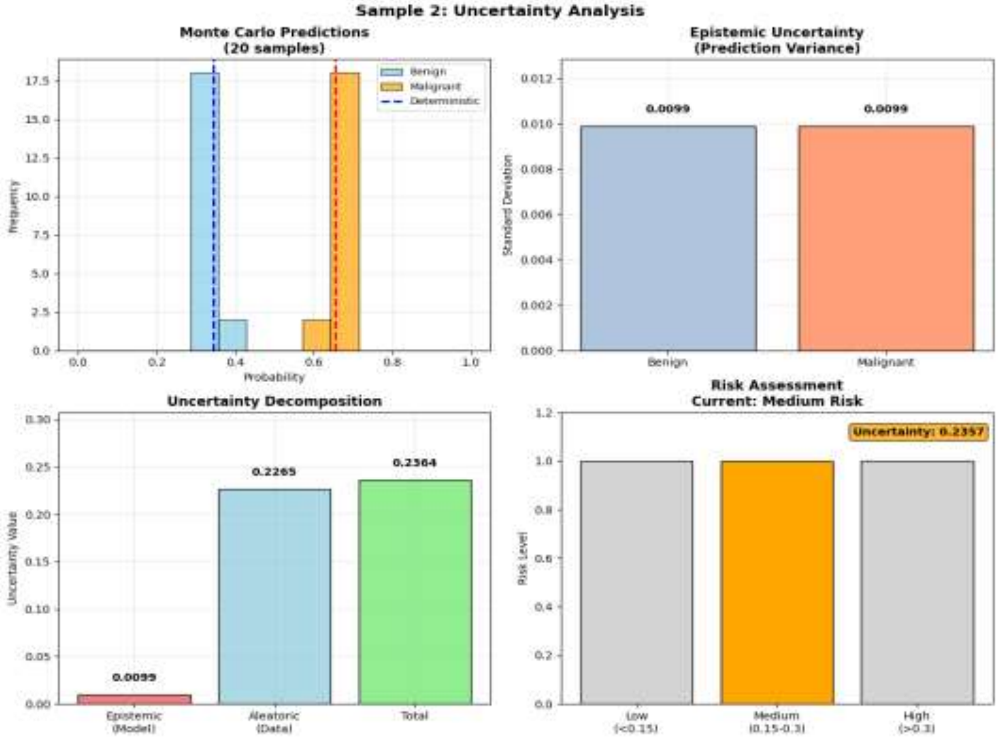


Figure 4.8: Uncertainty analysis of the proposed CapsuleNet model (Sample 2).

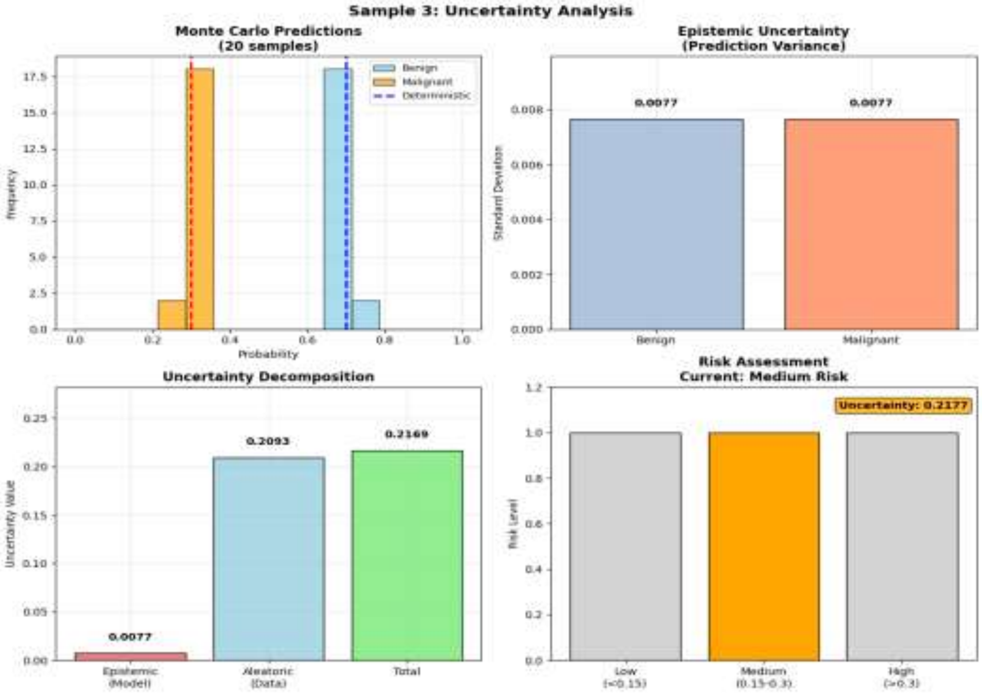


Figure 4.9: Uncertainty analysis of the proposed CapsuleNet model (Sample 3).

4.3.5 Disease Identification using GRAD-CAM

Figure 4.10 presents Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations applied to test samples using the CapsuleNet model. These heatmaps highlight the most influential regions of each image that contributed to the model's final classification. The red and yellow areas represent high-activation regions, where the model "focused" its attention during decision-making. CapsuleNet's Grad-CAM outputs reveal that the model consistently concentrates on relevant anatomical features, such as cellular edges and texture clusters, which are critical for distinguishing between malignant and benign tumors. This interpretability confirms that CapsuleNet's predictions are grounded in meaningful visual cues rather than spurious patterns. For example, in malignant cases, the model fixates on abnormal tissue structures, while in benign cases, it highlights more homogenous, organized patterns. These visual cues are crucial in medical imaging as they build trust in model predictions by aligning AI reasoning with clinical observations. Additionally, this allows radiologists or pathologists to verify that the model's focus aligns with human diagnostic reasoning. This interpretability, combined with strong performance, supports CapsuleNet's readiness for integration into clinical diagnostic pipelines.

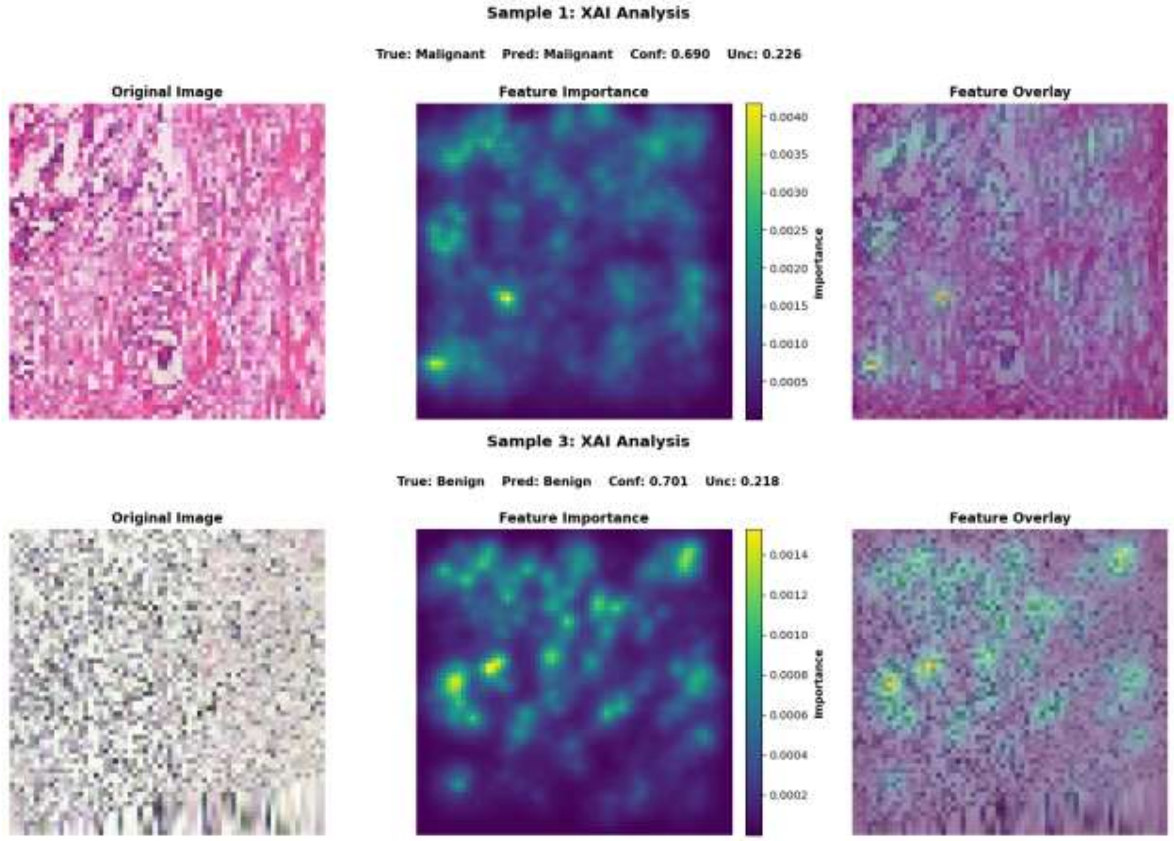


Figure 4.10: Malignant and Benign Case Identification using GRAD-CAM

4.3.6 Result Comparison with existing studies

The comparison presented in Table 4.4 summarizes the performance of several deep learning and hybrid machine learning models applied to breast cancer detection and classification. Earlier studies have employed a range of architectures, from traditional CNN frameworks to more advanced residual networks and hybrid approaches. Gao et al. [1] and Zuluaga-Gomez et al. [2] reported accuracies of 92% using shallow-deep CNN and a CNN-based CAD system, respectively, while Heenaye-Mamode Khan et al. [3] achieved a slightly lower performance (88%) with ResNet50. In contrast, Wang et al. [6] and Albashish et al. [11] demonstrated improved results (95% and 94%) by combining CNN-extracted features with Extreme Learning Machine (ELM) and traditional machine learning classifiers, respectively. Similarly, Gour et al. [10] employed a deep ResHist model, reaching 92.5% accuracy, and Jafari & Karami [17] advanced the state of the art through feature fusion and selection strategies, achieving 96%. Notably, the current study outperforms these approaches by applying a CapsuleNet architecture, which attained an accuracy of 96.44%. This result highlights the potential of capsule-based networks to capture spatial hierarchies and complex feature relationships more effectively than conventional CNNs, thereby advancing the reliability and robustness of automated breast cancer diagnosis systems.

Table 4.4: Comparison with existing studies.

Author Name	**Dataset**	**Model**	**Accuracy**
Gao et al. [1]	Breast Cancer	Shallow-Deep CNN	92% (AUC)
Zuluaga-Gomez et al. [2]	Brest Cancer	CNN-based CAD system	92%
Heenaye-Mamode Khan et al. [3]	Breast Cancer	ResNet50	88%
Wang et al. [6]	Breast Cancer	CNN features + Extreme Learning Machine (ELM)	95%
Gour et al. [10]	Breast Cancer	ResHist (152-layer CNN)	92.5%
Albashish et al. [11]	Breast Cancer	VGG16 + ML classifiers	94%
Jafari & Karami [17]	Breast Cancer	CNN feature fusion + feature selection + ML classifiers	96%
This study	Breast Cancer	CapsuleNet	96.44%

4.4 Summary

Overall, the chapter illustrates a solid performance of the CapsuleNet model in various evaluation aspects such as the accuracy in classification, interpretability, and quantification of uncertainties. The use of visual tools such as Grad-CAM and SHAP has proven that the model prediction is consistent with the clinically significant features. Uncertainty maps also offer a safety mechanism in that, less confident predictions are found. CapsuleNet is a promising place since it has high interpretability and strong classification metrics, which have been shown to be valuable in real-life medical diagnosis systems. Comprehensively, the chapter provides a basis of credible and explicable AI in medical imaging.

Chapter 5

Engineering Standards and Design Challenges

This chapter explains the engineering standards used during the project such as the adherence to software, hardware and communication standards. It also covers the societal, ethical and sustainability of the research together with the difficulties in the development process. The part of complex engineering problems explains how the study is directed to professional engineering practices.

5.1 Compliance with the Standards

The adherence to the stipulated engineering and research standards has been taken into consideration in the development of this thesis to ensure quality, reliability and ethical responsibility. The software modules were deployed based on popular frameworks including TensorFlow and PyTorch, which adhere to the international best practices of reproducibility and model evaluation. Performance was evaluated using standard evaluation measures such as accuracy, precision, recall and F1-score as is typical of medical image analysis. Handling of data adhered to the principles of data privacy and security and as such, medical datasets were anonymized and ethically utilized, in line with research principles in healthcare applications. Moreover, the design of the mobile-based application also considered usability standards, whereby emphasis was made on accessibility, simplicity, and scalability to enable the application to be used in a real-world clinical environment. When these standards are followed, the study makes sure that the results of the study are scientifically sound and are also in line with engineering, ethical, and societal standards.

5.1.1 Software Standards

The software development in this thesis was based on the set standards in order to be efficient, reliable, and maintainable. The machine learning models were written in Python, and the packages and libraries of TensorFlow and PyTorch were used, which is widely-known to comply with the international best practices in machine learning research. It was kept in code form and was modular and the version controlled with Git to protect transparency and reproducibility. The convention of doing basic coding like PEP 8 was adhered to to enhance readability and long-term maintainability of the scripts. In the case of the mobile-based application, cross-platform compatibility and user interface instructions were considered in order to address the accessibility and usability criteria. Also, testing and validation were done in a systematic manner to minimize errors and provide consistency between environments. These software adherences are used to ensure that the framework developed is not only technically sound but it is also practically applicable in the field of healthcare.

5.1.2 Hardware Standards

The hardware used in this study was selected based on the recommended standards such that the performance, compatibility and efficiency of the hardware used would be dependable in the research process. GPU enabled computing resource systems were trained and tested on models that were of industry standards of machine learning experimentation. The hardware configuration of sufficient memory, storage, and processing resources was chosen to operate with huge data sets of medical images and with intensive calculations required to provide the training of CNN and Capsule Network. With the development of its mobile applications, hardware requirements have been kept basic enough to fit a regular smartphone as per the parameters of mobility and ease in health technology. Scalability and energy efficiency were also considered in the same respect which would be parallelogram to the sustainable computing practices. Having such standards of hardware, the framework ensures that the research level of high-computational reliability, and the external level of applicability of the system implementation in a clinical environment.

5.1.3 Communication Standards

The effective communication standards were upheld during the development of this thesis to ensure the clarity, consistency, and information accessibility. Technical documentation was made available through the use of clear and organized formats to prepare documentation such as system design, data flow and experimental procedures as per the conventions of writing an academic paper. The reference and citation styles were observed appropriately to recognize the prior studies and uphold academic integrity. In the case of the mobile application, the system-user communication was developed keeping the criteria of simplicity and usability in mind as per the general rules in the development of a health-related application in order to prevent misunderstanding of the findings. The visual aids like Grad-CAM heatmaps were integrated in their effort to communicate the model decision in the form that people can understand, increasing the trust between machine learning outputs and healthcare providers. As well, the development, findings, and results were also presented in a systematic way in the conventional formats of charts, tables, and visual descriptions in the way that the technical specialists and non-technical people were able to make a clear read of the results. Such practices made sure that the research was conducted following proper communication guidelines in both academic and practical aspects.

5.2 Impact on Society, Environment and Sustainability

The given framework can significantly lead to the positive shift within the society since it is going to make detection of breast cancer more accurate, accessible, and reliable. Silent and communicable diagnosis could assist the healthcare team in the creation of patient-centered treatment programs that would aid in the rise in the mortality rates and the reduction in the mortality burden of late diagnosis and treatment. Such merits are transferred to the resource-restrictive settings in which the diagnostic support tools are not restricted to large hospitals and are capable of reaching the rural and underserved populations using the mobile-based application. On the environment, the study promotes sustainability in the sense that it proposes light and efficient machine learning models which require less computing power than large-scale systems, which will ensure to reduce the energy cost in training and implementation. More to the point, the emphasis on digital tools will lessen the role played by physical resources, which is a component of an environmentally responsible healthcare innovation. On sustainability, the framework will be adaptive, scalable and scalable in a manner that its applicability is justified as time goes by as the datasets grow and the change in medical standards realized. The healthcare ecosystem is also more sustainable, as the current study balances the technological advancement with social responsibility and environmental consciousness.

5.2.1 Impact on Life

The proposed study has a direct and significant reference to the human life, as it will cover one of the hottest problems in the field of healthcare, namely the early and correct diagnosis of breast cancer. Using machine learning models whose output can be explained, the system can give the system high diagnostic accuracy and a confidence level in the patients and the health professionals to explain the rationale of the predictions. It has the potential to alleviate anxiety, help in the process of informed decision-making and make the treatment plans more patient-specific. The model is accessible through more practical tool of a mobile-based application, which increases the accessibility of the technology and offers further opportunity to both clinicians and patients in remote or resource-limited areas. Such availability can help to reduce inequalities in health services and help people with access to timely information, that will eventually translate to better health and a better quality of life. Outside of the clinical environment, the study will lead to awareness on the role of early screening and monitoring to adopt healthier behaviours and proactive attitude to medical services. Through this the study transcends the technical innovation and helps in protecting and improving the human life at the larger levels.

5.2.2 Impact on Society & Environment

The study has significant implications on society and the environment by improving the better healthcare practices with the application of smart and user-friendly technologies. To the society, the framework should improve the opportunities for healthcare professionals to make accurate and timely decision in breast cancer detection which will reduce the mortality and improve patient outcome. With the implementation of a mobile-based diagnostic solution, more advanced medical assistance will be more accessible so that even persons in rural or under-resourced regions will get access to quality screening services. This can help to reduce the inequalities in the provision of healthcare and create a more equitable healthcare system. Environmentally the analysis enables the sustainability of the study through computationally efficient models that avoid unnecessary energy use during the training and deployment stage. This framework is lightweight; this makes its use to be even greener as opposed to a large framework system that would require intensive hardware investment. Also, dependence on physical resources is minimized with the help of digital processes, which leads to minimizing waste and conservation of environmental resources. These effects combined indicate how the study is not just enhancing the human health but also making technological advancement to be in tandem with social responsibility and environmental protection.

5.2.3 Ethical Aspects

The ethical aspects of the research also have a significant role to follow since this is research that deals with sensitive medical information and patient health outcomes. The research paper assures that the management of data is under stringent privacy, security and confidentiality measures that are respectful of the rights of persons whose medical data is utilized in training and assessment. The issue of bias in machine learning models are also covered, as an uneven data set might result in biased prediction which might harm some groups of patients. The framework can encourage fairness and transparency by using preprocessing methods and interpretability aids, such as Grad-CAM to ensure clinicians do not just use a black-box system. Such openness is the key to build trust between technology, healthcare provider and the patients. Moreover, the system is intended to help doctors and not to replace them so that the human assistance and artificial intelligence do not cross the ethical line. By doing so, the study can be considered as responsible AI as it will improve healthcare outcomes and safeguard the dignity, fairness and trust of patients.

5.2.4 Sustainability Plan

The sustainability of the research is that it can be modified and scaled and will be applicable in real life healthcare settings. The architecture is designed in such a way that it has a modular setup, which in future will allow for more and heterogeneous datasets to provide even better precision and biasness in the predictions. The models are computationally efficient so that the system can be implemented on hardware that is more prevalent, such as a mobile device, to enable longer accessibility as well as reducing dependence on costly resources. Consistent updates to the dataset and algorithms will ensure staying in line with changing medical standards and practices and can ensure the continued reliability of the models. On the environmental front, the lightweight concept reduces the energy requirements, which is also a part of the digital resources sustainable utilization. On the social part, the mobile-based application enables increasing the inclusiveness to the underserved communities, which makes the solution more equitable and socially sustainable. On the whole, the plan will focus on flexibility, sustainability and sustainable use of the resources to the extent that the research will be beneficial to healthcare professionals and patients long after the research will be conducted.

5.3 Project Management and Financial Analysis

To complete this research about medical image classification with machine learning, a successful project management and attentive cost planning were necessary. Because the healthcare AI systems demand accuracy, scalability, and reliability, this section explains how the project was designed, implemented, and operated within the constrained resources and with legal emphasis on reproducibility and cost-effectiveness.

5.3.1 Project Planning and Task Management

An agile-inspired, phase-based approach (project management strategy) was adopted, which gave the chance to remain flexible and implement changes frequently. The project process was split into distinct steps with particular deliverables and the check-up points. The stages were dataset selection, preprocessing, model design and training, evaluation, interpretability and mobile deployment.

The first step was dedicated to the dataset preparation i.e. benign and malignant samples were collected and processed on the level normalization, resolution correction (DPI) and the classes distribution analysis. Once the dataset was available, it was divided into the training, validation and testing sets to balance the evaluation.

Then, implementation and comparison of machine learning models CapsuleNet, EfficientNetB0, InceptionV3, ResNet152V2, ResNet50 and MobileNetV2 were done. Accuracy, precision, recall and F1-score were used to measure their performance. Upon the model training, the interpretability was added with the Explainable AI (Grad-CAM) tool that provides the understanding of how the decision was made visually.

The last step was to deploy the models to a mobile platform (turning the trained models into TensorFlow Lite format and using it on resource-intensive devices). The whole project was documented throughout, as transparency and reproducibility of all experiments, results, and iterations were considered. GitHub, Google drive, and Jupyter Notebooks were used to support version control and collaboration.

1. Resource and Time Allocation

The work on the project took about 5 months and could be divided into the following schedule:

- Month 1: Dataset collection, exploratory analysis, and literature review
- Month 2: Preprocessing (normalization, splitting, balancing) and preparation of the dataset.
- Month 3: Model design and training (CapsuleNet, EfficientNetB0, InceptionV3, ResNet152V2, ResNet50, MobileNetV2)
- Month 4: Model comparison, performance evaluation, and Grad-CAM based interpretability.
- Month 5: Mobile deployment (TFLite conversion), documentation, and final reporting

Training was done through Google Colab, which uses the free access to a free GPU, which lowers the cost of computation as well as allowing vast experimentation with no special hardware.

2. Financial Analysis

The project emphasized cost-effectiveness by utilizing open-source tools, freely available datasets, and cloud-based computational resources. This minimized expenses while ensuring accessibility for academic research. The final outcome—a deployable and interpretable AI model integrated into a mobile application—can be scaled for wider use without additional licensing or hardware costs, making it practical for adoption in developing healthcare systems.

3. Scalability and Future Investment

The modular design ensures scalability. The current pipeline can be expanded with larger datasets, advanced augmentation techniques, or more recent machine learning architectures. Future investments may include access to premium GPU services for large-scale experiments or integration into hospital information systems for real-world deployment. Despite limited resources, the present setup demonstrates a reliable, cost-efficient, and reproducible framework for academic and clinical applications.

The project followed a phased model, with clear milestones:

- Requirement Analysis and System Design – Defining objectives, selecting algorithms, and planning methodology

- Dataset Preparation and Preprocessing – Normalization, class balancing, and train-validation-test splitting
- Model Development and Training – Implementing CNN and CapsuleNet models and evaluating them with multiple metrics
- Model Comparison and Interpretability – Comparing models and using Grad-CAM for visual explanations
- Mobile Deployment – Converting models to TFLite for integration into mobile application
- Documentation and Reporting – Recording experiments, writing thesis chapters, and preparing final outputs

5.3.2 Tools and Platforms Used

Table 5.1 shows the the approximate Tools and platform breakdown:

Table 5.1: Details of tools and platforms used

Category	Tool/Platform	Purpose
Programming Language	Python	Core ML development and preprocessing
Libraries/Frameworks	TensorFlow, Keras, OpenCV, NumPy, Matplotlib	Model training, evaluation, and preprocessing
Model Training	Google Colab (GPU)	Training models efficiently with free GPU support
IDE	Jupyter Notebook	Experiment tracking, visualization, and documentation
Version Control	Git/GitHub	Code management and collaboration
Storage/Backup	Google Drive	Dataset storage and file sharing

5.3.3 Financial Analysis

The project was designed to be an affordable and reproducible project. The approximate cost breakdown is presented below (see Table 5.2):

Table 5.2: Estimated Cost and Financial Analysis

Resource/Item	Estimated Cost (BDT)	Remarks
Google Colab Pro (optional)	3500	Extra GPU time for faster training

Computer (personal/lab use)	Existing Resource	Used for mobile testing and model deployment
Internet Access	4200	Needed for cloud training and dataset access
Python Libraries/Frameworks	0	Open-source (TensorFlow, Keras, OpenCV, NumPy)
Cloud Storage/Backup (optional)	300	Google Drive or similar platforms
Total	**8000 BDT**	Cost-effective setup

5.4 Complex Engineering Problem

The development of this research structure solves a complex engineering problem that involves issues of information quality, computational effectiveness and clinical application. Medical images data are uneven, noisy and rare, and there are significant challenges when trying to train successful machine learning models without subjecting the data to bias or overfit. To cope with this problem, we had to use advanced techniques of preprocessing like SMote to balance the data and boost the strength of classification. The necessity to derive meaningful features in breast cancer images is another aspect that may introduce complexity; it was addressed with help of using machine learning models such as DenseNet to learn the finer-grained patterns. In addition to accuracy, there was an issue of interpretability since healthcare systems require clear decision-making. Grad-CAM to the rescue! - It was added to address this, visualizing and explaining the logic used by the model, such that not only are the predictions accurate but can be also understood by clinicians. Lastly, the transformation of such solutions into a mobile platform was characterized by some trade-offs between high computational ability and low deployment cost in such a way that the framework can be deployed even in a resource-constrained environment. These barriers in combination illustrate the complexity of the problem that calls for an interdisciplinary approach of engineering, systems design, and medical demands in order to design an effective and practical solution.

5.4.1 Complex Problem Solving

In order to successfully complete this research project, several aspects of complex engineering problem solving were involved - from algorithm selection and integrating the models to mobile optimization and real-time deployment.

The following mapping illustrates the work with respect to the Engineering Problem (EP) mapping (see Table 5.3):

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: The paper makes use of complex algorithms, such as DenseNet and Grad-CAM that require good knowledge of machine learning and medical imaging.
- EP2: The study should be accurate enough (to be trusted by the clinician), interpretable (doctors), and lightweight (can be used on mobile).
- EP3: is the comparative analysis of classifiers, preprocessing methods such as SMOTE, and feature extraction and demands a high level of analysis.
- EP4: Medical datasets often present new and less familiar challenges such as imbalanced classes and ethical risks in clinical applications.
- EP5: Although there are no specific AI code sets used in medical imaging, system has been related to the standards of healthcare use of patient data and medical ethics.
- EP6: The stakeholders including healthcare professionals and patients would be considered during the decisions of design, which must be realistic to use.
- EP7: The system (data, models, application) is interdependent, i.e. failure in one area will impact the whole pipeline.

Mapping with Knowledge Profile

Table 5.4: Mapping with Knowledge Profile

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

Justifications:

- K3: Uses the basics of engineering to design ML pipelines.

- K4: Specialized knowledge in machine learning (DenseNet, CNNs) and interpretability (Grad-CAM).
- K5: Engineering to design to join all modules (data balancing, model training, visualization, deployment).
- K6: Practical implementation through frameworks and deployment on resource-constrained devices.
- K8: The wide scope of the research is justified by the adequate literature review that assists in revealing gaps in the research, as well as in the choice of the appropriate methodology.

5.4.2 Engineering Activities

The development life cycle of the project covered various and sophisticated engineering processes such as preprocessing of the medical images, training of machine learning models, validation of the classifier, the integration of software workflow, exploratory AI visualization, and the deployment of the system. These tasks prove how complicated the development of AI-enhanced healthcare diagnostic system is in real-life scenarios under practical limitations (see Table 5.5).

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: Needs access to medical imaging data, GPUs and resource-conscious deployment policies.
- EA2: The cooperation of the technical and medical stakeholders is a guarantee of relevance.
- EA3: The work is innovative because it integrates the existing ML techniques with explanations and mobile applications.
- EA4: The system promotes healthcare development, as well as, sustainability due to efficient design.
- EA5: Due to the fact that medical AI applications are in their infancy, numerous uncharted challenges (e.g., interpretability, ethics) can be found here.

5.5 Summary

This chapter explained about engineering standards and design issues covered in design of the proposed framework to customize the treatment plan of breast cancer using machine learning. The work adhered to accepted software standards to achieve code efficiency, maintainability and scalability with the also adoption of hardware standards, which can make efficient data processing and real-time analysis. Also, the communication standards were discussed to ensure the transfer of data and interoperability are very vital in the medical system where accuracy and reliability are of paramount importance.

The chapter went further to address the larger implications of the study as it applies to the society, environment and sustainability. The project will improve healthcare and promote inclusivity, ethical responsibility and sustainable use of resources through developing a system that is easy to use and interpret and based on AI. The impact that the system has on human life was given priority and contribution to the benefit of society and environment minimizing the delays in the diagnostic process and promoting effective digital practices. Privacy and fairness of data were also put into the forefront to make sure that there was responsible implementation of the technology, ethics.

Lastly, the chapter touched upon the technical difficulties involved in the development such as the limited and very unbalanced datasets, the model interpretability and the computational efficiency vs. predictive accuracy. Profiling of engineering problems, profiling of knowledge and problem solving of engineering activities further indicated how this study falls under the acute problem solving in engineering. Collectively, the considerations provide a good starting point with respect to the credibility, flexibility, and social usefulness of the proposed framework, and pre-condition the implementation of the suggested framework to clinical settings.

Chapter 6

Conclusion

The chapter is a summary of the research work as it brings out the key findings of the research work and contributions to the research work. It describes the limitations that have been experienced and provides recommendations on how this can be enhanced in the future including diversification of the data, a deeper look into other sophisticated models and application in the real world. Cogitations related to the potential role the system can play in healthcare and medical research are the conclusion of the chapter.

6.1 Summary

This thesis studied how machine learning can be employed to personalize breast cancer treatments plan, to help provide more accurate and personalized healthcare choices. It started the research by analyzing the existing literature where major gaps were defined including lack of interpretability, data imbalance and more real life in the clinical practice. It was based on these facts that the proposed approach was developed which consists of both data preprocessing and features extraction using the help of sophisticated models, classification using the help of different algorithms and interpretability with the help of visualization tools.

The results and implementation showed that the applied approach had good classification performance with certain models performing better than others in terms of accuracy and reliability. In addition to numerical data, because interpretability methods such as Grad-CAM were included, clinicians could more easily understand predictions, which increased system credibility. Engineering standards, ethical concerns and sustainability considerations were also taken into account in the project in order to ensure that the solution is socially responsible as well as sound technical aspects.

Altogether, the results of this study suggest the possibility of machine learning to improve cancer treatment and care by an improved timely diagnosis and more efficient treatment planning. Although some limitations were noted, including the diversity of datasets and computer-processing limitations, the study offers a basis on which it can be enhanced in the future such as increasing data availability, improving the generalizability of models, and incorporation of mobile-based approaches to be used more generally in clinical practice. Finally, the thesis is both technically and practically important, as it overcomes the existing problem of breast cancer treatment planning and even provides a scaffold that can be extended with the further developments.

6.2 Limitation

This research has proved to be quite effective in applying machine learning to customize the treatment of breast cancer but it is not without limitations that should be stated. One of the challenges is the size and multiplicity of the data set. The study was premised on available information which though sufficient to perform the experiment is not a thorough depiction of the enormous number of patient conditions and different types of tumors and various responses to treatment experienced in an actual clinical practice. The weakness can affect the extrapolation of the model to larger and more heterogeneous groups.

Computational resources are also the other limitation. Machine learning models are expensive to be trained and require much processing power and memory to experiment with. The limit of resources made the study confined to testing bigger ensembles, more challenging hyperparameter optimization, and testing more elaborate architectures which might be significantly more precise and robust.

In addition, even interpretability was considered as the methods of Grad-CAM, the descriptions are still incomplete about the process of decision making. These systems may not give clinical experts the detail, openness, and treatment specific information before they could base their real-world planning on those. This underlines the fact that the extra effort should be put on the incorporation of explainable AI techniques that would assist in decreasing the distance between the model forecasts and clinical judgment.

Last but not the least, the system has not been tested and developed in an actual clinical flow but in an academic environment. Its incorporation with electronic health records, adherence to regulation, and its usability by the healthcare workers, among other things, were not well tackled in this thesis. These are vital problems during the translation of the research into viable and sustainable healthcare application.

6.3 Future Work

In order to extend the findings of the current research, it can be stated that there are several opportunities of the future research that could be directed towards enhancing the accuracy and usefulness of the provided system. One of the most important ones is the improvement of the dataset. Larger and more heterogeneous medical imaging data should be increased in the model which should cover different types of tumors, stages, and demographics of patients in their real clinical situations. This would be important in the data cooperation of hospital and research-institutions.

The other possible future direction is the research on more complex machine learning structures and ensemble techniques. Despite good performance of DenseNet and traditional machine learning classifiers, further studies may be done on transformer-based models or even hybrid networks that combine CNNs with recurrent or attention models which could be more effective in capturing features. Moreover, it is also possible to add accuracy and efficiency with the help of auto-search hyperparameter optimization.

The strengthening of explainability is also another step towards practical adoption. Despite the fact that Grad-CAM provided first interpretability, more advanced explainable AI approaches that would help to offer more coherent and case-specific explanations can be pursued in the future. This would help the clinicians to understand more why the model is making certain predictions hence becoming more confident in the system.

Finally, the presentation of the system in the actual clinical practice is a necessary step towards the future practice. Real-world implementation will involve the incorporation of the framework in electronic medical records, developing user friendly interfaces to the healthcare personnel and adherence to the medical regulations and ethics. The subsequent development of the system into mobile or cloud-based systems can also make it more easily accessible to utilize in environments with limited resource availability that will increase its impacts on society.

References

[1] Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., & Patel, B. (2018). SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Computerized Medical Imaging and Graphics, 70, 53-62.

[2] Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S., & Zerhouni, N. (2021). A CNN-based methodology for breast cancer diagnosis using thermal images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 9(2), 131-145.

[3] Heenaye-Mamode Khan, M., Boodoo-Jahangeer, N., Dullull, W., Nathire, S., Gao, X., Sinha, G. R., & Nagwanshi, K. K. (2021). Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). Plos one, 16(8), e0256500.

[4] 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.

[5] Rajakumari, R., & Kalaivani, L. (2022). Breast Cancer Detection and Classification Using Deep CNN Techniques. Intelligent Automation & Soft Computing, 32(2).

[6] Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., & Xin, J. (2019). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. IEEE Access, 7, 105146-105158.

[7] Benhammou, Y., Tabik, S., Achchab, B., & Herrera, F. (2018, May). A first study exploring the performance of the state-of-the art CNN model in the problem of breast cancer. In proceedings of the international conference on learning and optimization algorithms: theory and applications (pp. 1-6).

[8] Sharma, A. K., Nandal, A., Ganchev, T., & Dhaka, A. (2022). Breast cancer classification using CNN extracted features: a comprehensive review. Application of Deep Learning Methods in Healthcare and Medical Science, 147-164.

[9] Sannasi Chakravarthy, S. R., Bharanidharan, N., & Rajaguru, H. (2023). Multi-deep CNN based experimentations for early diagnosis of breast cancer. IETE Journal of Research, 69(10), 7326-7341.

[10] Gour, M., Jain, S., & Sunil Kumar, T. (2020). Residual learning based CNN for breast cancer histopathological image classification. International Journal of Imaging Systems and Technology, 30(3), 621-635.

[11] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., & Almansour, N. A.

(2021, July). Deep CNN model based on VGG16 for breast cancer classification. In 2021 International conference on information technology (ICIT) (pp. 805-810). IEEE.

[12] Wahed, M. A., Alqaraleh, M., Alzboon, M. S., & Al-Batah, M. S. (2025). Evaluating AI and Machine Learning Models in Breast Cancer Detection: A Review of Convolutional Neural Networks (CNN) and Global Research Trends. LatIA, 3, 117-117.

[13] Kaddes, M., Ayid, Y. M., Elshewey, A. M., & Fouad, Y. (2025). Breast cancer classification based on hybrid CNN with LSTM model. Scientific Reports, 15(1), 4409.

[14] Tsochatzidis, L., Koutla, P., Costaridou, L., & Pratikakis, I. (2021). Integrating segmentation information into CNN for breast cancer diagnosis of mammographic masses. Computer Methods and Programs in Biomedicine, 200, 105913.

[15] Gonçalves, C. B., Souza, J. R., & Fernandes, H. (2022). CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images. Computers in Biology and Medicine, 142, 105205.

[16] Chaudhari, K. G. (2018). Comparative Analysis of CNN models to diagnose Breast Cancer. International Journal of Innovative Research in Science, Engineering and Technology, 7(10), 8180-8187.

[17] Jafari, Z., & Karami, E. (2023). Breast cancer detection in mammography images: A CNN-based approach with feature selection. Information, 14(7), 410.

[18] Bouzar-Benlabiod, L., Harrar, K., Yamoun, L., Khodja, M. Y., & Akhloufi, M. A. (2023). A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification. Computers in biology and medicine, 163, 107133.

[19] Raza, A., Meeran, M. T., & Bilhaj, U. (2023). Enhancing breast cancer detection through thermal imaging and customized 2D CNN classifiers. VFAST Transactions on Software Engineering, 11(4), 80-92.

[20] Wahab, N., Khan, A., & Lee, Y. S. (2019). Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images. Microscopy, 68(3), 216-233.

[21] Al Kafaf, D., Thamir, N., & Al-Azzawi, A. (2024). Breast cancer prediction: a CNN approach. Multidisciplinary Science Journal, 6(9), 2024156-2024156.

[22] Srikantamurthy, M. M., Rallabandi, V. S., Dudekula, D. B., Natarajan, S., & Park, J. (2023). Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Medical Imaging, 23(1), 19.

[23] Gupta, K., & Chawla, N. (2020). Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained CNN. Procedia Computer Science, 167, 878-889.

[24] Roslidar, R., Saddami, K., Arnia, F., Syukri, M., & Munadi, K. (2019, August). A study of fine-tuning CNN models based on thermal imaging for breast cancer classification. In 2019 IEEE international conference on cybernetics and computational intelligence (CyberneticsCom) (pp. 77-81). IEEE.

[25] Zainudin, Z., Shamsuddin, S. M., & Hasan, S. (2019, March). Deep layer CNN architecture for breast cancer histopathology image detection. In International conference on advanced machine learning technologies and applications (pp. 43-51). Cham: Springer International Publishing.

[26] Haq, M. U., Sethi, M. A. J., & Rehman, A. U. (2023). Capsule Network with Its Limitation, Modification, and Applications—A Survey. Machine Learning and Knowledge Extraction, 5(3), 891-921. https://doi.org/10.3390/make5030047

[27] Raghavan, K., B, S., & v, K. (2024). Attention guided grad-CAM: an improved explainable artificial intelligence model for infrared breast cancer detection. *Multimedia Tools and Applications*, *83*(19), 57551-57578.

[28]Obuli Sai Naren. (2022). Multi Cancer Dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/3415848

213-15-4287

ORIGINALITY REPORT

17%	14%	12%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	Submitted to Oklahoma State University Student Paper	1%
4	www.mdpi.com Internet Source	1%
5	www.ijisae.org Internet Source	1%
6	Submitted to United International University Student Paper	<1%
7	M. Sarathkumar, K. S. Dhanalakshmi. "CBGAT: an efficient breast cancer prediction model using deep learning methods", Multimedia Tools and Applications, 2023 Publication	<1%
8	digitalcommons.unl.edu Internet Source	<1%
9	arxiv.org Internet Source	<1%
10	www.informatica.si Internet Source	<1%
11	Submitted to University of Portsmouth Student Paper	<1%