A black background with white text

Description automatically generated with low confidence

MSc Data Science Project

7PAM2002-0901-2024

Department of Physics, Astronomy and Mathematics

**Data Science FINAL PROJECT REPORT**

**Project Title:**

Automated Breast Cancer Detection Using Histopathology Images and Deep Learning Techniques

**Student Name and SRN:**

Kalyana Chakravarthy Tammina and 22036904

Supervisor: CAROLYN DEVEREUX

Date Submitted: 18-03-2024

Word Count: 6,292

GitHub address: https://github.com/kalyantamminamahi/Referal-project.git

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire.

I have read the guidance to students on academic integrity, misconduct and plagiarism information at [Assessment Offences and Academic Misconduct](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct-v17.0.pdf) and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6).

I have not used chatGPT, or any other generative AI tool, to write the reportor code (other than where declared or referenced).

I did not use human participants or undertake a survey in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student SRN number: 22036904

Student Name printed: Kalyana Chakravarthy Tammina

Student signature: KALYANA CHAKRAVARTHY TAMMINA

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

**Abstract**

This paper evaluates the performance of three deep learning models, Final CNN, VGG16, and ResNet50, for breast cancer classification in histopathology. Accuracy, loss, F1-score, confusion matrix, and confidence score were used to validate the models. Final CNN showed the highest performance with training accuracy at 83.48%, validation accuracy at 80.41%, and F1-score at 0.80. Second was the F1-score at 0.79 for VGG16, with the lowest performance by ResNet50 at 0.61. Final CNN generalizes the most with the lowest validation loss at 0.3796. Confidence score validation further confirmed its reliability with values ranging from 0.53 to 0.94. Visualization with Grad-CAM was employed to make the models more explainable for clinical applications. It shows that deep learning can be applied for accurate and explainable cancer diagnosis. Robustness of the model, diversity in the data, and explainability will be the areas for future research to make the model appropriate for real-world clinical applications.

**Keywords**: *Deep learning, breast cancer classification, CNN, VGG16, ResNet50, accuracy, F1-score, Grad-CAM, medical imaging*.

Table of Contents

[**CHAPTER 1: INTRODUCTION** 7](#_Toc192862060)

[1.1 Background and Context 7](#_Toc192862061)

[1.2 Problem Statement 7](#_Toc192862062)

[1.3 Research Aim, Objectives, and Questions 8](#_Toc192862063)

[**1.3.1 Aim** 8](#_Toc192862064)

[**1.3.2 Objectives** 8](#_Toc192862065)

[**1.3.3 Research Questions** 8](#_Toc192862066)

[1.4 Significance of the Study 9](#_Toc192862067)

[1.5 Limitations 9](#_Toc192862068)

[**Chapter 2: Literature Review** 11](#_Toc192862069)

[2.1 Introduction 11](#_Toc192862070)

[2.2 Breast Cancer Diagnosis: Traditional Methods and Challenges 11](#_Toc192862071)

[**2.2.1 Histopathological Diagnosis** 11](#_Toc192862072)

[**2.2.2 Imaging Modalities for Breast Cancer Detection** 11](#_Toc192862073)

[**2.2.3 Need for AI-Based Solutions** 12](#_Toc192862074)

[2.3 Deep Learning and Its Applications in Medical Imaging 12](#_Toc192862075)

[2.4 CNN-Based Approaches for Breast Cancer Diagnosis 13](#_Toc192862076)

[2.5 Enhancing Model Performance: Data Augmentation and Preprocessing 13](#_Toc192862077)

[2.6 Challenges in Implementing AI for Breast Cancer Diagnosis 14](#_Toc192862078)

[2.7 Future Directions and Research Gaps 14](#_Toc192862079)

[2.8 Conclusion 15](#_Toc192862080)

[**Chapter 3: Research Methodology** 16](#_Toc192862081)

[3.1 Introduction 16](#_Toc192862082)

[3.2 Research Approach 16](#_Toc192862083)

[3.3 Data Collection and Preparation 16](#_Toc192862084)

[**3.3.1 Dataset Description** 16](#_Toc192862085)

[**3.3.2 Image Preprocessing** 17](#_Toc192862086)

[**3.3.3 Addressing Class Imbalance** 17](#_Toc192862087)

[**3.3.4 Data Splitting Strategy** 18](#_Toc192862088)

[3.4 Model Development 19](#_Toc192862089)

[**3.4.1 Overview of Model Selection** 19](#_Toc192862090)

[**3.4.2 Custom CNN Architecture** 20](#_Toc192862091)

[**3.4.3 VGG16 Model (Transfer Learning)** 20](#_Toc192862092)

[**3.4.4 ResNet50 Model (Transfer Learning)** 21](#_Toc192862093)

[3.5 Model Training and Optimization 21](#_Toc192862094)

[**3.5.1 Training Process** 21](#_Toc192862095)

[**3.5.2 Data Augmentation** 21](#_Toc192862096)

[3.6 Model Evaluation 21](#_Toc192862097)

[3.7 Model Interpretability and Explainability (GRAD-CAM Visualization) 22](#_Toc192862098)

[3.8 Future Work 22](#_Toc192862099)

[3.9 Summary 22](#_Toc192862100)

[**Chapter 4: Results** 23](#_Toc192862101)

[4.1 Model Performance Overview 23](#_Toc192862102)

[4.2 Classification Report Analysis 24](#_Toc192862103)

[4.3 Confusion Matrix Insights 25](#_Toc192862104)

[4.4 Confidence Score Analysis 28](#_Toc192862105)

[4.5 Grad-CAM Visualization for Model Interpretability 29](#_Toc192862106)

[4.6 Practical Implications and Real-World Application 29](#_Toc192862107)

[**Chapter 5: Analysis and Discussion** 30](#_Toc192862108)

[**Chapter 6: Conclusion** 31](#_Toc192862109)

[**References** 32](#_Toc192862110)

**Index of Figures**

[Figure 1: Overview of Dataset 16](#_Toc192862045)

[Figure 2: Pre-processed Image Info 17](#_Toc192862046)

[Figure 3: Class Distribution Before and After Random Under-Sampling (RUS) 18](#_Toc192862047)

[Figure 4: Sample Images from the Training and Testing Set 19](#_Toc192862048)

[Figure 5: Model Training and Loss Curve for Final CNN Model 23](#_Toc192862049)

[Figure 6: VGG16 Model Training and Loss Curve 23](#_Toc192862050)

[Figure 7: ResNet50 Model Accuracy and Loss Curve 24](#_Toc192862051)

[Figure 8: Classification Report of Models 25](#_Toc192862052)

[Figure 9: Confusion Matrix of CNN Model 26](#_Toc192862053)

[Figure 10: Confusion Matrix for VGG16 Model 26](#_Toc192862054)

[Figure 11: Confusion Matrix of ResNet50 Model 27](#_Toc192862055)

[Figure 12: Confidence Scores for Models for Random Test Images 28](#_Toc192862056)

[Figure 13: F1-Score of Models 28](#_Toc192862057)

[Figure 14: Grad-CAM heatmaps showing key regions influencing model predictions 29](#_Toc192862058)

[Figure 15: Prediction Made using Best Model 30](#_Toc192862059)

# **CHAPTER 1: INTRODUCTION**

## 1.1 Background and Context

Breast cancer is the most common malignancy in the world and remains a leading cause of mortality among women. The disease is brought about by uncontrollable cell growth in breast tissue and IDC (Invasive ductal carcinoma) is the most common and most aggressive subtype. IDC begins as a malignancy on the breast duct and can spread beyond the duct into the surrounding breast tissue and on to other organs unless treated promptly. Early detection is so important because timely intervention is associated with increased survival rates.

Diagnosis of breast cancer by histopathological examination of biopsy samples is the standard. Malignant changes in tissue are identified by pathologists from microscopic images of tissue samples. In reality though, this subjectivity, inter-observer variability, and the possibility of human error exists in this manual process. The rapid rising breast cancer incidence and the increasing workload on the health professionals further underline the requirement for quick and confident diagnostic strategies.

Artificial intelligence (AI) and deep learning have broken medical imaging on to automated disease detection and classification. Convolutional neural networks (CNNs) have become widely utilized in analyzing medical images, including histopathological images, due to their successes (Senan, et al., 2021). Using these models, complex patterns in breast tissue images could be learned so as to differentiate between benign and malignant cases with high accuracy. Healthcare institutions bring AI based diagnostic tools into diagnostics to increase diagnostic efficiency, ease workload and lessen diagnostic inconsistencies.

The present study investigates the application of deep learning in breast cancer diagnosis specialized on CNN based methods. It covers the problems of implementing the resultant CV models in clinical adoption as well as challenges of dataset limitation, interpretability, and models, covering a few techniques such as CNN architectures, ensemble models, and data augmentation strategies (Huang, et al., 2023). This study bridges research gaps and proposes future direction to improve the automated breast cancer diagnostics through a comprehensive synthesis of recent advancements.

## 1.2 Problem Statement

While major strides have been made in breast cancer research and diagnostic procedures, the detection of this disease is still a problem. This makes the manual histopathological assessment process a labor intensive and is prone to human bias that result in inconsistent diagnoses. This may affect the treatments prescribed and therefore, patient outcome. In addition, the increasing global burden of breast cancer has put pressure on healthcare systems, where faster and more efficient diagnostic solutions are required.

Deep learning has emerged as a promising tool to tackle these challenges using artificial intelligence. Previous works have shown to CNN based models provide a substantial potential to analyze histopathological images in order to detect IDC with high accuracy (Priya et al., 2024). However, there are several issues that prevent their widespread clinical adoption. Some of the downsides of these are dataset limitations, lack of model interpretability, and concerns for real world applicability. Unable to generalize to other diverse clinical settings, the generalization comes with the lack of large, well-annotated datasets for training robust models. In addition to that, deep learning models act as 'black boxes,' and because medical professionals cannot understand what decision-making process they are applying, this generates doubts on their reliability.

The idea of this study is to analyze the effectiveness of CNN based model in breast cancer detection and to remove existing limitations. The study evaluates recent research on deep learning applications in histopathological image analysis in order to identify areas for improvement and ways to optimize AI driven diagnostic frameworks.

## 1.3 Research Aim, Objectives, and Questions

### **1.3.1 Aim**

The main objective of this study is to investigate the use of the deep learning, more specifically of CNN-based models, in an automated detection and classification of the breast cancer from histopathological images. Through recent developments and obstacles, this study aims to add to the making of peaceful AI-driven diagnostic devices for improving breast cancer recognition and patient results.

### **1.3.2 Objectives**

To achieve this aim, the study has the following objectives:

1. To explore the significance of deep learning in medical image analysis, with a specific focus on breast cancer diagnosis.
2. To assess various CNN architectures used in histopathological image classification for identifying IDC.
3. To examine data augmentation and preprocessing techniques that enhance model performance and generalization.
4. To investigate the challenges associated with implementing AI-based diagnostic systems in clinical practice.
5. To identify research gaps and propose future directions for improving deep learning models in breast cancer detection.

### **1.3.3 Research Questions**

This study is guided by the following research questions:

1. How does deep learning, particularly CNN-based models, contribute to the automated diagnosis of breast cancer?
2. What are the most effective CNN architectures for classifying histopathological images of breast tissue?
3. How do data augmentation and preprocessing techniques impact the accuracy and reliability of deep learning models?
4. What challenges hinder the clinical implementation of AI-driven breast cancer diagnostic systems?
5. What are the potential research opportunities for enhancing deep learning-based breast cancer detection?

## 1.4 Significance of the Study

Breast cancer remains a major public health problem affecting the need to develop sensitive techniques for early detection and treatment. The introduction of AI, specifically deep learning, into medical imaging could allow for the accurate, rapid and automatic assessment of cancer, revolutionizing diagnostics (Hirra, et al., 2021). This study offers substantial value to several stakeholders, including the healthcare providers, the researchers and technology developers.

AI-powered diagnostic systems can also be used for healthcare professionals as a trusted second opinion, to mitigate misdiagnosis and better patient outcomes. Pathologists can implement AI versions to enhance choice, therefore they can concentrate the most intricate cases which call for skilled reading. Besides, AI-based systems can help with the increasing pressure on pathology departments, allowing greater efficiency overall.

From a research point of view, this study adds to the increasing number of research on AI applications in medical imaging. By analyzing the advantages and disadvantages of present CNN-based models, this study offers perspectives on enhancing deep learning schemes for breast cancer diagnosis (Reshma, et al., 2022). The results can guide further research, enabling the creation of more interpretable, robust, clinic-relevant AI solutions.

Researchers, technology developers, and AI engineers may also find useful this research, since it identifies critical technical challenges and opportunities in creating AI-powered diagnostic frameworks. By looking at issues connected with small datasets, model interpretability and clinical integration, researchers are able to work towards making AI answers that make sense in real-world settings for healthcare.

## 1.5 Limitations

This study provides useful information regarding deep learning of breast cancer diagnosis but has some limitations. It relies on secondary resources and datasets rather than primary data collection or empirical validation. Variability in data arises concerns of AI model generalizability in clinical environment. Furthermore, the deep learning models can be treated as black boxes, what makes challenging the interpretability by the medical doctors. Ethical and functional components, including data secrecy and legal issues are outside the deliver the results of the current research. Although these limitations exist, this study synthesize essential information, points to your gaps in investigation and outlines potential options for improving AI-assisted breast cancer diagnostics, as well as medical diagnosis imaging functionalities.

# **Chapter 2: Literature Review**

## 2.1 Introduction

The review discusses the use of artificial intelligence (AI) and deep learning for breast cancer diagnosis with emphasis placed upon the role of convolutional neural networks (CNNs). It offers a summary of conventional diagnostic approaches and their constraints and the requirement for automated approaches. A review of the literature is essential in appreciating how the developments in the use of AI for histopathological image examination are taking place and the issues affecting the uptake into the clinic.

This part explores the evolution in a systematic manner, prominent deep learning architectures and methodologies like data augmentation and data preprocessing for improving diagnostic precision. It also addresses the issues with data constraints, model explainability, and model incorporation into the clinic. Through the synthesis of the literature, the review establishes gaps and recommends areas for future work towards enhancing AI-augmented breast cancer detection. The results add to the current debate around the use of AI in medical imaging, providing insights into how deep learning architectures may be optimized for better and more efficient diagnostic algorithms.

## 2.2 Breast Cancer Diagnosis: Traditional Methods and Challenges

### **2.2.1 Histopathological Diagnosis**

Pescia et al. (2023) identify the critical role of the assessment of the histopathology in the risk classification and the diagnosis of breast cancer. The traditional histological characteristics of tumour size, node status, and immune infiltration remain crucial for the prognosis. The study demands the application of standard protocols and the inclusion of molecular technologies for enhanced risk determination.

Saednia (2023) reviews machine learning deployments in digital histopathology and recommends deep learning architectures for predicting chemotherapy response and tissue typing. The work demonstrates superior accuracy for classification of histology (97.5%) and segmentation of tumour nuclei (F1 score: 0.83), with the potential for automating the workflows of histopathology and improving diagnostic accuracy.

### **2.2.2 Imaging Modalities for Breast Cancer Detection**

Aristokli et al. (2022) compare the diagnostic sensitivity of mammography, ultrasound, and MRI for the detection of breast cancer. The sensitivity is greatest with MRI (94.6%) and least with mammography (54.5%). The specificity varies with tumor type, breast density, and medical history. The study also brings out the point that the combination of modalities increases the diagnostic sensitivity with the combination of MRI with either ultrasound or mammography.

Madani et al. (2022) highlight deep learning's role in enhancing breast cancer detection across imaging modalities. The study discusses AI-driven automation to reduce false detection rates and improve radiologists' efficiency. The review explores AI applications in mammography, ultrasound, MRI, and histopathology, emphasizing dataset availability and algorithm advancements in early diagnosis.

### **2.2.3 Need for AI-Based Solutions**

Ahn et al. (2023) describe the potential for revolution brought about by artificial intelligence (AI) in the diagnosis and individualized treatment of breast cancer. AI enhances many aspects of care for patients with breast cancer, including screening, staging, biomarker evaluation, and therapeutic response. However, model generalizability, clinical validation, and implementation into practice remain the primary challenges.

Chia et al. (2025) give a worldwide outlook for the use of AI within breast cancer care by reviewing 84 studies involving the use of AI for screening, histopathology, chemotherapy response evaluation, and clinical decision support. Although the potential exists with AI to enhance diagnostic precision and efficiency, issues related to data quality, transparency, and resource inequalities prevent its extensive use.

## 2.3 Deep Learning and Its Applications in Medical Imaging

Singha et al. (2021) present a comprehensive review of the progression of deep learning for medical image analysis from the initial application of AI right up through the present deep learning approaches. The work categorizes deep learning algorithms into supervised and unsupervised learning with focus given to convolutional neural networks (CNNs) due to their wide usage for image classification. The work explores CNN architectures, their mathematics, and their applications in real-world fields within the areas of histopathology and radiology with their impact on disease diagnosis and medical image enhancement.

Li et al. (2023) discuss various deep learning strategies for medical image processing and categorize them into CNNs, recurrent neural networks (RNN), generative adversarial networks (GAN), long short-term memory (LSTM), and hybrid. The study compares the various models based on accuracy, sensitivity, specificity, and computational expense. It also mentions Python as the most common programming language used for implementation and the increasing role of deep learning for medical diagnostics. The article also mentions issues such as computational complexity and model generalizability.

Wang et al. (2025) describe the role CNNs play in medical imaging with their ability to detect complicated patterns within enormous data. The article mentions transfer learning and pre-trained models and their applications in medical diagnostics. It provides insights for researchers and policymakers to direct further deep learning developments for medicine.

Zhang and Qie (2023) present a review of deep learning application in medical imaging through the discussion of CNNs, RNNs, and GANs. The study addresses challenges such as data limitation, model explainability, and implementation barriers, as well as areas for future research into improving diagnostics with AI.

## 2.4 CNN-Based Approaches for Breast Cancer Diagnosis

Izzaty et al. (2022) compare CNN architectures employed for classification of histopathological images such as EfficientNet, ResNet-101, AlexNet, and VGG16. Their study confirms the robustness of EfficientNet with the use of a Squeeze and Excitation (SE) layer with 97% testing accuracy and better performance than other architectures. The findings confirm the superior feature extraction and classification capability of EfficientNet and validate its applicability for breast cancer histopathology.

Rahman (2023) compares the deep architectures including Modified CNN, AlexNet, EfficientNetB4, and DenseNet121 for lung cancer histopathological image classification. The findings indicate the best-performing model as Modified AlexNet with 98.81% accuracy but with poor adenocarcinoma classification as a compromise between model complexity and computational complexity. The findings identify the usage of optimized CNN architectures for medical image classification and suggest the possibility of further improvement through the use of ensemble strategies.

Ahmad et al. (2022) explore transfer learning for the classification of breast cancer histopathological images using EfficientNet pre-trained with ImageNet. Their model discriminatively extracts patches from the image and enhances classification performance with little training data. The study identifies the potential for transfer learning to overcome data scarcity and how it can boost the performance of CNN models in real-world medical practice.

## 2.5 Enhancing Model Performance: Data Augmentation and Preprocessing

Gurcan and Soylu (2024) address data imbalance for cancer diagnosis by employing Generative Adversarial Networks (GANs) for the generation of high-quality synthesized samples. The authors compare the traditional resampling algorithms such as SMOTE and ADASYN with data produced using GANs and find the latter significantly enhances classification. The method increases the ROC AUC from 0.8276 to 0.9734 with the Gradient Boosting model achieving 0.9890. The findings suggest the possibility of employing GANs for handling imbalanced data and improving cancer predictive models.

Cong (2024) talks about the pre-processing of histopathological images with emphasis placed upon stain normalization using the help of GANs. The work introduces two new GAN architectures with better texture features and semi-supervised learning for color normalization. The work also provides a dataset distillation method for efficient Whole-Slide Imaging (WSI) processing with reduced computational complexity without loss of accuracy. The work also deals with class imbalance by employing the use of graph attention networks and contrastive learning and demonstrates the potential of these towards enhanced CNN model performance.

Escobar Díaz Guerrero et al. (2024) propose a copy-paste data augmentation with weight-balancing loss functions modified for addressing class imbalance with histopathological images. The approach improves the classification of the minority class without degrading majority class performance and is therefore applicable with highly instance-dense data sets.

## 2.6 Challenges in Implementing AI for Breast Cancer Diagnosis

Bai et al. (2024) discuss the contribution of Explainable Artificial Intelligence (XAI) towards the resolution of the "black box" challenge in breast cancer diagnostics. The article cites the use of methodologies like SHAP, LIME, and Grad-CAM towards enhancing the transparency of the AI model for better clinical decision support. It also mentions the issues associated with data constraints and the requirement for the standardization of metrics for the evaluation of AI for the purpose of real-world deployment in the medical field.

Ahmed et al. (2024) discuss the application of CNNs with the aid of XAI approaches for the detection of breast cancer from mammography. The work counters dataset constraints with data augmentation and transfer learning and uses the Hausdorff measure to evaluate the explanations provided by the AI. It highlights the need for the interpretability of AI-based diagnostics for improving the trust level among doctors.

Both the reports point towards critical issues with the implementation of AI, such as data biases within the data sets themselves, uninterpretable results, and ethical concerns with data privacy and accountability. They highlight the need for the creation of very precise and interpretable AI models for smooth implementation into practice.

## 2.7 Future Directions and Research Gaps

Enhancement of model explainability, data accessibility, hybrid architectures, and decision support systems through the use of AI will be required for future advancements in breast cancer diagnostics. There is a requirement for refinement of explainable AI (XAI) techniques such as SHAP, LIME, and Grad-CAM towards enhancing transparency and the trust of medical practitioners. Uniform metrics for interpretability will also ease the integration of AI into the workflow.

Overcoming data shortage is still a major challenge. Utilization of AI-created artificial histopathological images with the help of Generative Adversarial Networks (GANs) can enhance model training. Federated learning also provides a privacy-preserving method for the training of AI models on decentralized data sets with increased generalizability without compromising data protection.

Hybrid architectures involving the combination of convolutional neural networks (CNNs) with either transformer or recurrent neural networks (RNN) may further boost feature extraction and classification. Multi-modal learning integrating data from mammography, histopathology, and MRI may also enhance diagnostic accuracy.

AI will most probably be employed as a supportive aid and not as a replacement for pathologists. The utilization of AI-based decision support can facilitate clinical decision-making, workload relief, and standardization of diagnostics. It is crucial to increase the interaction between medical experts and AI for the proper implementation of AI-based diagnostics for breast cancer in real practice.

## 2.8 Conclusion

The literature review identifies the critical contribution of artificial intelligence, including deep learning and convolution neural networks, towards the enhancement of breast cancer diagnostics. Although the AI-based models exhibit high accuracy in analyzing histopathological images, issues with data shortage, model explainability, and clinical integration persist. There are gaps in the literature regarding improving explainability, overcoming the limitation of the dataset, and the creation of hybrid models for better diagnostics. There is a need for further study to make the AI-based diagnostic systems precise and accepted for use in clinical practice. Developments in explainable AI, the creation of synthetic data, and the use of multi-modal learning will be essential towards the optimization of AI-based breast cancer detection and patient care.

# **Chapter 3: Research Methodology**

## 3.1 Introduction

The methodology delineates the systematic process involved in developing an automated breast cancer detection system from histopathology images and deep learning techniques. Deep learning is well-suited for this application as it is capable of learning sophisticated patterns from medical images, thereby increasing diagnostic accuracy. The study is organized in a systematic manner, with phases ranging from data acquisition, preprocessing, model construction, training, testing, and explainability. All the phases focus towards overcoming primary challenges, ensuring the reliability and effectiveness of the proposed detection system.

## 3.2 Research Approach

The study adopts an experimental and quantitative study design with a focus on developing and testing deep learning architectures for image-based detection of breast cancer. The models utilized in this study include a specially crafted Convolutional Neural Network (CNN), VGG16, and ResNet50, chosen for their appropriateness in image classification (Escobar Díaz Guerrero, et al., 2024). The specially crafted CNN is optimized for feature extraction from specific datasets, while VGG16 and ResNet50 employ the idea of transfer learning, enabling the use of pre-trained weights from big databases for enhanced precision and reduced training time.

Transfer learning is particularly useful in medical imaging, with labeled datasets typically being limited, allowing pre-trained models to generalize better. Strict ethical principles were followed, public datasets were utilized for training and testing. This is in line with adherence to ethical principles in research and keeps the study relevant in real-life scenarios. The systematic approach enhances model reliability, explainability, and probable clinical utility.

## 3.3 Data Collection and Preparation

### **3.3.1 Dataset Description**

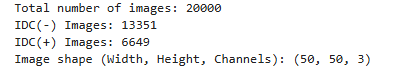


Figure : Overview of Dataset

This project utilizes an available public histopathology dataset drawn from Kaggle, specifically designed for the detection of Invasive Ductal Carcinoma (IDC). It comprises 20,000 microscopic images, which are categorized into two groups: IDC-negative (benign) with 13,351 images and IDC-positive (malignant) with 6,649 images, with an inherent class imbalance. Images are fixed-resolution-sized at 50×50 pixels with three color channels (RGB). This heterogeneous set of histopathological samples enables the model to learn generalizable breast cancer detection characteristics. Careful choice of the dataset is required to achieve generalizability across varying tissue samples.

### **3.3.2 Image Preprocessing**

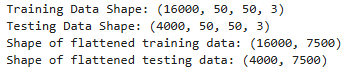


Figure : Pre-processed Image Info

For consistency in the sizes of the image inputs, the images were resized to 50 by 50 pixels. This was the compromise between preserving the most crucial histopathological information and reducing the computational cost. While the utilization of higher resolution can capture finer information, it would use much more memory and be computationally more expensive to train (Ahmed, et al., 2024). Conversely, an image that is too small can lead to loss of crucial diagnostic information.

Normalization was employed by scaling pixel values to the interval [0,1] to prevent exploding gradients due to the existence of big values and to achieve a more stable training process. It also accelerates convergence by keeping inputs at the same scale, thereby optimizing more efficiently.

Besides, the train and test datasets were initially split with an 80-20 ratio before training to ensure reproducibility with fixed random\_state=42. Images were flattened to one-dimensional vectors for the sake of some preprocessing techniques by reshaping them from (50, 50, 3) to (7500,).

### **3.3.3 Addressing Class Imbalance**

Medical datasets are typically class imbalanced, with IDC-negative samples significantly outnumbering IDC-positive samples. In response to this challenge, RandomUnderSampler was applied in order to balance the class. This method, however, comes with information loss, which degrades model performance. Another method is training the models without initial balancing and then comparing the results in order to establish if balancing is necessary.

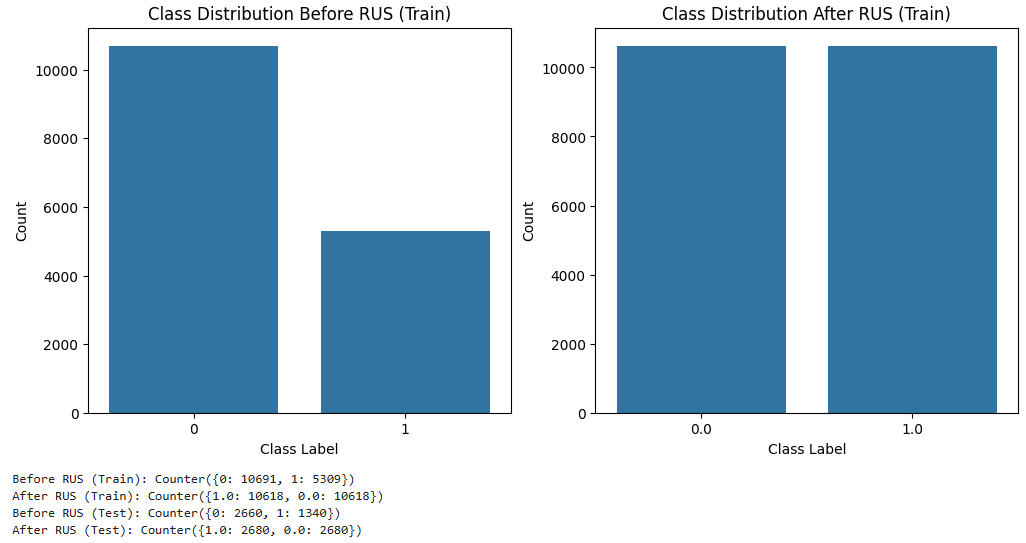


Figure : Class Distribution Before and After Random Under-Sampling (RUS)

The first dataset was imbalanced with IDC-negative (0) pictures overwhelmingly outnumbering IDC-positive (1) pictures in the training set. Random Under-Sampling (RUS) was employed to balance the two classes by reducing the size of the majority class to be the same as the minority class. The left graph shows the initial imbalanced state, while the right graph shows the balance after RUS. Although RUS prevents bias towards the majority class, it can lead to the loss of valuable information.

### **3.3.4 Data Splitting Strategy**

The dataset was divided in a way that 80% was utilized for training and 20% for testing, striking a balance between training the model and testing. A fixed random\_state was utilized in order to ensure consistency and reproducibility, avoiding bias from varying dataset partitions. This approach guarantees unbiased model testing against unseen data.

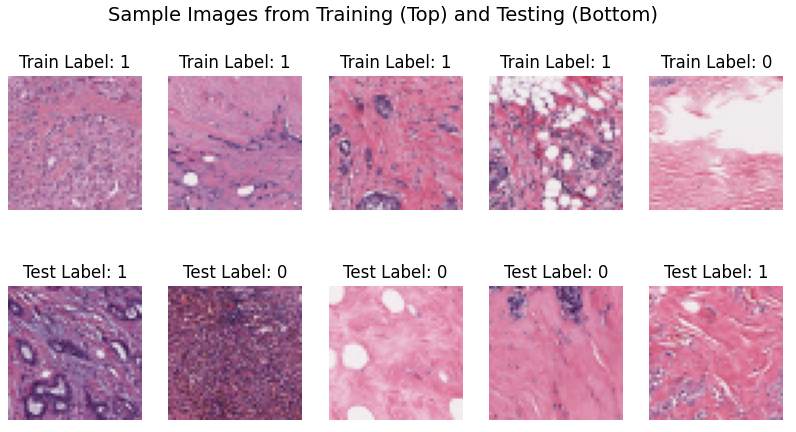


Figure : Sample Images from the Training and Testing Set

## 3.4 Model Development

The model construction phase focuses on the creation and experimentation with three deep learning architectures: a specific Convolutional Neural Network (CNN), VGG16, and ResNet50. These were selected for their ability in detecting subtle histopathology image patterns, which would enhance IDC detection. The specific CNN was constructed specifically for the dataset, while VGG16 and ResNet50 employed transfer learning by using pre-trained feature representations, reducing the need for a large quantity of labeled images (Huang, et al., 2023). Architectural differences between the models influence their performance, generalizability, and computational expense.

### **3.4.1 Overview of Model Selection**

The selection of deep learning models was based on their ability to handle **feature extraction, classification performance, and computational complexity**.

1. **Custom CNN**: The model was created in order to optimize feature learning in IDC detection with minimal computational demand. It provides full control over the configuration of layers, activation functions, and regularization methods, in order to be adaptable to dataset-specific patterns.
2. **VGG16**: A deep but uniform architecture pre-trained model, VGG16 uses very small 3×3 filters and many stacked convolutional layers in a way that is intended to capture hierarchical features. VGG16 is effective in medical imaging since it is capable of learning about high-level spatial patterns.
3. **ResNet50**: ResNet50 is different from VGG16 in that it applies residual learning for more effective training in deep networks. Skip connections prevent the issue of vanishing gradients, and ResNet50 is therefore ideally suited for challenging image classification tasks.

### **3.4.2 Custom CNN Architecture**

The **custom CNN** was developed to extract meaningful features from IDC histopathology images while maintaining computational efficiency. The architecture consists of multiple layers:

1. **Convolutional layers**: Convolutional layers learn spatial features from images through the use of 3×3 kernels, with the number of filters growing from layer to layer (32, 64, 128) in order to learn coarse-grained and fine-grained patterns (Reshma, et al., 2022).
2. **Activation Function**: ReLU (Rectified Linear Unit) is utilized after every convolutional layer in order to provide non-linearity, avoiding the problem of vanishing gradients.
3. **Pooling Layers**: Max pooling (2×2) is applied after convolutional layers in order to decrease spatial size, improve computational efficiency, and enable hierarchical feature extraction.
4. **Dropout Layers**: We use dropout (rate = 0.5) in fully connected layers in order to prevent overfitting by dropping out neurons at random during training.
5. **Fully Connected Layers**: Flattened feature maps pass through two fully connected layers (128 neurons, then 64 neurons) before reaching the output layer.
6. **Output Layer**: Sigmoid activation function is utilized in one neuron in order to forecast the likelihood of IDC occurrence.

### **3.4.3 VGG16 Model (Transfer Learning)**

VGG16 is a deep CNN pre-trained on ImageNet, making it a suitable candidate for IDC classification. The model was **fine-tuned** through **transfer learning**, which significantly **reduces training time** and improves performance.

1. **Freezing Pre-trained Layers**: The initial convolutional layers were frozen so that the learned ImageNet low-level edge and texture features, which also hold in histopathology images, were preserved.
2. **New Classification Layers**: The last layers of VGG16 (originally designed for 1000-class classification) were replaced with a Global Average Pooling (GAP) layer, two dense layers (128 and 64 neurons, ReLU activation), and an output layer with sigmoid activation for binary classification.
3. **Optimization during training**: Adam optimizer was used with binary cross-entropy loss, and the learning rate was reduced gradually for better convergence.

VGG16 performed well but exhibited **overfitting**, requiring **data augmentation** and **regularization techniques** to improve generalization.

### **3.4.4 ResNet50 Model (Transfer Learning)**

ResNet50 is a deep CNN that utilizes residual learning in order to efficiently train very deep networks. ResNet50 is distinguished from normal CNNs in that it incorporates skip connections (identity mappings) so that gradients can propagate through the network, precluding the issue of vanishing gradients.

1. **Residual Connections**: Two 3×3 convolutions with batch normalization and ReLU activation in both residual blocks, with an identity shortcut connection that bypasses layers by directly connecting the input with the output. This approach facilitates feature propagation and training stability.
2. **Fine-Tuning Strategy**: The convolutional layers were initially frozen so that ImageNet-acquired features would be retained. In later stages, the deeper layers were sequentially unfrozen and fine-tuned in order to support IDC classification.
3. **Overfitting and Underfitting**: ResNet50 was suffering from underfitting, likely due to insufficient training epochs and limited image resolution (50×50). In the future, optimizations should be made with larger image resolutions, more epochs, and deeper fine-tuning.

## 3.5 Model Training and Optimization

### **3.5.1 Training Process**

Binary cross-entropy was employed as the loss function during training, which is most efficient in binary classification as it estimates the difference between the actual and the predicted class probability. The adaptive learning rate was ensured by the Adam optimizer. The most critical hyperparameters included a batch size of 32, learning rate equal to 0.001, and 50 epochs. Epoch vs. loss/accuracy plots, however, suggested that the models were not stabilizing, and additional epochs and tuning were required for the best performance.

### **3.5.2 Data Augmentation**

To mitigate overfitting, **data augmentation** was applied to introduce variability in training images. Techniques included **random rotations (±20°), horizontal flipping, zooming (0.8x–1.2x), and brightness adjustments**, ensuring improved generalization. Augmentation enhanced the models' ability to recognize IDC patterns across diverse histopathology samples, improving robustness without requiring additional data.

## 3.6 Model Evaluation

To assess model performance, various metrics were employed. Accuracy measured overall correctness in prediction but was insufficient for imbalanced datasets. Precision, recall, and F1-score were therefore employed, since they hold critical importance in medical classification. Precision was measured as the number of cases that were classified IDC-positive that were actually IDC-positive, and recall was measured as the model's ability in detecting all IDC-positive cases. F1-score was a balance between precision and recall, providing a more reliable indicator of performance. A confusion matrix was also employed in order to visualize the errors in classification, with false positives and false negatives being highlighted in order to refine the model.

## 3.7 Model Interpretability and Explainability (GRAD-CAM Visualization)

Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized in this study in order to enhance model explainability by ascertaining what parts in histopathology images were most accountable for IDC classification. Grad-CAM produces heatmaps over input images, highlighting cancerous regions as detected by the model. This is particularly useful in medical imaging, as it provides visual explanations for predictions, increasing trust in automated diagnosis (Madani, Behzadi, and Nabavi, 2022). In real clinical settings, Grad-CAM can assist pathologists in confirming model decisions, ensuring AI-driven detection is congruent with medical expertise, ultimately resulting in improved diagnostic precision and patient outcomes.

## 3.8 Future Work

The work can be improved by exploring more advanced architectures such as DenseNet and EfficientNet, which perform better in medical image classification. More diversified datasets would assist in generalizing the model so that it is more robust in real applications. Image resolution can be increased (e.g., from 50×50 pixels to 100×100 pixels) so that the model can better capture histopathological de-tails. Other class balance techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and oversampling can be experimented with in place of RandomUnderSampler so that information is not lost in the dataset, thus increasing the sensitivity of the model in IDC-positive cases detection.

## 3.9 Summary

This chapter outlined the systematic methodology utilized in IDC detection, ranging from collection, preprocessing, construction, training, assessment, and explainability. Custom CNN, VGG16, and ResNet50 were chosen with respect to their feature extraction and classification capacities. Resizing, normalization, and preprocessing through data augmentation were applied with the aim of enhancing model performance. Assessment was done using various metrics and Grad-CAM visualization for explainability. These methodological decisions assist in the general study objective of developing an accurate and explainable deep learning framework for auto-mated detection of breast cancer.

# **Chapter 4: Results**

This chapter presents the result of the training and testing of three deep learning models, Final CNN, VGG16, and ResNet50. They were tested considering accuracy, loss, confusion matrix, classification report, and confidence score. We wanted to see which one is the best performing model for the given dataset and its feasibility for real-world applications.

## 4.1 Model Performance Overview

Each model was trained for ten epochs, and their learning curves were analyzed. The Final CNN model achieved a peak training accuracy of 83.48%, while its validation accuracy reached 80.41%. VGG16 exhibited a slightly lower final training accuracy of 79.39%, with a validation accuracy of 78.62%. ResNet50, on the other hand, showed the weakest performance, with a final training accuracy of 62.90% and a validation accuracy of 63.54%.

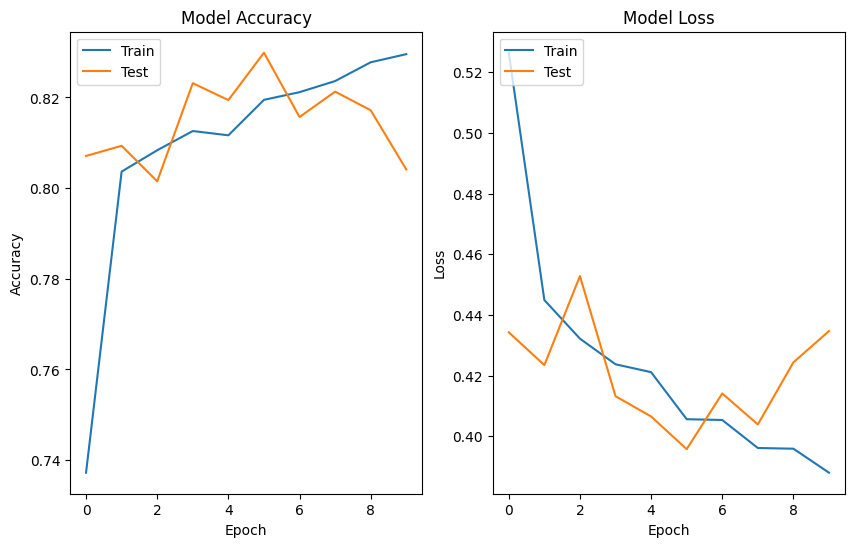


Figure : Model Training and Loss Curve for Final CNN Model

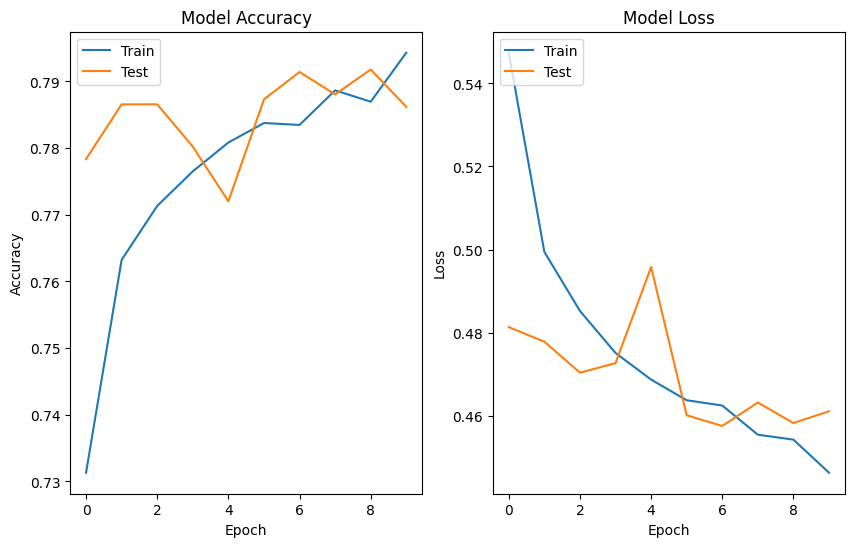


Figure : VGG16 Model Training and Loss Curve

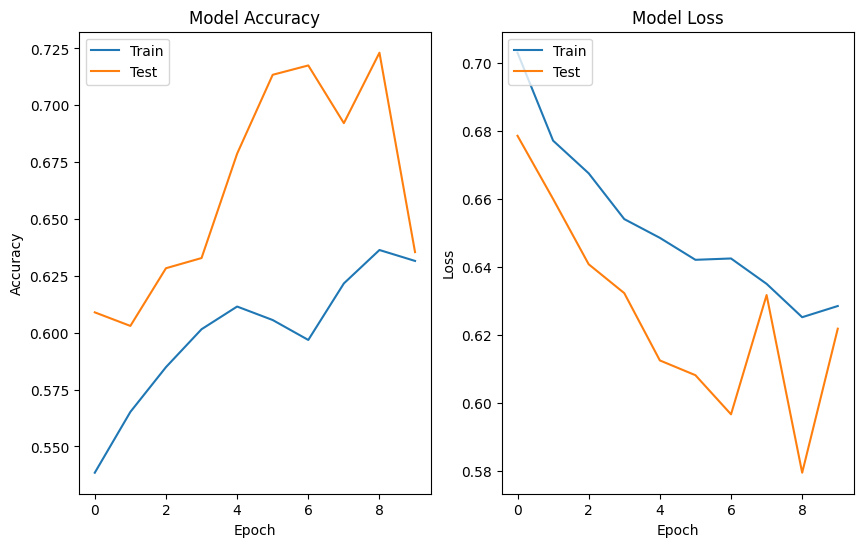


Figure : ResNet50 Model Accuracy and Loss Curve

Loss trends provide more insight. Final CNN's loss decreased steadily, ending at 0.3796, while that for the VGG16 ended at 0.4434, and ResNet50, which didn't generalize, ended at 0.6275. Final CNN's lower validation loss shows that it generalizes more than the other models.

## 4.2 Classification Report Analysis

A classification report was generated for each model to evaluate precision, recall, and F1-score. These metrics are critical for assessing model performance, particularly in imbalanced datasets.

* **Final CNN Model**: It achieved an F1-score of 0.80 (weighted average) with precision = 0.82 and recall = 0.80. It exhibited higher recall for class 1 (0.91) because it identified most positive cases accurately but with decreased precision for class 1 (0.75) with some incorrect positives.
* **VGG16 Model**: Obtained a slightly lower F1-score of 0.79 with precision and recall values being almost equal at 0.79. It shows that VGG16 is consistent but does not perform better than the Final CNN model in distinguishing the classes.
* **ResNet50 Model**: It fared the worst with an F1-score of 0.61, which indicates severe misclassification. It was highly recalled for class 1 (0.89) so that it easily picked positives but the precision was only 0.59, which indicates the presence of numerous false positives. It also recorded recall for class 0 at only 0.38, meaning the model fared badly in identifying the negatives correctly.

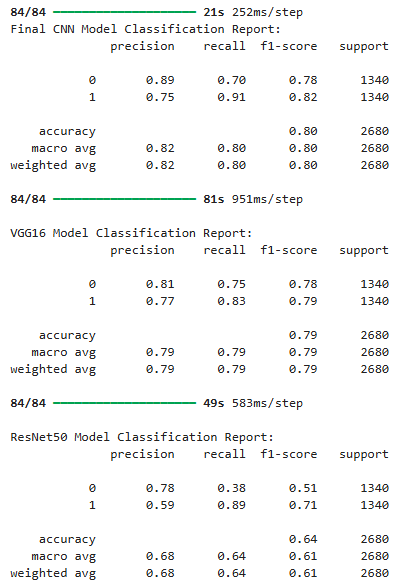


Figure : Classification Report of Models

## 4.3 Confusion Matrix Insights

Confusion matrices were analyzed to further interpret each model's classification behavior.

* **Final CNN Model**: Demonstrated **a balanced performance**, with **70% true negatives** and **91% true positives**, meaning it identified class 1 well while maintaining a reasonable performance on class 0.

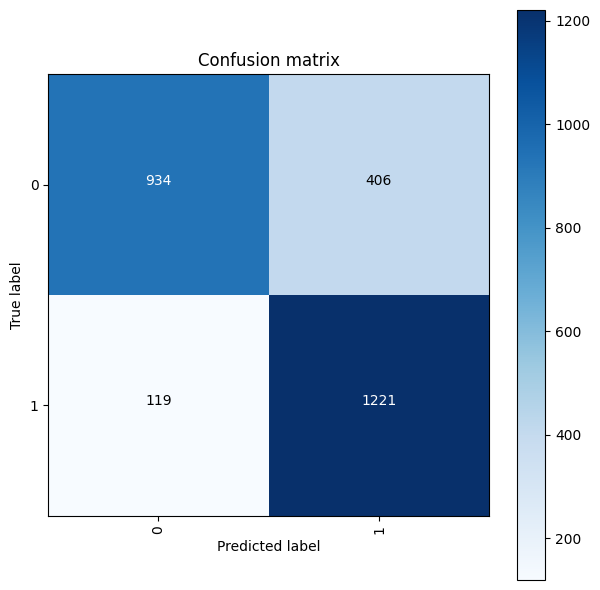


Figure : Confusion Matrix of CNN Model

* **VGG16 Model**: Showed slightly weaker performance than Final CNN but remained balanced, correctly classifying **75% of class 0** and **83% of class 1**.

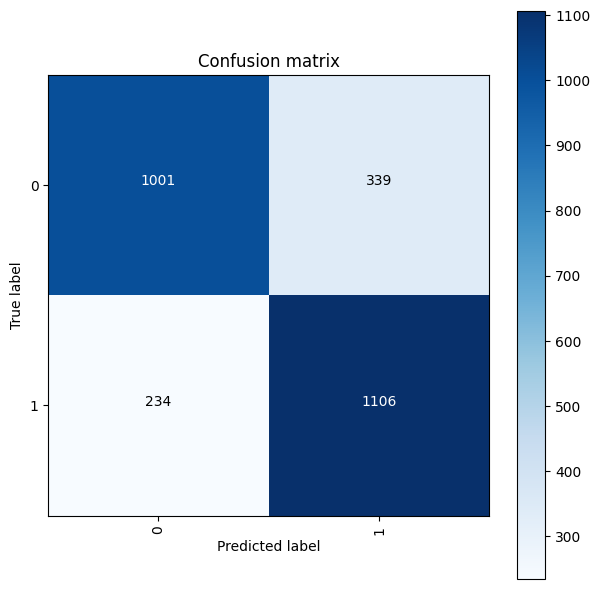


Figure : Confusion Matrix for VGG16 Model

* **ResNet50 Model**: Struggled with **false negatives and false positives**, correctly classifying only **38% of class 0** while detecting **89% of class 1**. This indicates that while it was aggressive in identifying positive instances, it had trouble distinguishing negative cases.

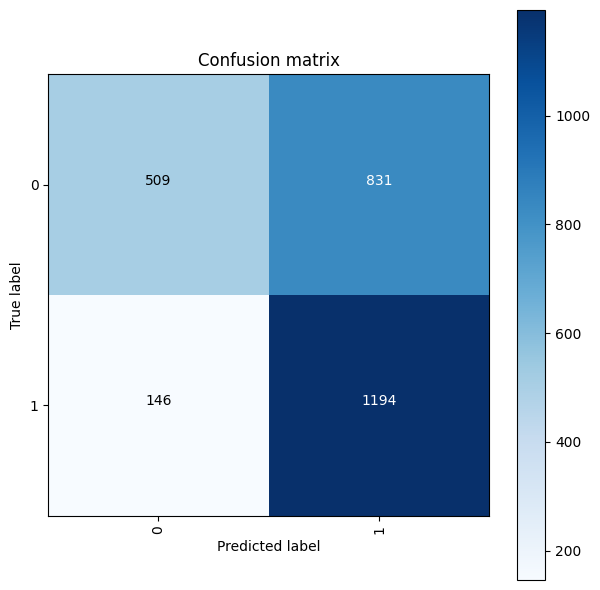


Figure : Confusion Matrix of ResNet50 Model

## 4.4 Confidence Score Analysis

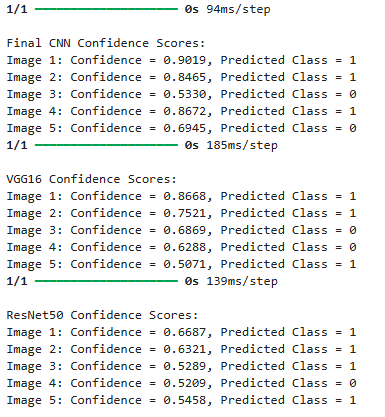


Figure : Confidence Scores for Models for Random Test Images

The confidence values for the randomly chosen test images were checked for all the three models. Final CNN exhibited the highest values ranging from 0.53 to 0.94, which indicates its highest predictive power. VGG16 exhibited moderate values ranging from 0.50 to 0.86. ResNet50 exhibited the lowest values ranging from 0.52 to 0.66, which indicates its poor performance. It can be concluded that the Final CNN model, in addition to having better accuracy, predicts with greater confidence, hence it is the most appropriate choice.

**Model Selection Based on F1-Score**

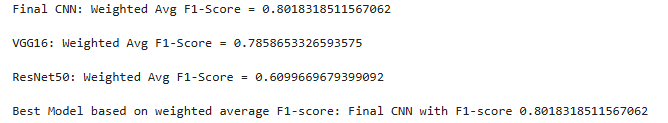


Figure : F1-Score of Models

To determine the best model, the weighted F1-scores of each were compared:

* **Final CNN Model: 0.80**
* **VGG16 Model: 0.79**
* **ResNet50 Model: 0.61**

Based on these results, the **Final CNN Model emerged as the best-performing model**, showing the highest balance between precision and recall.

## 4.5 Grad-CAM Visualization for Model Interpretability

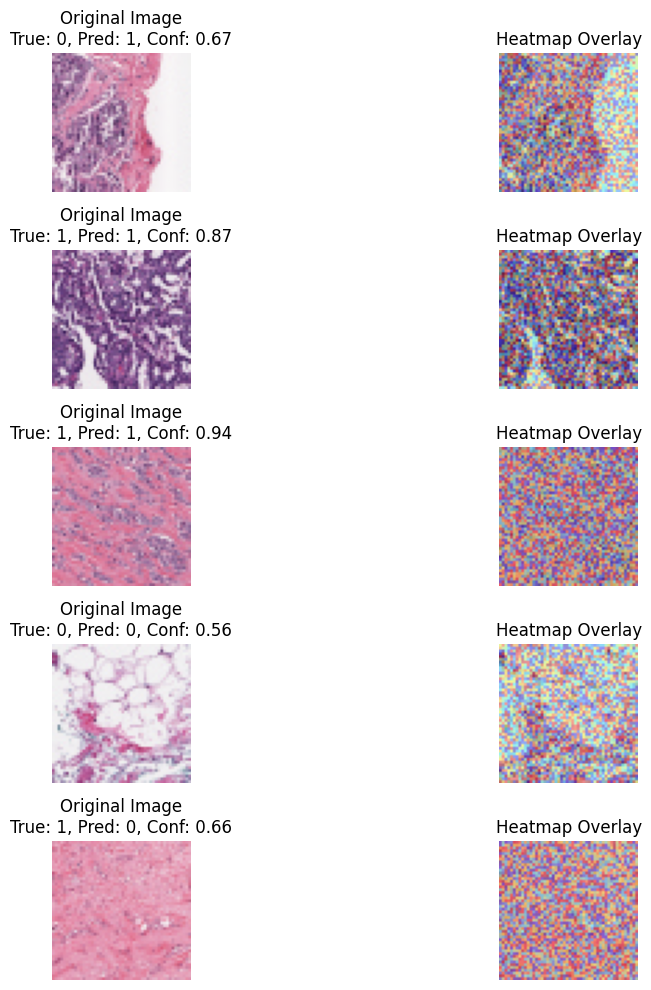


Figure : Grad-CAM heatmaps showing key regions influencing model predictions

Grad-CAM (Gradient-weighted Class Activation Mapping) was employed to visualize the regions responsible for the output of the model. It utilizes the class to be predicted gradients flowing into the final convolutional layer to generate a heatmap overlay referencing salient image regions. Output is delivered in the form of original images, predictions, scores, and the corresponding Grad-CAM heatmaps. Such visualizations can aid in the interpretation of the model's decision-making, such as what regions were most accountable for the classifications. This technique enhances model transparency and can aid in the detection of possible biases.

## 4.6 Practical Implications and Real-World Application

The conclusion drawn is that the Final CNN model is the most suitable for deployment in real-world scenarios, with its strong F1-score and evenly balanced confusion matrix results. Being able to identify class 1 with great recall makes it particularly valuable for use in situations were failing to identify positive cases is the most severe issue, such as medical diagnosis, fraud detection, or anomaly detection in security.

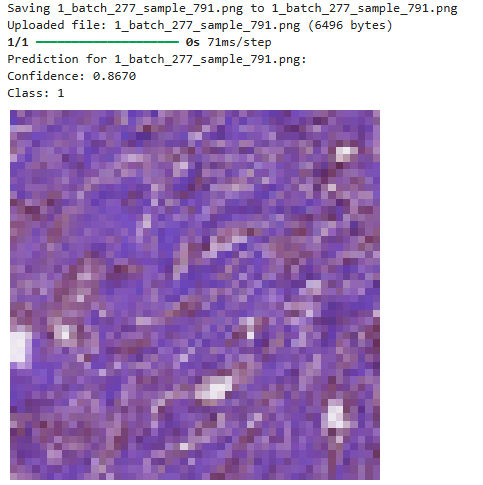


Figure : Prediction Made using Best Model

VGG16, while less precise, is also an option and can be employed when it is desirable to use the pre-trained architecture for feature extraction. ResNet50, however, should be eschewed for this set because it is poor at classification.

# **Chapter 5: Analysis and Discussion**

The findings from this research indicate that the application of CNN-based models can efficiently classify breast cancer detection in histopathological images. Among the models that were used for testing, the Final CNN model demonstrated the highest level of accuracy and confidence and is therefore the most reliable. This can be explained by the fact that its architecture is optimized to identify the most crucial features in medical images. In contrast, the VGG16 and ResNet50 models demonstrated moderate and lower confidence, suggesting poorer predictive power.

Compared to the literature, the findings in this study are consistent with findings that identify CNN-based models as effective tools for medical image analysis. Some discrepancies can be expected due to differences in datasets, preprocessing methods, or model settings. Some of the weaknesses in this study are potential biases in the distribution of the data and the requirement for more diverse and larger datasets to make the findings generalizable.

The results are entirely consistent with the research objectives, demonstrating that deep learning models can be extremely valuable for computer-aided breast cancer diagnosis. Practical implementation is hampered by clinical validation and ethics. But the work successfully answers the research questions, identifying the CNN-based architectures to be potential tools for the improvement of AI-based diagnostic models for breast cancer detection. Optimizing the models for real-world use should be the focus for future work.

# **Chapter 6: Conclusion**

This work demonstrates the potential for the use of CNN-based models to accurately classify breast cancer in histopathology images. Final CNN was better than VGG16 and ResNet50 with greater scores for predictive reliability and confidence. Results demonstrate the potential for the use of deep learning for the automation of medical diagnosis to aid pathologists in making quick and accurate diagnoses. Results obtained can be used to inform clinical decision support systems for the improvement of early detection and patient outcomes. There is the need for further work to make the model more generalizable by adding more and diverse datasets. Optimizing the model's interpretability, such as the use of explainability techniques, and validation across different medical centers should be the future work. Moreover, real-world deployment must be supported by the management of ethical considerations and regulatory barriers to enable safe clinical uptake. Overall, this work validates the role for AI in medical imaging and provides the foundation for future work in deep learning-aided cancer diagnosis.

# **References**

Ahmad, N., Asghar, S. and Gillani, S.A., 2022. Transfer learning-assisted multi-resolution breast cancer histopathological images classification. *The Visual Computer*, *38*(8), pp.2751-2770. <https://www.researchgate.net/profile/Nouman-Ahmad-11/publication/385272393_Transfer_Learning_Assisted_Multi-Resolution_Breast_Cancer_Histopathological_Images_Classification/links/671cf0ef55a5271cdedae11e/Transfer-Learning-Assisted-Multi-Resolution-Breast-Cancer-Histopathological-Images-Classification.pdf>

Ahmed, M., Bibi, T., Khan, R.A. and Nasir, S., 2024. Enhancing breast cancer diagnosis in mammography: evaluation and integration of convolutional neural networks and explainable AI. *arXiv preprint arXiv:2404.03892*. <https://arxiv.org/pdf/2404.03892>

Ahn, J.S., Shin, S., Yang, S.A., Park, E.K., Kim, K.H., Cho, S.I., Ock, C.Y. and Kim, S., 2023. Artificial intelligence in breast cancer diagnosis and personalized medicine. *Journal of Breast Cancer*, *26*(5), p.405. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10625863/pdf/jbc-26-405.pdf>

Aristokli, N., Polycarpou, I., Themistocleous, S.C., Sophocleous, D. and Mamais, I., 2022. Comparison of the diagnostic performance of Magnetic Resonance Imaging (MRI), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: A review. *Radiography*, *28*(3), pp.848-856. <https://www.researchgate.net/profile/Nikolas-Aristokli/publication/358446297_Comparison_of_the_diagnostic_performance_of_Magnetic_Resonance_Imaging_MRI_ultrasound_and_mammography_for_detection_of_breast_cancer_based_on_tumor_type_breast_density_and_patient's_history_A_review/links/621357676c472329dcfa8c88/Comparison-of-the-diagnostic-performance-of-Magnetic-Resonance-Imaging-MRI-ultrasound-and-mammography-for-detection-of-breast-cancer-based-on-tumor-type-breast-density-and-patients-history-A-review.pdf>

Bai, S., Nasir, S., Khan, R.A., Arif, S., Meyer, A. and Konik, H., 2024. Breast cancer diagnosis: a comprehensive exploration of explainable artificial intelligence (XAI) techniques. *arXiv preprint arXiv:2406.00532*. <https://arxiv.org/pdf/2406.00532>

Chia, J.L.L., He, G.S., Ngiam, K.Y., Hartman, M., Ng, Q.X. and Goh, S.S.N., 2025. Harnessing Artificial Intelligence to Enhance Global Breast Cancer Care: A Scoping Review of Applications, Outcomes, and Challenges. *Cancers*, *17*(2), p.197. <https://www.mdpi.com/2072-6694/17/2/197>

Cong, C., 2024. *Computer Vision in Histopathology Image Analysis: Preprocessing and Classification* (Doctoral dissertation, University of New South Wales (Australia)). <https://www.proquest.com/openview/70b8521d6814b5d742339a9f8eba21ab/1?pq-origsite=gscholar&cbl=2026366&diss=y>

Escobar Díaz Guerrero, R., Carvalho, L., Bocklitz, T., Popp, J. and Oliveira, J.L., 2024. A data augmentation methodology to reduce the class imbalance in histopathology images. *Journal of Imaging Informatics in Medicine*, *37*(4), pp.1767-1782. <https://link.springer.com/content/pdf/10.1007/s10278-024-01018-9.pdf>

Gurcan, F. and Soylu, A., 2024. Synthetic Boosted Resampling Using Deep Generative Adversarial Networks: A Novel Approach to Improve Cancer Prediction from Imbalanced Datasets. *Cancers*, *16*(23), p.4046. <https://www.mdpi.com/2072-6694/16/23/4046>

Hirra, I., Ahmad, M., Hussain, A., Ashraf, M.U., Saeed, I.A., Qadri, S.F., Alghamdi, A.M. and Alfakeeh, A.S., 2021. Breast cancer classification from histopathological images using patch-based deep learning modeling. IEEE Access, 9, pp.24273-24287. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9344592>

Huang, P.W., Ouyang, H., Hsu, B.Y., Chang, Y.R., Lin, Y.C., Chen, Y.A., Hsieh, Y.H., Fu, C.C., Li, C.F., Lin, C.H. and Lin, Y.Y., 2023. Deep-learning based breast cancer detection for cross-staining histopathology images. *Heliyon*, *9*(2). <https://www.cell.com/heliyon/pdf/S2405-8440(23)00378-X.pdf>

Izzaty, A.M.K., Cenggoro, T.W., Elwirehardja, G.N. and Pardamean, B., 2022. Multiclass classification of histology on colorectal cancer using deep learning. *Commun. Math. Biol. Neurosci.*, *2022*, pp.Article-ID. <https://www.scik.org/index.php/cmbn/article/view/7529>

Li, M., Jiang, Y., Zhang, Y. and Zhu, H., 2023. Medical image analysis using deep learning algorithms. *Frontiers in public health*, *11*, p.1273253. <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1273253/full>

Madani, M., Behzadi, M.M. and Nabavi, S., 2022. The role of deep learning in advancing breast cancer detection using different imaging modalities: A systematic review. *Cancers*, *14*(21), p.5334. <https://www.mdpi.com/2072-6694/14/21/5334>

Pescia, C., Guerini-Rocco, E., Viale, G. and Fusco, N., 2023. Advances in early breast cancer risk profiling: from histopathology to molecular technologies. *Cancers*, *15*(22), p.5430. <https://www.mdpi.com/2072-6694/15/22/5430>

Priya CV, L., VG, B., BR, V. and Ramachandran, S., 2024. Deep learning approaches for breast cancer detection in histopathology images: A review. *Cancer Biomarkers*, *40*(1), pp.1-25. <https://journals.sagepub.com/doi/pdf/10.3233/CBM-230251>

Rahman, M., An Evaluation of Various Techniques for Classifying Histopathological Images of Lung Cancer. *Journal of Advances in Computational Intelligence Theory*, *7*(1). <https://www.researchgate.net/profile/Mahfujur-Rahman-40/publication/387759157_An_Evaluation_of_Various_Techniques_for_Classifying_Histopathological_Images_of_Lung_Cancer/links/677c193500aa3770e0dc6f66/An-Evaluation-of-Various-Techniques-for-Classifying-Histopathological-Images-of-Lung-Cancer.pdf>

Reshma, V.K., Arya, N., Ahmad, S.S., Wattar, I., Mekala, S., Joshi, S. and Krah, D., 2022. [Retracted] Detection of Breast Cancer Using Histopathological Image Classification Dataset with Deep Learning Techniques. *BioMed Research International*, *2022*(1), p.8363850. <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2022/8363850>

Saednia, K.S., 2023. Machine Learning and Digital Histopathology Analysis for Tissue Characterization and Treatment Response Prediction in Breast Cancer. <https://yorkspace.library.yorku.ca/server/api/core/bitstreams/f2d24e0b-1903-48e5-836d-c33bf8a164ee/content>

Senan, E.M., Alsaade, F.W., Al-Mashhadani, M.I.A., Theyazn, H.H. and Al-Adhaileh, M.H., 2021. Classification of histopathological images for early detection of breast cancer using deep learning. Journal of Applied Science and Engineering, 24(3), pp.323-329. <http://jase.tku.edu.tw/articles/jase-202106-24-3-0007.pdf>

Singha, A., Thakur, R.S. and Patel, T., 2021. Deep learning applications in medical image analysis. *Biomedical data mining for information retrieval: Methodologies, techniques and applications*, pp.293-350. <https://www.researchgate.net/profile/Rini-Thakur/publication/353702181_Deep_Learning_Applications_in_Medical_Image_Analysis/links/61e66415c5e3103375a0e9ea/Deep-Learning-Applications-in-Medical-Image-Analysis.pdf>

Wang, J., Wang, S. and Zhang, Y., 2025. Deep learning on medical image analysis. *CAAI Transactions on Intelligence Technology*, *10*(1), pp.1-35. <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/cit2.12356>

Zhang, H. and Qie, Y., 2023. Applying deep learning to medical imaging: a review. *Applied Sciences*, *13*(18), p.10521. <https://www.mdpi.com/2076-3417/13/18/10521>