IDENTIFICATION OF BREAST CANCER FROM HISTOPATHOLOGICAL IMAGES USING IMAGE CLASSIFICATION

GROUP MEMBERS

1. Andrew Mutuku
2. Joseph Karumba
3. Amina Saidi
4. Winnie Osolo
5. Wambui Githinji
6. Margaret Njenga

Contents

LIST OF ABBREVIATIONS

CNN: Convolutional Neural Network

LMICs: Low- and Middle-Income Countries

ResNet: Residual Network

IDC: Invasive Ductal Carcinoma

DEFINITION OF TERMS USED

Breast Cancer: A disease where malignant cells form in the tissues of the breast.

Histopathological Images: Microscopic images of tissue samples used for diagnosing diseases.

Image Classification: A computer vision task that involves categorizing images into predefined classes.

Precision: (how many of the + cases were actually positive) Measures how many of the predicted positive cases were actually positive.

Recall: (how many of the true positives were correctly identified) Measures how many of the actual positive cases were correctly identified.

F1 Score: The harmonic mean of precision and recall, providing a single metric that balances both. It is useful when you need a balance between precision and recall, especially in cases of imbalanced class distributions.

Invasive Ductal Carcinoma: A common type of breast cancer that begins in the cells lining the milk ducts and then spreads to the surrounding breast tissue. It can invade nearby tissues and may spread to other parts of the body if not treated.

Magnetic Resonance Imaging: A medical imaging technique that uses strong magnets and radio waves to create detailed images of the inside of the body.

Immunohistochemistry: A laboratory technique used to detect specific proteins in tissue samples using antibodies. It helps identify and analyze the presence and location of these proteins, which can aid in diagnosing diseases and understanding tissue characteristics.

Surgical Open Biopsy: A procedure where a surgeon removes a sample of tissue from a specific area of the body to examine it for signs of disease. The sample is analyzed under a microscope to help diagnose conditions like cancer.

ABSTRACT

Background

Breast cancer is a leading cause of cancer-related deaths worldwide, with significant impacts in low- and middle-income countries like Kenya due to late diagnosis and limited healthcare resources. Traditional diagnostic methods are labor-intensive and prone to human error. This project explores the use of machine learning models to automate and enhance the diagnostic process through image classification of histopathological images.

Objective

The primary goal of this study is to develop a machine learning model that can accurately distinguish between benign and malignant breast tumor images from histopathological slides. Ultimately, the aim is to create an affordable and accessible diagnostic tool that enhances early detection and treatment of breast cancer, particularly in resource-limited settings.

Method

We utilized the Breast Cancer Histopathological Image Classification (BreakHis) database containing 9,109 images of breast tumor tissue. Data preprocessing involved normalization and augmentation to ensure uniformity. We tested several models, including Convolutional Neural Networks (CNNs) and ResNet models to determine the most effective approach for image classification.

Results

The ResNet model achieved the highest accuracy, demonstrating superior performance in distinguishing between benign and malignant images. Precision and recall metrics further validated the model's effectiveness by reducing false positives and negatives, resulting in an improved F1 score.

Conclusion

Machine learning models significantly enhance the efficiency of breast cancer diagnosis from histopathological images. Our best model predicts individual images in about five seconds, which reduces diagnostic time and enables quicker, more effective treatment. These advancements promise substantial improvements in healthcare delivery in Kenya and similar regions by providing timely and accurate diagnostics.

INTRODUCTION

1.0 Background Information

Cancer remains a significant global health issue, causing more deaths than HIV, tuberculosis, and malaria combined. LMICs like Kenya bear 70% of the global cancer burden. Early detection is crucial, as approximately 30% of cancers are curable with early identification, another 30% can be treated for prolonged survival, and the remaining 30% benefit from effective symptom management and palliative care (1).

In 2020, there were approximately 19 million new cancer cases and 10 million deaths worldwide. In Kenya, cancer ranked as the third leading cause of death, contributing to 7% of total mortality with 42,116 new cases and 27,092 deaths reported. Breast cancer, the most common cancer globally with over 2.2 million cases in 2020, is the leading cancer type in Kenya. The country reported 6,799 new cases of breast cancer and an age-standardized rate of 41 per 100,000 people. Data from the Kenya National Cancer Registry (2014-2019) show that 70% of breast cancer cases are diagnosed at advanced stages (stage III and IV), often at a younger age (35-50 years) compared to Western countries (50-55 years). Invasive ductal carcinoma (IDC) is the most common type, representing up to 75% of cases. Major challenges include limited access to preventive, diagnostic, and treatment services (2).

Risk factors for breast cancer include age (most cases occur in women over 50), physical inactivity, genetic mutations, alcohol consumption, obesity, and dense breast tissue. Symptoms to monitor include a new lump, breast thickening, skin changes, nipple pain, and abnormal discharge.

Early detection through screening is essential for effective treatment. Screening methods include mammograms for women aged 40 to 74, breast magnetic resonance imaging (MRI) for high-risk individuals, clinical breast exams, and breast self-exams.

Confirmation of breast cancer is typically done through a biopsy, which involves obtaining a tissue sample for microscopic examination. The process includes fixation in formalin, embedding in paraffin wax, breast tissue sectioning into thin slices, and staining with dyes like Hematoxylin and Eosin. Pathologists analyze these stained slides to identify cancerous cells, determine the cancer type, grade, and stage, and may use additional tests like immunohistochemistry (IHC) for further details (3).

The pathologist’s report, which includes details about the cancer's presence, type, and other features, is sent to the treating physician. The physician then creates a treatment plan based on this information. Follow-up care includes regular monitoring and additional tests to evaluate how well the treatment is working and to check for any recurrence. Treatment options may include surgery, chemotherapy, radiation therapy, hormone therapy, or targeted therapy (3).

Machine learning models can significantly enhance diagnostic accuracy by detecting subtle patterns in tissue images and improve efficiency by reducing image analysis time to about five seconds per image. This leads to quicker, more accurate diagnoses and faster treatment planning, promising substantial improvements in healthcare delivery.

BUSINESS UNDERSTANDING

2.0 Introduction

Breast cancer is the most common cancer in Kenya, often diagnosed at advanced stages due to limited access to timely diagnosis and treatment. Using machine learning in diagnosis can help by identifying patterns in histopathological images that might be missed by traditional methods. It can also speed up the analysis process, leading to quicker and more accurate diagnoses.

Machine learning can ease the workload on medical professionals and improve treatment effectiveness. This technology has the potential to transform healthcare in Kenya by enabling earlier detection and treatment, resulting in better patient outcomes and more efficient use of resources.

2.1 Problem Statement and Justification

The primary challenge is to develop a machine learning model that can accurately identify breast cancer from histopathological images. This will enhance diagnostic accuracy, reduce the workload on medical professionals, and improve early detection and treatment outcomes for breast cancer patients in Kenya and other LMICs.

2.2 Research Questions

1. How accurately can machine learning models classify benign and malignant breast tumor histopathological images?
2. Which machine learning algorithms yield the highest accuracy in classifying breast cancer images?
3. Which machine learning algorithms yield the highest precision and recall in classifying breast cancer images?
4. What are the key challenges in using machine learning for breast cancer histopathological image classification?
5. How can machine learning improve the diagnostic process for breast cancer in LMICs?

2.3 Study Objectives

Primary Objectives

1. Develop a robust image classification model to distinguish between benign and malignant breast tumor histopathological images.
2. Enhance diagnostic accuracy through automated analysis.

Secondary Objectives

1. Evaluate the performance of various machine learning algorithms.
2. Improve early detection rates by providing a reliable tool for regular screening.
3. Develop a cost-effective diagnostic tool for resource-limited settings.

DATA UNDERSTANDING

3.1 Data source

We sourced our data from the Breast Cancer Histopathological Image Classification (BreakHis) database, a key resource for medical image analysis focused on breast cancer detection and classification. Developed in collaboration with the P&D Laboratory in Parana, Brazil, this dataset is invaluable for benchmarking and evaluating breast cancer classification models.

You can directly access the Breast Cancer Histopathological Image Classification dataset from the following link: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/

3.2 Data description

The Breast Cancer Histopathological Image Classification dataset comprises 9,109 microscopic images of breast tumor tissue from 82 patients, captured at 40X, 100X, 200X, and 400X magnifications. It includes 2,480 benign and 5,429 malignant samples, each with a resolution of 700x460 pixels in 3-channel RGB PNG format.

The dataset is categorized into benign and malignant tumors. Benign tumors are non-cancerous and typically remain localized, while malignant tumors are cancerous and can invade or spread to distant sites. All samples were collected using the Surgical Open Biopsy (SOB) method, which provides larger tissue samples compared to needle biopsies. Image filenames include details such as biopsy method, tumor classification, tumor type, patient ID, and magnification factor.

3.3 Data preparation

3.3.1 Data cleaning and pre – processing

1. **Reorganization of data:** Sort images into 'benign' and 'malignant' folders and further organize them into training, validation, and test sets.
2. **Data Transformation and Loading:** The dataset images are resized, normalized, and converted to tensors. Datasets and Data Loaders are created for training, validation, and test sets to facilitate model training and evaluation.
3. **Handling Missing or Corrupted Data:** Identify and handle any corrupted or incomplete image files. Images with errors will be removed or repaired if possible.
4. **Ensure Accurate Labels:** Double-check that all annotations (benign or malignant) are accurate and consistently applied

METRICS OF SUCCESS

To assess the effectiveness and performance of our breast cancer detection model, we will use clear and measurable metrics to evaluate accuracy, robustness, and impact in distinguishing between benign and malignant tumors.

4.1 Accuracy

The ratio of correctly predicted instances to the total number of instances, providing a general measure of the model's overall performance.

4.2 Precision and recall

Precision measures how accurately the model identifies patients with breast cancer, ensuring that those predicted to have cancer truly have it, which reduces unnecessary treatments. Recall, on the other hand, assesses the model’s ability to detect all actual cases of breast cancer, ensuring that no patients with cancer are missed. Together, precision and recall provide a comprehensive view of the model’s effectiveness in both minimizing false positives and capturing all true cases.

4.3 F1 score

The F1 score combines precision and recall into a single metric, balancing both the accuracy of positive predictions and the ability to identify all relevant cases. In breast cancer detection, the F1 score is valuable for evaluating the model’s overall effectiveness, particularly when the dataset has imbalanced classes or when both false positives and false negatives have significant consequences.

MODELLING

A combination of DenseNet, CNN, and VGG16 architectures was selected for our image classification project due to their distinct advantages and effectiveness in handling complex image classification tasks.

DenseNet improves information flow between layers through dense connections, which reduces the number of parameters and enhances training efficiency by preventing overfitting. CNNs (Convolutional Neural Networks) form the backbone of modern image classification, with their ability to automatically learn spatial hierarchies of features through convolutional layers, making them ideal for extracting meaningful patterns from images. VGG16 was included for its deep architecture and simplicity, which offer a uniform design that achieves high accuracy in capturing complex patterns within images.

By leveraging these models, we aim to achieve high accuracy and robust performance in classifying images into their respective categories, capturing a diverse range of features from our dataset.

5.1 Model selection

1. **DenseNet-121:** Selected for its efficient design that enhances feature propagation while keeping the model lightweight. With an adapted final layer for binary classification, DenseNet-121 effectively utilized pre-trained weights, resulting in improved accuracy and performance.

2. **Custom CNN:** Designed to balance complexity and efficiency, this model features two convolutional layers and fully connected layers to extract key patterns from the images. It is tailored to address the class imbalance in the dataset, helping to distinguish between classes more effectively.

3. **VGG16:** Known for its strong performance in image classification, this model was used with a frozen base and a customized top layer for binary classification. VGG16's deep architecture captures intricate patterns, which aids in dealing with the class imbalance and enhancing classification accuracy.

5.2 Model architecture (CNN Model)

```
----------------------------------------------------------------
      Layer (type)               Output Shape         Param #
================================================================
          Conv2d-1         [-1, 32, 128, 128]             896
            ReLU-2         [-1, 32, 128, 128]               0
       MaxPool2d-3           [-1, 32, 64, 64]               0
```

Conv2d-4	[-1, 64, 62, 62]	18,496
ReLU-5	[-1, 64, 62, 62]	0
MaxPool2d-6	[-1, 64, 31, 31]	0
Linear-7	[-1, 1000]	61,505,000
Linear-8	[-1, 2]	2,002

==========

Total params: 61,526,394

Trainable params: 61,526,394

Non-trainable params: 0

Input size (MB): 0.19

Forward/backward pass size (MB): 13.23

Params size (MB): 234.70

Estimated Total Size (MB): 248.12

- This network is designed to handle complex image classification tasks with a large number of parameters, indicating its capacity for learning detailed features from the input data.
    1. **Layers:** The network includes convolutional layers for feature extraction, activation layers (ReLU), pooling layers for downsampling, and fully connected layers for classification.
    2. **Output Shape:** The dimensions of the data after each layer, showing how the size changes through the network.
    3. **Parameter Counts:** The total number of parameters (weights and biases) in the network is 61,526,394, all of which are trainable.
    4. **Memory Usage:** The network requires 0.19 MB of input size, 13.23 MB for the forward/backward pass, and 234.70 MB for parameters, totaling an estimated 248.12 MB of memory.

5.3 Model performance

Model parameters	**Dense Net Model**	**CNN Model**	**VGG16 Model**
Accuracy	0.9858	0.9048	

Precision	0.9850	0.9040	
Recall	0.9821	0.9048	
F1 score	0.9836	0.9037	

5.3.1 Dense Net model

The performance of the Dense Net model is as shown below:

1. **High Accuracy (98.58%):** Most predictions made by the model are correct.
2. **High Precision (98.50%):** Most predictions of malignant cases are accurate, meaning few benign cases are wrongly classified as malignant.
3. **High Recall (98.21%):** Most actual malignant cases are correctly identified by the model, meaning few malignant cases are missed.
4. **High F1 Score (98.36%):** Indicates a strong balance between precision and recall, suggesting overall excellent performance.

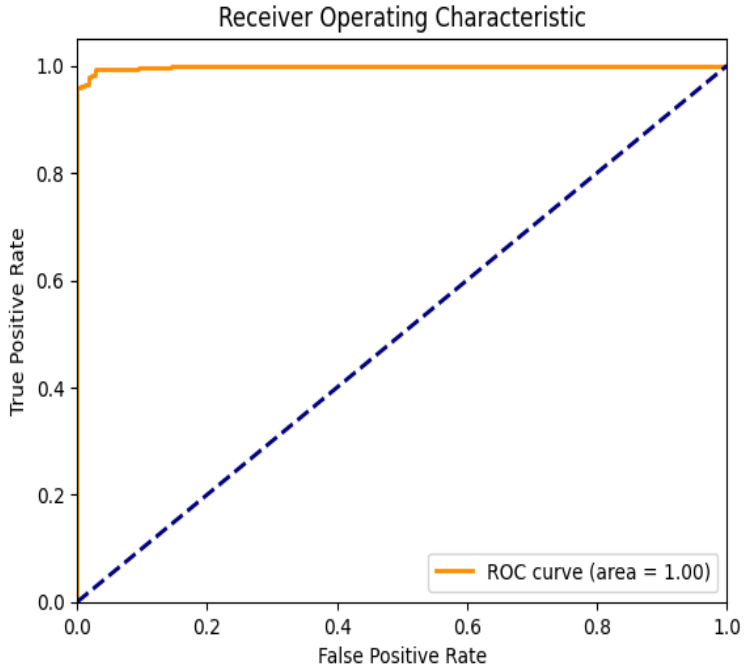
5.3.2 CNN model

The performance of the CNN model is as shown below:

1. **Test Accuracy (0.9048):** The model achieved an accuracy of 90.48% on the test set, indicating that most predictions were correct.
2. **Precision (0.9040):** The model's precision of 90.40% shows that most positive predictions were accurate, with few benign cases misclassified as positive.
3. **Recall (0.9048):** With a recall of 90.48%, the model successfully identified most actual positive cases, missing very few.
4. **F1 Score (0.9037):** The F1 Score of 90.37% reflects a strong balance between precision and recall, indicating robust overall performance.

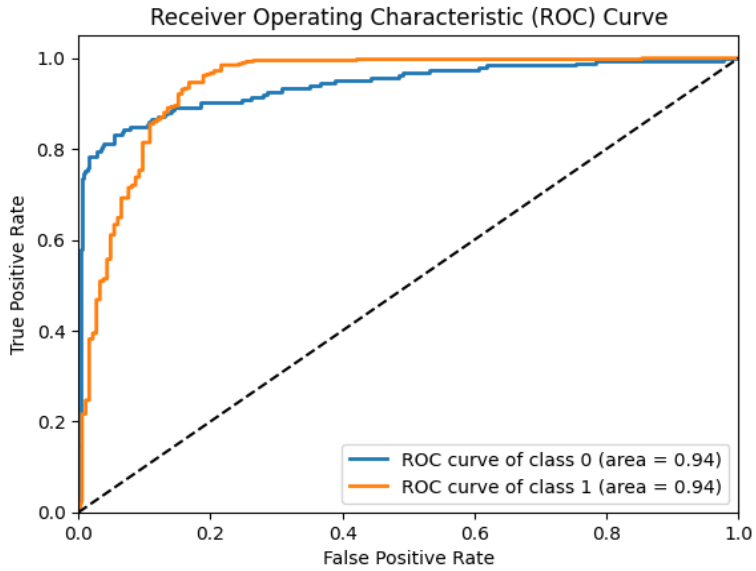
5.4 Model evaluation

5.4.1 ROC/AUC curve



- The DenseNet model achieved an exceptional ROC AUC score of 1.0, indicating perfect performance in distinguishing between the classes.

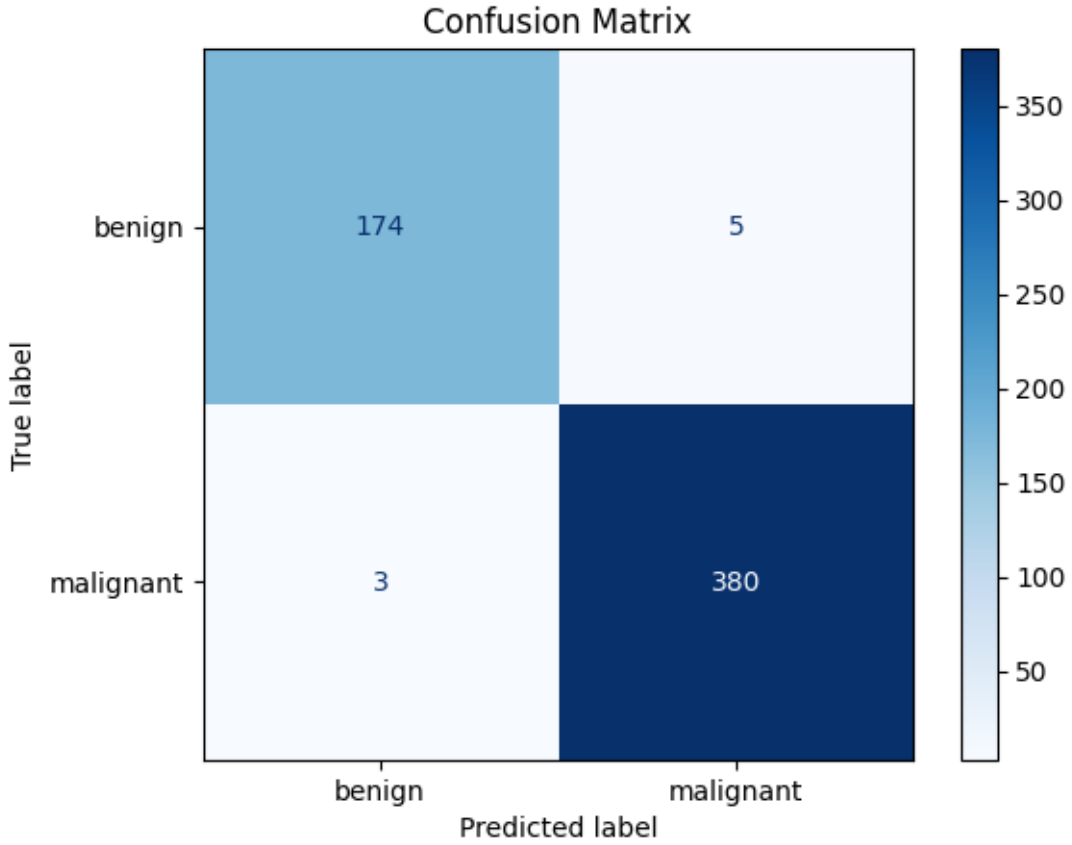
- This perfect score suggests that the model can flawlessly separate positive and negative cases based on the receiver operating characteristic curve. However, while an ROC AUC of 1.0 is impressive, it may also indicate potential overfitting



- The CNN model achieved a robust ROC AUC score of 0.94, demonstrating strong performance in distinguishing between the classes.

- This high score reflects the model's effective ability to differentiate between positive and negative cases based on the receiver operating characteristic curve.

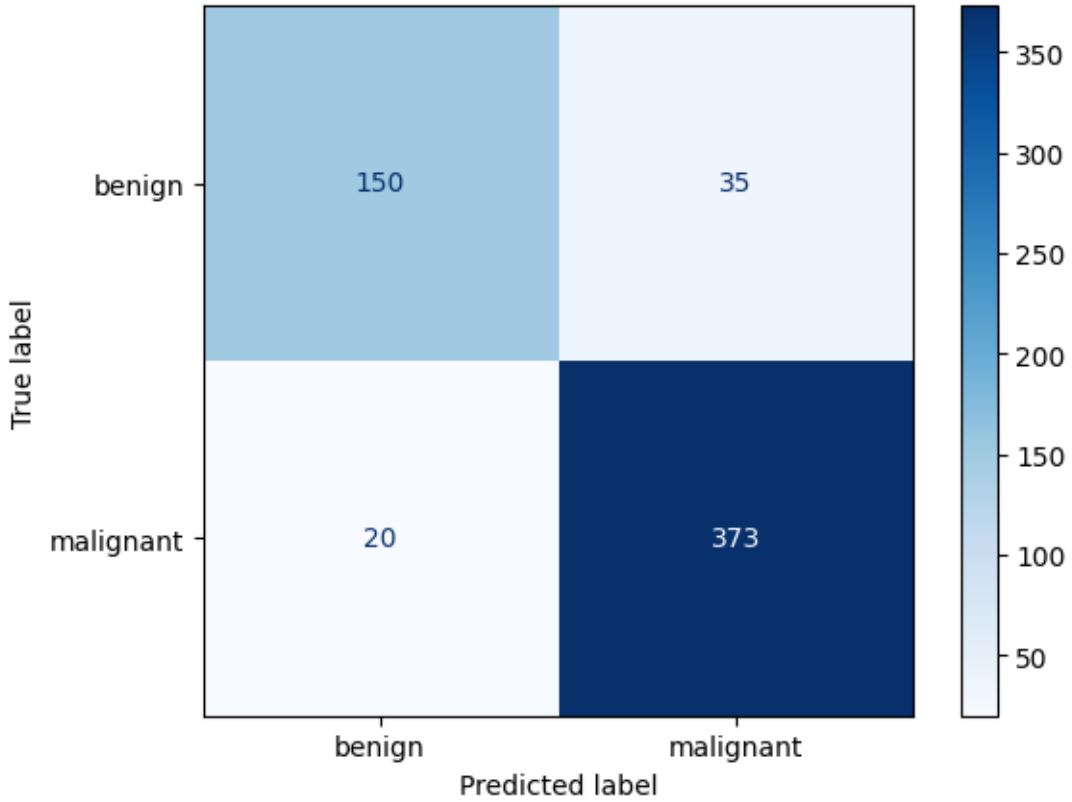
5.4.2 Confusion matrix



This matrix provides a detailed view of the Dense Net model's performance:

- **True Negatives (174):** The model correctly identified 174 benign cases.
- **False Positives (5):** There were 5 instances where the model incorrectly classified benign cases as malignant.
- **False Negatives (3):** The model missed 3 malignant cases, classifying them as benign.
- **True Positives (380):** The model correctly identified 380 malignant cases.

The high number of true positives and true negatives indicates that the model is performing exceptionally well, with only a few misclassifications in both positive and negative categories.



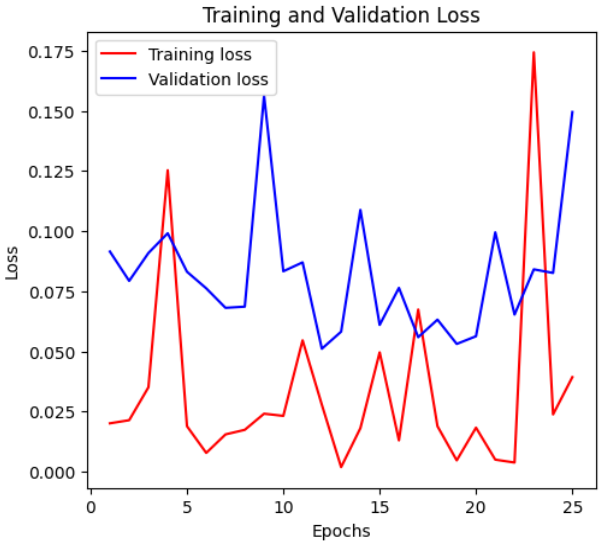
This matrix provides a detailed view of the CNN model's performance:

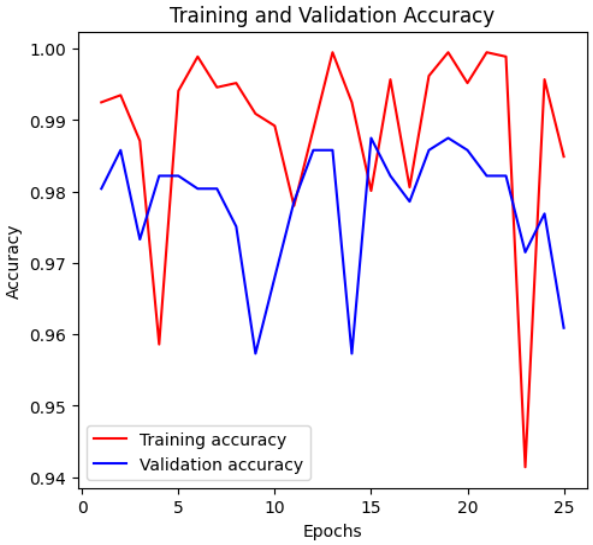
- **True Negatives (150):** The model correctly identified 150 benign cases as benign.
- **False Positives (35):** There were 35 instances where the model incorrectly classified benign cases as malignant.
- **False Negatives (20):** The model missed 20 malignant cases, classifying them as benign.
- **True Positives (373):** The model correctly identified 373 malignant cases as malignant.

The model shows a strong performance overall, with a high number of true positives and true negatives, though there are some misclassifications, particularly in false positives and false negatives.

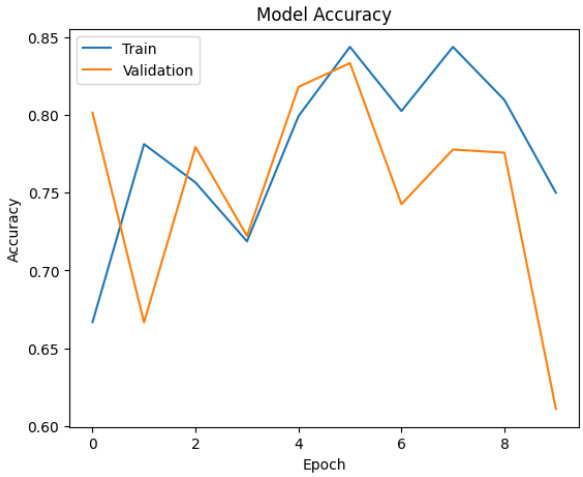
5.4.3 Model loss and accuracy

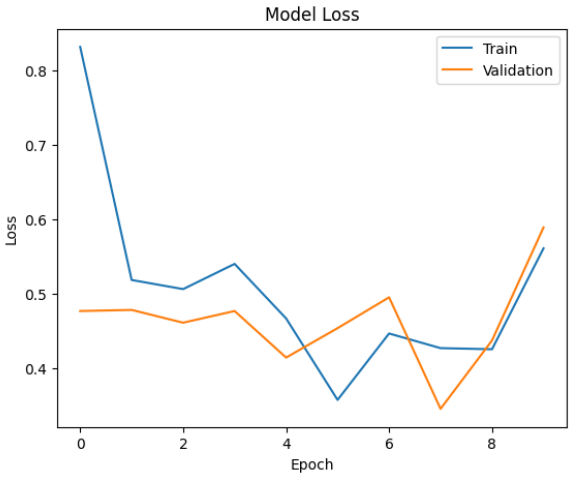
1. Dense net model





2. CNN model





- The comparison of model loss and accuracy graphs highlights why the CNN model is preferred over the DenseNet model.
- For the DenseNet model, the training loss is lower and the training accuracy is higher than the validation loss and accuracy, respectively, with noticeable high variance. This suggests that the DenseNet model may be overfitting, meaning it performs well on the training data but less reliably on new, unseen data.
- In contrast, the CNN model exhibits a more stable performance with higher validation accuracy compared to training accuracy and lower validation loss compared to training loss. This stability indicates that the CNN model generalizes better to new data, making it a more reliable choice for our image classification task.

5.5 Best performing model

All models performed well, showing high precision, recall, and accuracy. However, we selected the CNN model as the best performer.

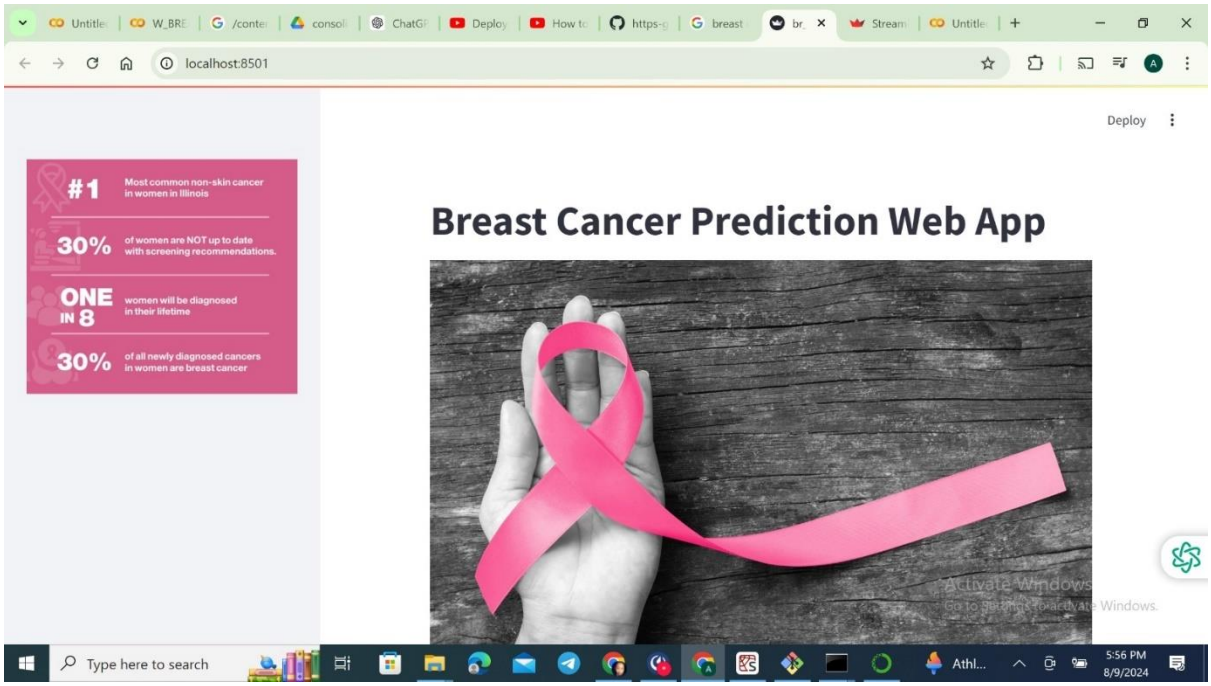
The DenseNet model exhibited significant variance in training and validation loss and accuracy, suggesting potential overfitting. In contrast, the CNN model demonstrated more stability and consistency, indicating it is likely to generalize better to new and unseen data. Therefore, we chose the CNN model for its robust performance and reliability.

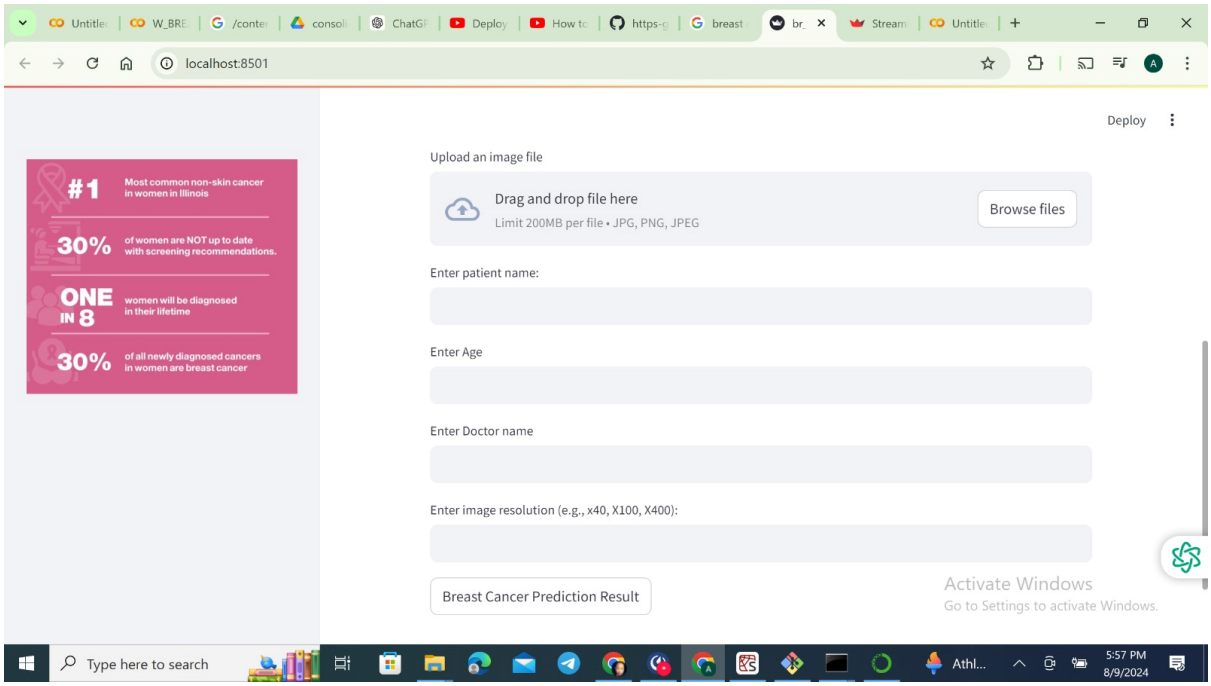
DEPLOYMENT

The model was deployed using Streamlit, allowing medical practitioners to upload histopathological images for real-time diagnosis.

- Performance Monitoring: Implement logging and monitoring to track the model's performance and schedule periodic retraining with updated data.

Below, are images of our user interface;





RECOMMENDATIONS

1. **Create Public Awareness:** Increase public awareness about the importance of early breast cancer detection and the role of advanced diagnostic technologies in improving outcomes.
2. **Integration into Healthcare Systems:** Integrate the developed model into the Kenyan healthcare system to assist pathologists and enhance diagnostic accuracy and reduce turnaround time
3. **Training and Education:** Provide training for healthcare professionals on the use and interpretation of the machine learning model to ensure effective implementation.

NEXT STEPS

1. **Data Expansion:** Continuously expand and update the dataset with new histopathological images to improve the model's robustness and accuracy over time.
2. **Model Refinement:** Explore advanced techniques such as transfer learning, fine-tuning, and additional ensemble methods to further enhance model performance.
3. **Integration with Clinical Systems:** Develop interfaces for integrating the model with existing clinical systems to streamline workflow and facilitate real-time diagnosis.
4. **Continuous Monitoring and Feedback Loop:** Establish a feedback loop with medical professionals to regularly refine the model. Utilize their insights to enhance predictions and update the model with new data over time.

CHALLENGES AND LIMITATIONS

1. **Computational Resources:** Training deep learning models on large datasets demands substantial computational power and memory, which can be resource-intensive.
2. **Model Overfitting:** Complex models like DenseNet may overfit the training data, reducing their ability to generalize to new data. Due to time constraints, we couldn't explore alternative methods to address this issue.

CONCLUSION

Successfully developed an image classification model for breast cancer detection with an accuracy of 90.4%. Additionally, a web-based app was created that allows medical practitioners to upload histopathological images, providing predictions on whether the condition is benign or malignant. This model aids in the early detection and management of breast cancer.

With a detection time of 45 seconds, the model aids in faster diagnoses, potentially shortening turnaround times and enhancing patient outcomes.

STUDY DISSEMINATION PLAN

Timeline	Activity
Week 1 (12th – 18th August)	Data sourcing
Week 2 (19th – 25th August)	Data cleaning and EDA
Week 3 (26th – 31st August)	Modelling
Week 4 (1st – 9th August)	Deployment

REFERENCES

1. Kenya Cancer Network. (n.d.). Kenya cancer facts. Retrieved August 9, 2024, from https://kenyacancernetwork.wordpress.com/kenya-cancer-facts/
2. Kenya Ministry of Health. (2021). *Kenya breast cancer action plan 2021-2025*. ICCP Portal. https://www.iccp-portal.org/system/files/plans/Kenya%20Breast%20Cancer%20Action%20Plan%202021-2025_compressed.pdf
3. Faraja Cancer Support. (n.d.). Breast cancer therapy. Faraja Cancer Support.https://farajacancersupport.org/therapy/Breast/index.html#:~:text=Biopsy%20%2D%20This%20is%20done%20when,shows%20a%20suspicious%20breast%20mass.