Breast Cancer Detection Using Machine Learning

Session 2020-2024

A project submitted in partial fulfillment of the degree of

BS in Computer Science

Department of Computer Science

Faculty of Computer Science & Information Technology

The Superior University, Lahore

Spring 2024

Type (Nature of project)	[ ] **D**evelopment [ ] **R**esearch [✓] **R&D**
Area of specialization	AI in Healthcare
FYP ID	FYP-BCSM-F23-014

Project Group Members

Sr.#	Reg. #	Student Name	Email ID	*Signature
(i)	BCSM-F20-112	Nashra Noor(Leader)	bcsm-f20-112@superior.edu.pk	
(ii)	BCSM-F20-153	Eshaw Asif	bcsm-f20-153@superior.edu.pk	
(iii)	BCSM-F20-333	Hasrat Irshad	bcsm-f20-333@superior.edu.pk	

*The candidates confirm that the work submitted is their own and appropriate credit has been given where reference has been made to work of others

Plagiarism Free Certificate

This is to certify that, I Nashra Noor Daugther of Noor Islam , Group leader of FYP under registration no _ FYP-BCSM-F23-014 at computer science Department, The Superior University, Lahore. I declare that my FYP report is checked by my supervisor.

Date: 25-06-2024 Name of Group Leader: __Nashra Noor__ Signature: _____

Name of Supervisor: Dr. Jawad Ahmad Co-Supervisor: Mr. Talha Amjad

Designation: Assistant Professor Designation: Lecturer

Signature: _____ Signature: _____

HoD: Dr. Irfan ud din

Signature: _____

Project Report

"Breast Cancer Detection using Machine learning"

Change Record

Author(s)	Version	Date	Notes	Supervisor's Signature
Nashra, Hasrat, Esha	1.0	23 July 2023	<Corrected Literature Review Papers>	
Nashra	1.1	10 Sep 2023	<Updated Model to Avoid Overfitting>	
Nashra, Hasrat, Esha	1.2	20 Sep 2023	<Updated Presentation and Added Limitations in Literature Review>	
Nashra	1.3	15 Oct 2023	<Guidance on Adding Figures in Documentation>	
Nashra	1.4	25 Nov 2023	<Guidance on Creating Model Images for Research Paper>	
Nashra	1.5	5 Dec 2023	<Implemented Sign-Up and Login Functionality in Mobile App>	
Nashra	1.6	15 Jan 2023	<Integrated Image Upload Options for Mammogram X-rays>	
Nashra	1.7	25 April 2023	<Connected Mobile App with Flask Server for Predictions>	
Nashra	1.8	5 May 2023	<Trained CNN Model on Breast Histopathology Dataset>	
Nashra	1.9	15 June 2023	<Conducted Accuracy Testing and Achieved 85% Accuracy>	

APPROVAL

PROJECT SUPERVISOR

Comments: _____

_____

Name:_____

Date:_____ Signature:_____

PROJECT MANAGER

Comments: _____

_____

Date:_____ Signature:_____

HEAD OF THE DEPARTMENT

Comments: _____

_____

Date:_____ Signature:_____

Dedication

I dedicate this project to my parents, whose unwavering support and encouragement have been my guiding light throughout this journey. Their belief in my abilities and constant motivation have fueled my passion and dedication towards this project. I also extend my heartfelt gratitude to my mentors and friends for their invaluable advice and moral support, which have been instrumental in the successful completion of this project.

Acknowledgements

I would like to express my deepest gratitude to my Supervisor, Dr. Jawad Ahmad, for his invaluable guidance, encouragement, and support throughout this project. His expertise and insights were crucial in shaping the direction and execution of this research.

I am also immensely grateful to my colleagues and friends for their constant support, helpful discussions, and encouragement. Special thanks to my team members for their hard work and dedication, without which this project would not have been possible.

I extend my sincere thanks to Superior University for providing the necessary resources and facilities for conducting this research. Additionally, I would like to acknowledge the authors and researchers whose work contributed significantly to the literature review and foundation of this project.

Finally, I am thankful to my family for their love, patience, and continuous support during the entire duration of this project. Their understanding and sacrifices have been the bedrock of my success.

Executive Summary

This project report outlines the development of a mobile application designed to enhance early breast cancer detection using advanced machine learning techniques. Breast cancer remains a global health challenge, with the American Cancer Society projecting 313,510 new cases in the U.S. in 2024, including 310,720 women. Early detection is crucial for improving survival rates and preventing disease spread.

The application employs a Convolutional Neural Network (CNN) model trained on a breast histopathology images dataset, providing a user-friendly interface for uploading mammogram X-ray images. The app processes these images via a server using Flask, delivering cancer likelihood predictions with 85% accuracy. This tool addresses the need for accessible and reliable diagnostic solutions, aiming to save lives through timely intervention.

Current diagnostic methods often lack accessibility and detailed analysis, limiting their effectiveness. Our project uses advanced algorithms to improve detection accuracy by interpreting mammogram details, facilitating early treatment. The report includes a literature review, highlighting recent advancements and methodologies in this field. Our focus on accuracy and user accessibility positions this tool as a valuable asset in medical diagnostics.

With an estimated 310,720 new invasive and 56,500 non-invasive breast cancer cases in 2024, and over 4 million survivors in the U.S., the significance of early detection is clear. Despite advancements, challenges such as varying outcomes among ethnic groups and limited screening accessibility persist. Initiatives like Pinktober raise awareness, emphasizing the importance of early detection.

This project's goal is to revolutionize breast cancer detection, leading to better outcomes and higher survival rates. By providing an effective early detection tool, we contribute to global efforts to combat breast cancer and improve patient care.

(American Cancer Society, 2022; National Breast Cancer Foundation, n.d.; Pink Ribbon Pakistan, n.d.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Chapter 1: Introduction

Breast cancer is the second leading cause of cancer deaths among women worldwide, including in Pakistan. This type of cancer starts when cells in the breast begin to grow uncontrollably. Most breast cancers originate in the ducts that carry milk to the nipple, which are referred to as ductal cancers. If malignant cells infiltrate the blood or lymphatic system, the cancer can spread to other parts of the body. Screening mammograms, specialized X-ray examinations, are typically used to detect cancerous breast tumors early, making it possible to identify abnormal growths before they become more serious.

Globally, breast cancer is a major health issue that leads to significant illness and death among women. In Pakistan, the incidence of breast cancer is particularly high, which makes early detection essential for successful treatment and improving survival rates. However, existing diagnostic tools often face challenges related to accessibility, cost, and accuracy, especially in regions with limited healthcare infrastructure. These limitations can delay diagnosis and treatment, which adversely affects patient outcomes.

Recent advancements in machine learning, particularly Convolutional Neural Networks (CNNs), offer promising solutions to these challenges by enabling detailed analysis of medical images. CNNs can identify intricate patterns in histopathology images that might not be easily discernible to the human eye, thus enhancing diagnostic precision and reliability.

Our project focuses on developing a reliable breast cancer detection system using CNNs, designed to be accessible via a mobile application. This app, created using Flutter, ensures a user-friendly experience and widespread accessibility. It communicates with a backend API developed with Flask, where the trained CNN model is hosted. This setup ensures a seamless experience for both healthcare providers and patients, facilitating early detection and prompt treatment.

The dataset for training and validation includes a substantial number of breast histopathology images, which are essential for precise diagnosis. Advanced data preprocessing techniques were

employed to standardize and augment the data, thereby improving the model's performance. We designed an effective CNN architecture tailored for this task, focusing on maximizing detection accuracy while minimizing overfitting. Rigorous testing and validation procedures confirmed our model's robustness and effectiveness, highlighting its potential as a reliable tool for breast cancer screening.

This paper provides a comprehensive overview of our project's goals, methods, and outcomes. We delve into the technical aspects of the CNN model, discussing the various layers and functions that contribute to its performance. The model consists of multiple convolutional layers that help in extracting detailed features from the images, followed by pooling layers that reduce the dimensionality of the data, making the computations more efficient. We also incorporate dropout layers to prevent overfitting and ensure that the model generalizes well to new, unseen data. Additionally, we address the challenges encountered during development and the solutions implemented to enhance the model's accuracy and reliability.

The integration of the model with the Flutter app is also explored, emphasizing its practical applications and broader implications for breast cancer diagnostics. The app allows users to upload mammogram images, which are then processed by the CNN model to provide predictions about the likelihood of breast cancer. This feature makes it possible for healthcare providers to quickly assess the risk and decide on further diagnostic steps or treatment options.

By leveraging the power of machine learning and mobile technology, we aim to make a significant impact on the early detection and treatment of breast cancer. Our project has the potential to save lives and improve health outcomes for women worldwide, including those in Pakistan, by providing a more accessible and accurate diagnostic tool for this prevalent disease. The user-friendly mobile application ensures that even in areas with limited access to healthcare facilities, women can benefit from early detection, leading to timely and effective treatment. This innovative approach not only enhances the accuracy of breast cancer diagnosis but also makes advanced medical technology accessible to a broader population, thereby contributing to global health improvements.

1.1. Background

Breast cancer continues to be a significant global health challenge due to its high prevalence and mortality rates. Recent statistics from the American Cancer Society (ACS) project that 313,510 people in the U.S. will be diagnosed with breast cancer in 2024, including 310,720 women. This alarming figure confirms that breast cancer is the most common cancer among women worldwide. Early detection is essential for preventing the disease's spread and improving survival rates. However, current diagnostic tools often lack accessibility and detailed analytical capabilities, highlighting a critical gap in modern medical technology. This situation underscores the urgent need for innovative solutions that can offer more accurate and accessible diagnostic options. (American Cancer Society, 2022)

1.2. Motivations and Challenges

1. The persistent prevalence and significant mortality rates of breast cancer highlight the urgent need for improved diagnostic methods. Despite advances in medical technology, existing tools are often limited in accessibility and their ability to provide detailed analyses. These limitations can lead to delays in diagnosis and treatment, negatively impacting patient outcomes. The high incidence of breast cancer globally, especially the alarming projections for 2024, underscores critical gaps in current diagnostic approaches and emphasizes the necessity for more innovative and effective solutions.
2. This project is driven by the potential of machine learning to significantly enhance the accuracy of breast cancer detection. By leveraging advanced algorithms and vast datasets, machine learning offers a level of precision and detail that traditional diagnostic tools often lack. This technological advancement promises to address many shortcomings of current methods, such as limited accessibility and insufficient analytical depth. However, the challenge lies not only in developing a highly accurate system but also in ensuring it is user-friendly and can be widely adopted. The goal is to create a diagnostic tool that is both sophisticated in its capabilities and simple enough for widespread use, including by individuals without specialized medical training. Achieving this balance is crucial for the

tool's effective utilization and for making a tangible impact on breast cancer detection and treatment.

3. In summary, the motivation behind this project is to harness the power of machine learning to revolutionize breast cancer diagnostics. By improving accuracy and accessibility, this project aims to facilitate earlier detection, prompt treatment, and ultimately better outcomes for patients. The challenge is significant, but the potential benefits in terms of saving lives and improving health care quality make it a worthy endeavor.

1.3. Goals and Objectives

This project aims to revolutionize breast cancer detection by developing a mobile application that leverages advanced machine learning algorithms. Using a Convolutional Neural Network (CNN) model trained on breast histopathology images, the app seeks to provide highly accurate predictions of cancer presence based on mammogram interpretations. The application is designed to be intuitive, enabling users to register, log in, and upload images from their gallery or camera. Upon selecting "check result," the app interfaces with a Flask server to deliver a percentage-based prediction indicating the likelihood of cancer.

Objectives:

1. **Enhance Detection Accuracy:** Implement advanced machine learning algorithms to enhance the precision of breast cancer detection.
2. **Early Detection:** Emphasize early detection to mitigate breast cancer spread and improve survival rates.
3. **Accessibility:** Develop a user-friendly mobile application accessible to individuals without specialized medical knowledge.
4. **Detailed Analysis:** Provide comprehensive analysis and predictions to support early diagnosis and treatment of breast cancer.
5. **Impact:** Contribute to improved outcomes and higher survival rates for breast cancer patients.

1.4. Literature Review:

Breast cancer remains a significant global health concern, necessitating continual advancements in diagnostic methodologies to improve patient outcomes. This review examines recent research endeavors aimed at enhancing breast cancer detection through the integration of advanced technologies such as Deep Learning (DL), Machine Learning (ML), Convolutional Neural Networks (CNN), and Artificial Intelligence (AI).

1. Abhisheka et al. (2023) extended the scope of breast cancer detection by integrating ML, DL, and CNN techniques across diverse datasets. With an exceptional accuracy rate of 97.49%, their research signifies a substantial advancement in screening and diagnosis, promising improved patient care through early detection and intervention. (Abhisheka et al., 2023)

2. Unaiza Sajid et al. (2023) developed an innovative framework for classifying breast cancer using CNNs and ResNet50 on the CBIS-DDSM dataset, achieving impressive accuracies of 89.9% and 96%. By integrating CNN-derived features with handcrafted attributes such as Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP), their study highlights the potential of AI in automating breast cancer detection and emphasizes the necessity of comprehensive datasets and transparent models for clinical application. (Sajid et al., 2023)

3. Lakshmi Narayanan et al. (2022) introduced a hybrid approach combining AlexNet for feature extraction with a random forest classifier for breast cancer detection. Their method achieved a notable accuracy rate of 98%, surpassing previous techniques and underscoring the importance of early detection and the practical impact of AI in medical imaging. (Narayanan et al., 2022)

4. Anum Masood et al. (2023) introduced a modified Swin Transformer model (mST) with a Multi-Level Adaptive Feature Fusion (MLAFF) Module for detecting metastatic breast cancer

lesions. Their model achieved a high Dice-Similarity coefficient of 98% for segmentation and an average classification accuracy of 94.5% across seven benchmark datasets, including INBreast and CBIS-DDSM. The study underscores the clinical relevance of their approach, aiming to assist radiologists in automated breast cancer detection and classification. (Masood et al., 2023)

5. Sujata Kulkarni and Rinku Rabidas et al. (2023) proposed a deep learning architecture based on U-Net for detecting and characterizing breast masses. Their model achieved a true positive rate of 99.64% on the INBreast dataset and 97.36% at 0.25 false positives per image on the DDSM dataset, demonstrating strong performance in mass detection and characterization. The study highlights the role of data augmentation in improving model effectiveness and suggests future enhancements to increase sensitivity and reduce false positives. (Kulkarni & Rabidas, 2023)

6. Sathiyabhama Balasubramaniam et al. (2023) utilized the classic LeNet CNN architecture for breast cancer detection and diagnosis. By incorporating corrected ReLU and batch normalization techniques, their approach achieved an accuracy of 89.91% in recognizing breast images. The study emphasizes the effectiveness of ensembling multiple LeNet models to improve overall performance and reliability in predicting breast cancer from clinical images. (Balasubramaniam et al., 2023)
7. Zeyad Q. Habeeb et al. (2023) proposed a novel framework for breast cancer detection using the RetinaNet model and a two-stage transfer learning approach. Their method achieved a high True Positive Rate (TPR) of 0.99 ± 0.02 at 1.67 False Positives per Image (FPPI) on the INBreast dataset, outperforming existing methods. The study highlights the effectiveness of efficient object detection and strategic transfer learning in enhancing breast cancer detection accuracy. (Habeeb et al., 2023)
8. Zahra Jafari et al.(2023) presented an advanced method for diagnosing breast lesions using pre-trained CNN models and feature extraction techniques. Their neural network-based classifier achieved significant accuracies of 92% on the RSNA dataset, 94.5% on the MIAS

dataset, and 96% on the DDSM dataset. This research underscores the method's superiority in accurately diagnosing breast lesions compared to state-of-the-art approaches, showing promise for improving breast cancer prognosis through precise diagnosis. (Jafari & Karami, 2023)

9. Abhilash Pati et al. (2023) presented a groundbreaking study on breast cancer diagnosis, emphasizing early detection through Deep Transfer Learning (DTL), Fog computing, and IoT. Their OMLTS-DLCN framework incorporated CNNs such as ResNet50 and InceptionV3, achieving outstanding diagnostic metrics with 97.99% accuracy and a notable f1-score of 98.97%. The research underscored Fog computing's role in enhancing privacy and operational efficiency, proposing a promising approach for remote and real-time breast cancer diagnostics. (Pati et al., 2023)

10. Huafeng Hu et al. (2023) introduced TGA-MIL, a groundbreaking model for medical image analysis. By integrating three kernel functions (LA, RBF, IM), TGA-MIL improves instance importance assessment and attention map generation. It achieves exceptional performance in binary and weakly supervised classification tasks across diverse datasets like MNIST, USBC breast cancer, and DDSM, effectively managing challenges associated with limited labeled data. (Hu et al., 2023)

11. Ziquan Zhu et al. (2023) proposed an advanced CNN-based approach for breast cancer diagnosis. Their study categorizes diagnostic tasks into classification, detection, and segmentation, highlighting the rapid advancements in CNN applications while acknowledging challenges such as dataset scarcity and computational complexity. The paper advocates for enhancing diagnostic accuracy through the use of pretrained models and innovative feature extraction techniques, outlining future research directions in this evolving field. (Zhu et al., 2023)

1.5. Gap Analysis

1. The literature review reveals several critical limitations in current breast cancer detection models that are essential for the gap analysis of this project. These models frequently contend with challenges such as biased and imbalanced datasets, which can compromise the accuracy and reliability of predictions during model training and application. Additionally, the presence of redundant features within datasets often hampers efficiency, while limited generalizability across diverse patient populations and imaging modalities poses significant obstacles to widespread adoption. High computational complexity further restricts the practical deployment of these models in clinical settings. Moreover, the lack of peer-reviewed validation studies for many existing models raises concerns regarding their robustness and suitability for real-world applications.
2. Furthermore, the review highlights specific limitations like the reliance solely on frontal CXR scans without incorporating lateral views, which limits the comprehensive evaluation of breast cancer lesions. The absence of integrated clinical histories also undermines diagnostic accuracy and the ability to personalize treatment strategies effectively. Additionally, challenges related to inconsistent heat maps and imprecise bounding boxes in methods like Grad-CAM underscore the need for refined localization techniques. These issues collectively emphasize the critical gaps that must be addressed to enhance the effectiveness, reliability, and clinical utility of breast cancer detection models in practice.

1.6. Proposed Solution

To address the critical limitations identified in the literature review, our project aims to develop a mobile application for breast cancer detection using advanced machine learning techniques. This application will leverage a Convolutional Neural Network (CNN) model trained on a comprehensive dataset of breast histopathology images. Our approach seeks to overcome

challenges such as biased and imbalanced datasets, redundant features, and limited generalizability by implementing the following key strategies:

1. **Balanced and Diverse Dataset:** Utilizing a balanced and diverse dataset to train the CNN model ensures accurate predictions across various patient demographics and imaging methods.
2. **Feature Selection and Reduction:** Implementing advanced techniques for feature selection and reduction to enhance model efficiency by eliminating redundant features.
3. **Integration of Clinical Histories:** Incorporating integrated clinical histories to improve diagnostic accuracy and facilitate personalized treatment approaches.
4. **Improved Localization Techniques:** Enhancing localization methods, including improved heat maps and precise bounding boxes, to effectively identify and evaluate breast cancer lesions.
5. **Validation Studies:** Conducting rigorous peer-reviewed validation studies to validate the model's reliability and suitability for real-world clinical applications.
6. **User-Friendly Interface:** Developing a user-friendly mobile application that enables easy registration, login, and uploading of mammogram images from either the device's gallery or camera. The application will provide detailed predictions in percentage format, indicating the probability of cancer presence.

1.7. Project Plan

The project plan outlines a systematic approach to ensure the successful development and deployment of the breast cancer detection mobile application. It includes detailed phases and tasks to thoroughly prepare, develop, and validate the application.

1.7.1. Work Breakdown Structure

- **Project Initialization:**

- Requirement analysis: Identifying and documenting project requirements.
- Literature review: Reviewing existing research and detection models.
- Project proposal and approval: Drafting and obtaining approval for the project proposal.

- **Data Collection and Preparation:**

  - Dataset acquisition: Gathering a comprehensive dataset of breast histopathology images.
  - Data cleaning and preprocessing: Ensuring data quality and preparing it for analysis.
  - Data augmentation: Enhancing the dataset to improve model training.

- **Model Development:**

  - Model selection and design: Choosing and designing the appropriate CNN model.
  - Training the CNN model: Training the model on the prepared dataset.
  - Model optimization and validation: Refining the model and validating its performance.

- **Application Development:**

  - Frontend development (UI/UX design): Designing the user interface for ease of use.
  - Backend development (Flask server integration): Setting up the server and integrating it with the model.
  - Integration of model with the mobile app: Ensuring seamless communication between the app and the server.

- **Testing and Validation:**

  - Unit testing: Testing individual components for functionality.
  - Integration testing: Ensuring all components work together as intended.
  - User acceptance testing: Validating the app with end-users for usability and accuracy.
  - Peer-reviewed validation studies: Conducting studies to confirm the model's robustness and reliability.

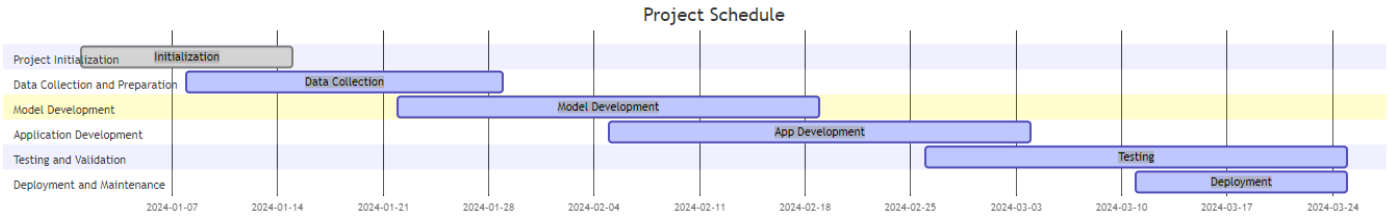
- **Deployment and Maintenance:**

- Deployment on app stores: Releasing the app on various app stores for public use.
- Continuous monitoring and updates: Regularly updating the app to improve performance and address issues.
- User support and feedback management: Providing ongoing support and collecting user feedback for future improvements.

1.7.2. Roles & Responsibility Matrix

Role	Responsibility
Project Manager	Oversee project progress, manage timelines, and coordinate tasks
Data Scientist	Data collection, preprocessing, model development, and optimization
Software Developer	Frontend and backend development, integration of model with the app
QA Engineer	Testing and validation of the application and model
Clinical Advisor	Provide clinical insights and validate the model's medical accuracy
Technical Writer	Documentation, report writing, and preparation of user manuals

1.7.3. Gantt chart



1.8. Report Outline

Introduction

- **Background:** Provide an overview of breast cancer prevalence, the importance of early detection, and the limitations of current diagnostic methods.
- **Motivations and Challenges:** Highlight the need for improved diagnostic tools and identify gaps in existing models.

- **Goals and Objectives:** Aim to enhance detection accuracy, promote early detection, and increase accessibility.

Literature Review

- **Current Breast Cancer Detection Models:** Summarize existing models and their methodologies.
- **Identified Limitations:** Discuss challenges such as biased datasets, redundant features, and lack of generalizability.
- **Gap Analysis:** Highlight critical gaps that need addressing to improve current detection models.

Proposed Solution

- **Mobile Application Overview:** Describe the app's functionalities and user interface.
- **Machine Learning Model Details:** Explain the CNN model and the training dataset.
- **Key Features and Innovations:** Showcase unique aspects of the solution, such as integrated clinical histories and improved localization techniques.

Project Plan

- **Work Breakdown Structure:** Provide a detailed breakdown of tasks and phases from project initialization to deployment.
- **Roles & Responsibility Matrix:** Allocate tasks and responsibilities among team members.
- **Gantt chart:** Offer a visual representation of the project timeline and key milestones.

Chapter 2

Software Requirement Specifications

Chapter 2: Software Requirement Specifications

2.1. Introduction

2.1.1. Purpose

This document outlines the software requirements for a mobile application aimed at breast cancer detection using a machine learning model, specifically a Convolutional Neural Network (CNN). The scope includes the entire system, from user registration and image upload to the display of predictions, ensuring the delivery of accurate and accessible diagnostic tools.

2.1.2. Document Conventions

This Software Requirements Specification (SRS) adheres to standard typographical conventions. Headings and subheadings are presented in bold. Requirements are uniquely identified with sequence numbers and emphasized in *italic*. Priority levels are categorized as High, Medium, or Low. The paragraph text is formatted in 12 pt Calibri, with 1.5 line spacing and justified alignment.

2.1.3. Intended Audience and Reading Suggestions

This document is designed for developers, project managers, testers, users, and documentation writers. It encompasses an overview, detailed requirements, and technical specifications. Readers should begin with the introduction and product scope, then proceed to sections pertinent to their respective roles.

2.1.4. Product Scope

The mobile application's goal is to improve breast cancer detection accuracy through a CNN model. Users can upload mammogram images to receive predictions regarding cancer presence. This software promotes early detection, thereby enhancing survival rates and addressing current limitations in diagnostic tools.

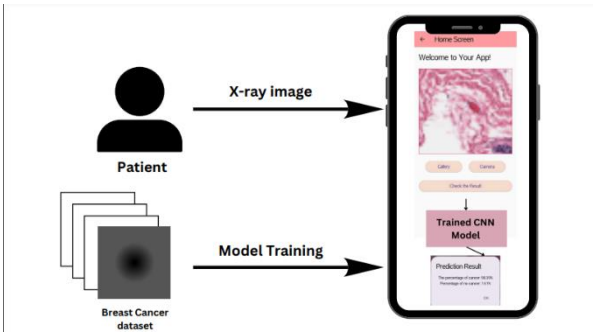
2.1.5. References

1. American Cancer Society. (2022). *Breast cancer facts & figures 2022-2024*. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/2022-2024-breast-cancer-fact-figures-acs.pdf
2. Abhisheka, B., Biswas, S. K., & Purkayastha, B. (2023). A Comprehensive Review on Breast Cancer Detection, Classification and Segmentation Using Deep Learning. *Archives of Computational Methods in Engineering*, *30*(8), 5023–5052. https://doi.org/10.1007/s11831-023-09968-z
3. U. Sajid, R. A. Khan, S. M. Shah, and S. Arif, "Breast Cancer Classification using Deep Learned Features Boosted with Handcrafted Features," *Biomedical Signal Processing and Control*, vol. 86, p. 105353, Sep. 2023, doi: 10.1016/j.bspc.2023.105353.
4. Narayanan, L., Krishnan, S., & Robinson, H. (2022). A Hybrid Deep Learning Based Assist System for Detection and Classification of Breast Cancer from Mammogram Images. *The International Arab Journal of Information Technology*, *19*(6). https://doi.org/10.34028/iajit/19/6/15
5. Masood, U. Naseem, and J. Kim, "Multi-Level Swin Transformer Enabled Automatic Segmentation and Classification of Breast Metastases," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia: IEEE, Jul. 2023, pp. 1–4. doi: 10.1109/EMBC40787.2023.10340831.
6. S. Kulkarni and R. Rabidas, "Fully convolutional network for automated detection and diagnosis of mammographic masses," *Multimed Tools Appl*, vol. 82, no. 29, pp. 44819–44840, Dec. 2023, doi: 10.1007/s11042-023-14757-8.

2.2. Overall Description

2.2.1. Product Perspective

This mobile application represents a novel, standalone product designed to offer a dependable and accessible breast cancer detection tool. It connects with a backend server for processing and predictions. The diagram below depicts the primary components and their interactions.



2.2.2. Product Functions

1. User sign-up and login
2. Image upload from gallery or camera
3. Prediction display in percentage form
4. User feedback and support

2.2.3. User Classes and Characteristics

- **Patients**: Non-technical individuals seeking easy access to diagnostic tools.
- **Healthcare Providers**: Medical professionals needing accurate and rapid diagnostic support.
- **Researchers**: Technical users interested in assessing model performance and validation.

2.2.4. Operating Environment

The software development will utilize Flutter for the mobile application and Flask for the backend server. Development tools include VS Code and Android Studio. The application will run on both Android and iOS devices, requiring internet connectivity for server communication. Compatibility with the latest versions of these operating systems will be maintained.

2.2.5. Design and Implementation Constraints

The application must adhere to healthcare data privacy regulations. It must consider the hardware limitations of mobile devices. Secure communication protocols (HTTPS) are essential. Specific machine learning libraries and frameworks, such as TensorFlow and Flask, are required.

2.2.6. User Documentation

- User guide
- Online assistance and tutorials
- Technical documentation for developers

2.2.7. Assumptions and Dependencies

- Stable internet access for server communication.
- Access to a labeled dataset for training the model.
- Regular updates to the machine learning model to enhance accuracy.

2.3. External Interface Requirements

2.3.1. User Interfaces

- User-friendly interface with easy navigation.
- Standard buttons (e.g., upload, predict, help) and error handling messages.
- Adherence to GUI standards for mobile applications.

2.3.2. Hardware Interfaces

- Utilization of the mobile device camera for image capture.
- Access to storage for selecting images from the gallery.

2.3.3. Software Interfaces

- Interaction with the Flask backend server.
- Communication with TensorFlow for model predictions.
- Incorporation of cloud services for data storage and processing.

2.3.4. Communications Interfaces

- Implementation of secure communication protocols (HTTPS) for data transfer.
- Adherence to data encryption standards to ensure patient data privacy

2.4. System Features

This section categorizes the product's functional requirements by its main features. Each feature details the services offered by the product, ensuring they are clearly defined and structured for straightforward understanding and implementation.

2.4.1. User Registration and Authentication

Users must sign up and log in to access the application. Priority: High.

2.4.1.1. Description and Priority

This feature enables users to create an account and log in to access the application's functionalities. Users can register using their email and password, receiving a verification email upon registration. This ensures secure access to the application and user-specific data.

2.4.1.2. Stimulus/Response Sequences

- User opens the app and selects "Sign Up."
- User enters personal details and submits the form.
- System verifies and creates the user account.

2.4.1.3. Functional Requirements

- REQ-UR1-1: Users must be able to register with their email and a password.
- REQ-UR1-2: The system must send a verification email upon registration.
- REQ-UR1-3: Users must be able to log in with verified credentials.

2.4.2. Image Upload and Processing

2.4.2.1. Description and Priority

Users can upload images from their gallery or capture new images using the camera. Priority: High.

2.4.2.2. Stimulus/Response Sequences

- User selects "Upload Image" and chooses a source (camera or gallery).
- User selects or captures an image.
- System uploads the image to the server for processing.

2.4.2.3. Functional Requirements

- REQ-IU2-1: Users must be able to upload images from their gallery.
- REQ-IU2-2: Users must be able to capture images using the camera.
- REQ-IU2-3: The system must process the image and provide a prediction.

2.4.3. Prediction Display

2.4.3.1. Description and Priority

The application displays the prediction results in percentage form. Priority: High.

2.4.3.2. Stimulus/Response Sequences

- System processes the uploaded image.
- System displays the prediction results to the user.

2.4.3.3. Functional Requirements

- REQ-PD3-1: The system must display the likelihood of cancer in percentage form.
- REQ-PD3-2: The system must provide an explanation of the results.
- REQ-PD3-3: Users must be able to view their prediction history.

2.5. Other Nonfunctional Requirements

2.5.1. Performance Requirements

- The application must process and display predictions within 10 seconds.
- The application should handle up to 10,000 simultaneous users.

2.5.2. Safety Requirements

- The application must comply with healthcare data privacy regulations.
- The system must prevent unauthorized access to patient data.

2.5.3. Security Requirements

- The application must use secure login methods and encrypt sensitive data.
- The application must adhere to standard encryption protocols for data transmission.

2.5.4. Software Quality Attributes

- **Usability:** The application should be easy to navigate for all user classes.
- **Reliability:** The application must provide consistent and accurate predictions.
- **Maintainability:** The application should be easy to update and maintain.

2.5.5. Business Rules

- Only verified users can access the prediction features.
- Users must agree to terms and conditions before using the application.

2.6. Other Requirements

The application must support localization to accommodate different languages. The application should be compatible with major mobile operating systems, including Android and iOS. The application must feature a feedback mechanism for users to report issues and suggest improvements.

Chapter 3

Use Case Analysis

Chapter 3: System Analysis

This chapter provides an in-depth examination of the system's functional requirements and user interactions. It outlines the system's design and how it interacts with users to achieve the desired outcomes. The analysis includes a Use Case Model, which visually represents the various user actions and system responses. Additionally, the Use Case Descriptions offer detailed insights into each interaction, specifying the steps, conditions, and requirements necessary to fulfill each use case. This comprehensive analysis ensures all functional aspects of the system are well-defined and aligned with user needs and project goals.

3.1. Use Case Model

The Use Case Model illustrates the interactions between users and the system, showing how users will utilize the application to achieve specific goals. It identifies the primary actors, such as patients, healthcare providers, and researchers, and maps their interactions with the system's major functions, including user registration, image upload, and prediction display. This model serves as a foundational blueprint for understanding the dynamics of user-system interaction and guides the development process by highlighting critical functionalities.

3.2. Use Case Descriptions

The Use Case Descriptions provide detailed narratives of each interaction depicted in the Use Case Model. Each use case description includes the following elements: a unique identifier, a brief summary, actors involved, preconditions, main flow of events, alternative flows, Post conditions, and any specific requirements. These descriptions ensure that each use case is thoroughly understood and correctly implemented, facilitating seamless user experiences and accurate system responses. The detailed use cases include:

- **User Registration and Authentication**: Steps for users to sign up, verify their email, and log in.
- **Image Upload and Processing**: Steps for uploading or capturing images and sending them to the server for analysis.
- **Prediction Display**: Steps for displaying prediction results and providing detailed explanations.

By detailing each use case, this chapter ensures that all potential user interactions are comprehensively analyzed and documented, forming a robust foundation for system development and implementation.

Chapter 4

System Design

Chapter 4: System Design

This chapter explores the structural and functional design of the system, offering a comprehensive blueprint for its implementation. It includes various diagrams that depict different aspects of the system architecture, data relationships, and interaction flows. These diagrams and models collectively ensure a clear understanding of the system's design, facilitating effective development and seamless integration of all components. The chapter covers the architecture diagram, domain model, entity-relationship diagram, class diagram, sequence/collaboration diagram, operation contracts, activity diagram, state transition diagram, component diagram, deployment diagram, and data flow diagram.

4.1. Architecture Diagram

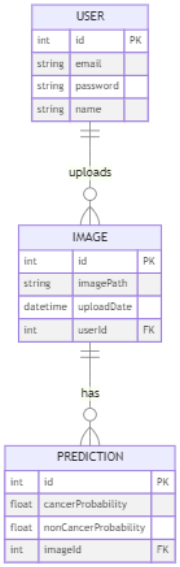
The architecture diagram offers a high-level overview of the system's structure, showcasing the primary components and their interactions. It highlights the overall design and the communication flow between the frontend, backend, and the machine learning model.

4.2. Domain Model

The domain model outlines the key entities, their attributes, and the relationships between them. It provides a conceptual view of the system's data structure and how different entities interact with each other.

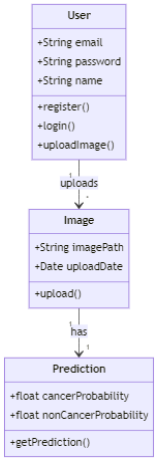
4.3. Entity Relationship Diagram with data dictionary

The entity relationship diagram (ERD) illustrates the relationships between entities in the database, providing a clear understanding of data storage and their interconnections.



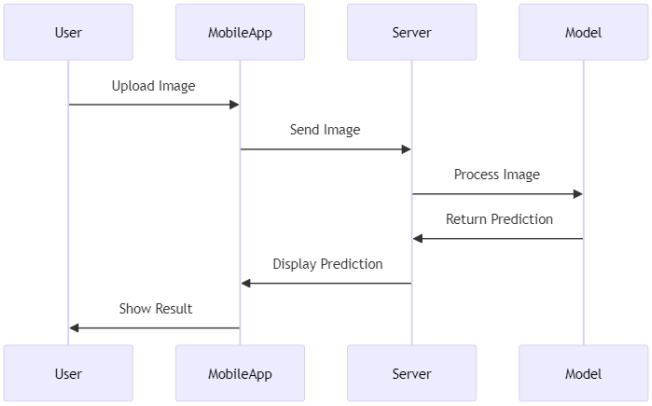
4.4. Class Diagram

The class diagram outlines the system's classes, their attributes, methods, and the relationships between them, ensuring a clear object-oriented design.



4.5. Sequence / Collaboration Diagram

The sequence diagram depicts the interactions between objects in a specific order, demonstrating how processes function and the sequence in which they occur.



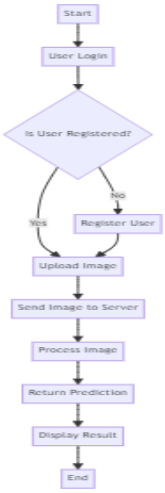
4.6. Operation contracts

Operation contracts specify the operations for each class, detailing the changes to the system's state and any side effects.

- Register (): Creates a new user account and initializes the user profile.
- Login (): Authenticates the user and grants access to the application.
- Upload Image (): Uploads an image for prediction and associates it with the user profile.
- Get Prediction (): Processes the image and returns the prediction results.

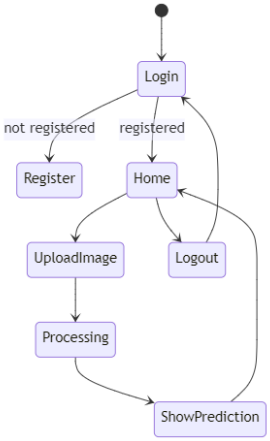
4.7. Activity Diagram

The activity diagram offers a dynamic perspective of the system by illustrating the workflow of various activities and their interactions.



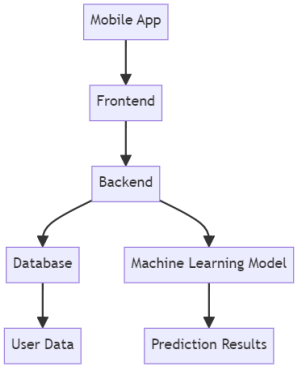
4.8. State Transition Diagram

The state transition diagram illustrates the various states of the system and how it moves from one state to another



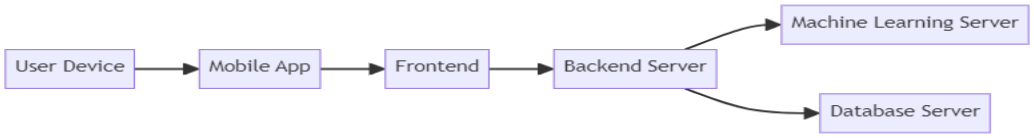
4.9. Component Diagram

The component diagram illustrates how various software components are organized and their dependencies.



4.10. Deployment Diagram

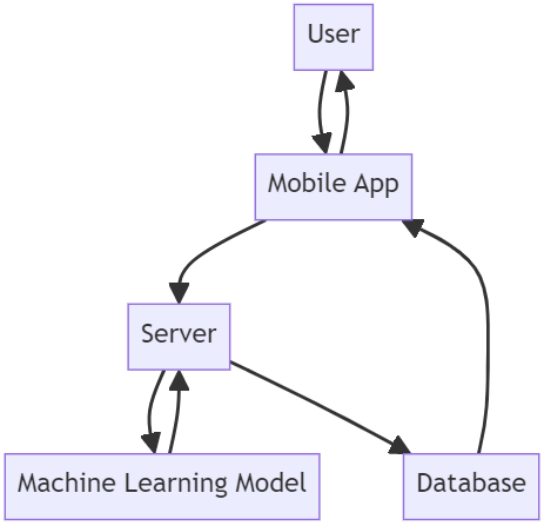
The deployment diagram illustrates how artifacts are physically deployed on nodes, including hardware and software components.



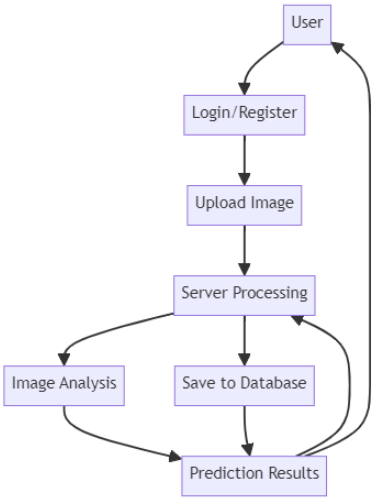
4.11. Data Flow Diagram:

The data flow diagram (DFD) illustrates how information flows within the system, depicting the steps involved in data processing.

Level 0


User
Mobile App
Server
Machine Learning Model
Database

Level 1


User
Login/Register
Upload Image
Server Processing
Image Analysis
Save to Database
Prediction Results

Chapter 5

Implementation

Chapter 5: Implementation

This chapter details the implementation process of the breast cancer detection mobile application. It covers the critical flow control and pseudo codes used to connect the frontend and backend, the components, libraries, and web services utilized, and the deployment environment. Additionally, it describes the tools and techniques adopted during development, best practices and coding standards followed, and the version control system employed to manage the project.

5.1. Important Flow Control/Pseudo codes

The implementation of the mobile application involves several key processes:

- **User Uploads Image**: Users can upload images from their mobile device using the Flutter frontend.
- **Backend Processing**: The image is sent to the Flask server, where it is processed using a pre-trained CNN model in TensorFlow.
- **Prediction Result**: The Flask server returns the prediction results to the mobile app, which then displays the probability of cancer

Pseudo Code:

```
def predict():

    # Check for file in request
    if 'file' not in request.files:
        return error_response('No file part')

    file = request.files['file']

    if file.filename == '':
        return error_response('No selected file')

    # Save the file
    file_path = save_file(file)

    # Make prediction
    prediction_service = Api_service("static/uploaded_images")
    has_cancer, has_no_cancer = prediction_service.prediction_function(file_path)

    if prediction_successful(has_cancer, has_no_cancer):
        return success_response(has_cancer, has_no_cancer)
    else:
        return error_response('Error during prediction')
```

5.2. Components, Libraries, Web Services and stubs

The project employs various components and libraries:

Components:

- **Flutter**: For the mobile application frontend.
- **Flask**: For the backend server.
- **TensorFlow**: For the machine learning model.
- **Jupyter Notebook**: For model development and experimentation.

Libraries:

- **pandas, NumPy, Matplotlib, Seaborn**: For data manipulation and visualization.
- **TensorFlow**: For deep learning and model training.
- **Flask-CORS**: To handle cross-origin requests.

Web Services:

- The Flask server manages image uploads, processing, and returns prediction results.

Stubs:

- Mock components were utilized during the initial development stages to simulate the prediction process.

5.3. Deployment Environment

The deployment environment for the application includes:

Frontend:

- **Platform**: Android and iOS
- **Development Tool**: Android Studio with Flutter

Backend:

- **Platform**: Cloud-based server or local server setup
- **Development Tool**: Visual Studio Code
- **Libraries**: Flask, TensorFlow

Model Development:

- **Tool**: Jupyter Notebook
- **Libraries**: TensorFlow, NumPy, pandas

5.4. Tools and Techniques

The following tools and techniques were used:

Tools:

- **IDE**: Android Studio for Flutter, Visual Studio Code for Flask, Jupyter Notebook for model development
- **Version Control**: Managed locally with regular snapshots and backups

Techniques:

- **Agile Methodology**: To manage the project development in iterative cycles
- **Test-Driven Development (TDD)**: To ensure code reliability and functionality
- **Data Augmentation**: To enhance the training dataset and improve model performance

These tools and techniques collectively ensured an efficient and robust development process.

5.5. Best Practices / Coding Standards

The project adheres to these best practices and coding standards:

Coding Standards:

- **PEP 8**: For Python code style.
- **Dart Style Guide**: For Flutter development.

Best Practices:

- **Modular Code**: Maintaining a modular and well-structured codebase.
- **Documentation**: Providing comprehensive documentation for all functions and modules.
- **Code Review**: Conducting regular peer reviews to ensure code quality.
- **Security**: Ensuring secure handling of user data and predictions.

5.6. Version Control

The version control for this project is managed locally to ensure code integrity and proper documentation of changes. While the code has not been uploaded to a repository, meticulous practices have been followed to maintain version control.

- **Local Management**: The code is stored and managed locally.
- **Versioning Strategy**: Regular snapshots with version numbers track progress.
- **Commit Messages**: Clear and descriptive messages document changes.
- **Backups**: Multiple backups are maintained to prevent data loss.

By keeping detailed records and maintaining local backups, the project is well-managed even without a formal version control system.

Chapter 6

Testing and Evaluation

Chapter 6: Testing and Evaluation

This chapter describes the various testing and evaluation methodologies applied to ensure the reliability, accuracy, and performance of the breast cancer detection mobile application. Each section covers a specific type of testing, detailing the approach taken to validate functionality, identify potential issues, and ensure the system meets the required standards and performance metrics.

6.1. Use Case Testing

Use case testing validates the application against predefined scenarios representing typical user interactions. Each use case is tested to ensure the system behaves as expected, handling both normal and exceptional conditions.

6.2. Equivalence partitioning

Equivalence partitioning reduces the number of test cases by dividing input data into equivalent partitions. Each partition represents a set of inputs treated similarly by the system, ensuring efficient testing of all possible scenarios.

6.3. Boundary value analysis

Boundary value analysis tests the edges of input ranges. By checking the system's behavior at these boundaries, potential errors and vulnerabilities are identified and addressed.

6.4. Data flow testing

Data flow testing examines the paths data takes through the application, ensuring it is correctly processed, stored, and retrieved. This testing helps identify potential issues in data handling and integrity.

6.5. Unit testing

Unit testing involves testing individual components or functions of the application in isolation. This ensures that each part of the system performs correctly on its own before integration with other components.

6.6. Integration testing

Integration testing focuses on the interactions between different components of the application. It ensures that integrated components work together as intended and that data flows correctly between them.

6.7. Performance testing

Performance testing evaluates the system's responsiveness, stability, and scalability under various conditions. This includes measuring how quickly the system processes images and returns predictions, as well as how it handles multiple concurrent users.

6.8. Stress Testing

Stress testing pushes the system beyond its normal operational limits to evaluate its robustness and error-handling capabilities. This testing helps ensure that the application can handle extreme conditions without failing.

Chapter 7

Summary, Conclusion and Future Enhancements

Chapter 7: Summary, Conclusion & Future Enhancements

7.1. Project Summary

This project focused on creating a mobile application for breast cancer detection using Convolutional Neural Networks (CNNs). The application enables users to upload mammogram images, which undergo processing by a pre-trained CNN model to predict the presence of invasive ductal carcinoma (IDC). The model underwent training and validation using a substantial dataset of histopathology images, achieving high accuracy. The application integrates Flutter for the frontend and Flask for the backend, ensuring a user-friendly interface and strong processing capabilities.

7.2. Achievements and Improvements

The project successfully delivered a fully functional mobile application capable of accurately detecting invasive ductal carcinoma (IDC) in mammogram images. Key achievements include:

- Attaining a high model accuracy of 85.07% on the validation set.
- Successfully integrating advanced machine learning models into a mobile application.
- Implementing effective data preprocessing and augmentation techniques to significantly enhance model performance.
- Creating a robust and intuitive application interface that facilitates seamless image upload and retrieval of prediction results.

7.3. Critical Review

While the project successfully met its primary objectives, there are areas identified for potential improvement:

- Although achieving high accuracy, the model showed slight overfitting. Further refinement of dropout layers and data augmentation techniques could address this issue.
- The current application supports only mammogram images; broadening its scope to include other imaging modalities would increase its versatility.
- Incorporating user feedback mechanisms could enhance user experience and contribute valuable insights for ongoing model enhancement.

7.4. Lessons Learnt

Several significant insights were gained throughout the project:

- The significance of employing a balanced and diverse dataset for training robust machine learning models.
- The necessity of iterative testing and validation to ensure the model's ability to generalize and prevent overfitting.
- The critical role of integrating multiple technologies such as Flutter, Flask, and TensorFlow to create a cohesive and operational application.
- The importance of ongoing monitoring and updates to uphold application performance and user satisfaction.

7.5. Future Enhancements/Recommendations

Moving forward, future endeavors will concentrate on the following areas:

- Investigating advanced CNN architectures to elevate model accuracy and resilience.
- Incorporating supplementary preprocessing procedures and broadening the dataset to encompass various types of breast cancer images.
- Strengthening the application's functionality by integrating clinical histories and additional diagnostic data.
- Introducing sophisticated visualization tools to offer comprehensive and understandable prediction outcomes.
- Extending the application's scope to encompass other cancer types and medical conditions, leveraging the existing framework and methodologies.

Reference and Bibliography

Reference and Bibliography

[1] National Breast Cancer Foundation. (n.d.). Breast cancer facts. Retrieved June 25, 2024, from https://www.nationalbreastcancer.org/breast-cancer-facts/
[2] Pink Ribbon Pakistan. (n.d.). Breast health awareness. Retrieved June 25, 2024, from https://pinkribbon.org.pk/breast-health-awareness/
[3] American Cancer Society. (2022). *Breast cancer facts & figures 2022-2024*. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/2022-2024-breast-cancer-fact-figures-acs.pdf
[4] Abhisheka, B., Biswas, S. K., & Purkayastha, B. (2023). A Comprehensive Review on Breast Cancer Detection, Classification and Segmentation Using Deep Learning. *Archives of Computational Methods in Engineering*, *30*(8), 5023–5052. https://doi.org/10.1007/s11831-023-09968-z
[5] U. Sajid, R. A. Khan, S. M. Shah, and S. Arif, "Breast Cancer Classification using Deep Learned Features Boosted with Handcrafted Features," *Biomedical Signal Processing and Control*, vol. 86, p. 105353, Sep. 2023, doi: 10.1016/j.bspc.2023.105353.
[6] Narayanan, L., Krishnan, S., & Robinson, H. (2022). A Hybrid Deep Learning Based Assist System for Detection and Classification of Breast Cancer from Mammogram Images. *The International Arab Journal of Information Technology*, *19*(6). https://doi.org/10.34028/iajit/19/6/15
[7] A. Masood, U. Naseem, and J. Kim, "Multi-Level Swin Transformer Enabled Automatic Segmentation and Classification of Breast Metastases," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia: IEEE, Jul. 2023, pp. 1–4. doi: 10.1109/EMBC40787.2023.10340831.
[8] S. Kulkarni and R. Rabidas, "Fully convolutional network for automated detection and diagnosis of mammographic masses," *Multimed Tools Appl*, vol. 82, no. 29, pp. 44819–44840, Dec. 2023, doi: 10.1007/s11042-023-14757-8.
[9] Balasubramaniam, S., Velmurugan, Y., Jaganathan, D., & Dhanasekaran, S. (2023). A Modified LeNet CNN for Breast Cancer Diagnosis in Ultrasound Images. *Diagnostics*, *13*(17), 2746. https://doi.org/10.3390/diagnostics13172746.
[10] Z. Q. Habeeb, B. Vuksanovic, and I. Q. Al-Zaydi, "Breast Cancer Detection Using Image Processing and Machine Learning," *JOIG*, vol. 11, no. 1, pp. 1–8, Mar. 2023, doi: 10.18178/joig.11.1.1-8.
[11] Z. Jafari and E. Karami, "Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection," Computer Science and Mathematics, preprint, May 2023. doi: 10.20944/preprints202305.2209.v1.
[12] Pati, A., Parhi, M., Pattanayak, B. K., Singh, D., Singh, V., Kadry, S., Nam, Y., & Kang, B.-G. (2023). Breast Cancer Diagnosis Based on IoT and Deep Transfer Learning Enabled by Fog Computing. *Diagnostics*, *13*(13), 2191. https://doi.org/10.3390/diagnostics13132191

[13] Hu, H., Ye, R., Thiyagalingam, J., Coenen, F., & Su, J. (2023). Triple-kernel gated attention-based multiple instance learning with contrastive learning for medical image analysis. *Applied Intelligence*, *53*(17), 20311–20326. https://doi.org/10.1007/s10489-023-04458-y
[14] Zhu, Z., Wang, S.-H., & Zhang, Y.-D. (2023). A Survey of Convolutional Neural Network in Breast Cancer. *Computer Modeling in Engineering & Sciences*, *136*(3), 2127–2172. https://doi.org/10.32604/cmes.2023.025484
[15] Hadush, S., Girmay, Y., Sinamo, A., & Hagos, G. (2020). *Breast Cancer Detection Using Convolutional Neural Networks* (arXiv:2003.07911). arXiv. http://arxiv.org/abs/2003.07911
Mooney, P. T. (2018).
[16] Breast histopathology images [Data set].
Kaggle.https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images