Breast Cancer Prediction Using Machine Learning

Team Members

Kahlel Cardona kcardona2023@my.fit.edu

Woroma Dimkpa wdimkpa2021@my.fit.edu

Taratong Dolinsky tdolinsky2022@my.fit.edu

Faculty Advisor

Dr. Zahra Nematzadeh znematzadeh@fit.edu

Date of Meeting with the Client for Developing this Plan

1/15/2026

Goal and Motivation

The overall goal of the project is to develop a machine learning-based system that assists medical professionals and researchers by automatically classifying breast tumors as benign or malignant using mammogram images from the CBIS-DDSM dataset. By improving accuracy, consistency, and efficiency in tumor classification, the system aims to support earlier detection and reduce diagnostic burden, ultimately improving clinicians' and patients' confidence in screening outcomes.

Breast Cancer is one of the most prevalent cancers worldwide and remains a leading cause of cancer-related deaths among women. Early detection is critical, as it significantly increases the chances of successful treatment and long-term survival. Mammography is one of the most widely used screening methods; however, interpreting mammogram images is a challenging and time consuming task that requires significant expertise. Radiologists and clinicians often face issues such as image noise, subtle tumor patterns, and high workloads, which can lead to delayed diagnosis or misclassification of tumors.

Current systems and manual analysis methods are limited by human subjectivity, fatigue, and variability in experience. Additionally, traditional computer-aided diagnosis systems may rely on handcrafted features that fail to generalize well across diverse imaging data. Advances in deep learning, particularly Convolutional Neural Networks (CNNs), provide an opportunity to overcome these limitations by learning discriminative features directly from medical images. This project is motivated by the need for more reliable, scalable, and data-driven diagnostic support tools in healthcare.

Approach (Key Features of the System)

The proposed system will provide several key features designed to help achieve the overall project goal. These features focus on user-facing functionality and outcomes rather than implementation details.

- **Feature 1: Automated Mammogram Image Preprocessing Pipeline**

  A core feature of the system is the ability to automatically classify mammogram images as either benign or malignant. Users can input preprocessed mammogram images from the CBIS-DDSM dataset, and the system will apply a trained machine learning model to generate a classification result. This functionality reduces the need for fully manual image inspection and provides a consistent baseline for evaluating tumor characteristics.

- **Feature 2: Model Training and Comparison Framework**

  Another key feature of the system is its support for training and comparing multiple CNN architectures. Users can experiment with different models, such as ResNet, EfficientNet, ect, and evaluate how each architecture affects performance. This includes the use of transfer learning, where pretrained models are adapted to the breast cancer classification task.

- **Feature 3: Performance Evaluation and Result Visualization**

  The system also provides functionality for evaluating and analyzing model performance. Users can assess trained models using metrics such as accuracy and precision. These metrics help users understand not only how often the model is correct, but also how well it handles false positives and false negatives, which are critical considerations in medical diagnosis. In addition, users can analyze misclassified images to identify patterns or challenges within the dataset. This feature supports informed decision-making and encourages critical evaluation of machine learning outcomes rather than blind trust in predictions.

Novel Features and Functionalities

While CNN-based image classification is a well studied area, this project introduces novelty through systematic comparison of multiple deep learning architectures on the CBIS-DDSM dataset and in depth analysis of misclassified cases. The integration of transfer learning, augmentation strategies, and detailed performance visualization within a unified experimental framework allows for insights beyond basic classification. Additionally the project emphasizes interpretability and evaluation rather than treating the model as a non-transparent system.

Algorithms and Tools

Potentially useful algorithms and tools include:

- **Algorithms**
  - Convolutional Neural Networks
  - Transfer Learning with pretrained models
  - Binary Classification Techniques
- **Software Tools**
  - Python
  - PyTorch or TensorFlow/Keras
  - NumPy, Pandas
  - Matplotlib / Seaborn for visualization
  - Scikit-learn for evaluation metrics
- **Dataset**
  - CBIS-DDSM mammogram image dataset

Technical Challenges

Limited Experience with Medical Image Data: The team has limited prior experience working with medical imaging datasets, which often contain noise, artifacts, and large image sizes. Understanding appropriate preprocessing techniques will be a key challenge.

Training Deep Learning Models Efficiently: CNNs require significant computational resources and careful hyperparameter tuning. Selecting appropriate architectures and avoiding overfitting will be challenging, especially given dataset size constraints.

Model Evaluation and Interpretation: Beyond achieving high accuracy, interpreting misclassifications and understanding model weaknesses is nontrivial. Designing meaningful evaluation and visualization strategies will require careful planning and experimentation.

Milestone 1 (Feb 23): Planning and Setup

- Compare and select technical tools for:
    - Image preprocessing
    - Model training framework
    - Visualization and evaluation
- Develop small “hello world” demos:
    - Load and display sample CBIS-DDSM images
    - Apply basic preprocessing and augmentation
    - Compute preliminary evaluation metrics
    - Train a simple CNN on a small data subset
- Resolve initial technical challenges:
    - Dataset loading and preprocessing
    - Basic CNN implementation
- Compare and select collaboration tools for:
    - Software Development
    - Documentation and presentations
    - Communication and task scheduling
- Create:
    - Requirements Document
    - Design Document
    - Test Plan

Milestone 2 (Mar 30): Core Implementation

- Implement and test full preprocessing pipeline
- Implement and train an initial CNN model
- Implement transfer learning using a pretrained architecture
- Evaluate and compare initial models
- Demo implemented features and intermediate results

Milestone 3 (Apr 20): Refinement and Analysis

- Implement additional CNN architectures
- Fine-tune hyperparameters and augmentation strategies
- Perform detailed evaluation and visualization
- Analyze misclassified images
- Finalize documentation and results.

Task Matrix for Milestone 1

Task	Kahlel	Woroma	Tara
Compare & Select Technical Tools	Data	Models	Visualization
“Hello World” Demos	Preprocessing	CNN Training	Metrics
Resolve Technical Challenges	Dataset	Architecture	Evaluation
Requirements Document	50%	25%	25%
Design Document	25%	25%	50%
Test Plan	25%	50%	25%

1. Approval from Faculty Advisor
    - "I have discussed with the team and approved this project plan. I will evaluate the progress and assign a grade for each of the three milestones."
    - Signature: _____ Date: _____