Contents

Project Title:

Medical Diagnostic System for Breast Cancer Detection

Group Members:

1. Mohammed Adil
2. Ali Soltan Dileyta
3. Kena Teshome
4. Dagim Faji
5. Mahdi Abdi Rayaleh
6. Impundu Joyeux Thevenin

1. Project Overview

Breast cancer remains the most common cancer among women worldwide, with over 2.3 million new cases diagnosed in 2020 alone, according to the World Health Organization (WHO). Despite advances in medical technology and treatments, breast cancer continues to be a leading cause of death among women globally. Early detection is critical in improving survival rates, as it allows for timely treatment and intervention.

The **Medical Diagnostic System for Breast Cancer Detection** project aims to leverage the power of machine learning (ML) and deep learning (DL) to create a robust, accurate, and accessible diagnostic tool. This tool will primarily target healthcare providers but can also be adapted for use by patients in resource-limited settings where access to advanced medical diagnostics is not readily available.

The project utilizes Convolutional Neural Networks (CNNs) to analyze mammography images and determine whether they indicate benign or malignant conditions. In parallel, a chatbot guides users through a series of questions based on recognized breast cancer symptoms. This dual approach ensures that the system can provide both image-based diagnostics and symptom-based assessments, making it a comprehensive tool for early detection.

Given the widespread impact of breast cancer, this project aligns with Sustainable Development Goal 3 (Good Health and Well-Being), aiming to reduce mortality rates through early detection and accessible diagnostic tools. By integrating advanced ML/DL techniques with user-friendly interfaces, the project aspires to contribute significantly to global breast cancer detection efforts.

2. Objectives

1. Development of an Accurate Diagnostic Tool

- **Image Analysis**: Utilize a Convolutional Neural Network (CNN) trained on large datasets of mammography images to classify breast tissue as benign or malignant. The CNN will be fine-tuned for high accuracy, with continuous model evaluation and retraining as more data becomes available.
- **Symptom Analysis**: Develop a chatbot that interacts with users, asking a series of symptom-related questions based on established medical guidelines. The chatbot will analyze the responses to provide a preliminary risk assessment.

2. Integration of Image and Symptom-Based Diagnostics

- Combine the image analysis from the CNN and the symptom analysis from the chatbot into a unified diagnostic tool. This will allow users to upload mammography images and answer diagnostic questions on a single platform, providing a comprehensive assessment.

3. Deployment as a Web-Based Application

- Develop a user-friendly web application using React and Django, ensuring that the tool is accessible to healthcare providers and potentially patients. The web app will feature a modern, responsive design, making it accessible across various devices, including smartphones and tablets.

4. Alignment with WHO Standards

- Ensure that the diagnostic processes align with WHO standards and guidelines for breast cancer detection. Continuous consultation with medical experts will refine the diagnostic criteria and ensure clinical relevance.

5. Enhancement of Global Breast Cancer Detection Efforts

- Contribute to global breast cancer detection initiatives by providing a tool that is both scalable and adaptable to different healthcare settings, particularly in low-resource environments.

3. Literature Review

The literature review focuses on various machine learning algorithms used for breast cancer prediction and diagnosis, as well as the effectiveness of deep learning models in processing medical images. Previous research has demonstrated the potential of ML and DL models in improving diagnostic accuracy, particularly in distinguishing between benign and malignant tumors.

Key references include:

- **Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis**
- **Prediction of Breast Cancer using Machine Learning Approaches**

These studies provide a strong foundation for the development of our diagnostic tool, highlighting the capabilities and limitations of different algorithms.

4. Datasets

The project utilizes two primary datasets:

1. **The rsna-mammography-768-vl-perlabel:** from Kaggle, which provides an image of mammography data labeled as 0(for malignant) and malignant

2. **CBIS-DDSM: Mass Case Mammograms PNG Dataset:** also, from Kaggle, containing labeled images of mammography image samples.

These datasets are crucial for training the machine learning or deep learning models used in this project.

5. Methodology

1. Convolutional Neural Networks (CNNs) for Image Classification

Convolutional Neural Networks (CNNs) are a class of deep learning models particularly effective for image processing tasks. They are designed to automatically and adaptively learn spatial hierarchies of features from input images. CNNs have been widely used in medical imaging, where they have proven to be highly effective in tasks like tumor detection, segmentation, and classification.

In this project, the CNN is tasked with classifying mammography images into two categories: benign or malignant. The architecture of the CNN includes multiple layers:

- **Convolutional Layers**: These layers apply convolution operations to the input image, creating feature maps that highlight various aspects of the image, such as edges, textures, and shapes. Each convolutional layer is followed by an activation function (typically ReLU) that introduces non-linearity, allowing the model to learn complex patterns.
- **Pooling Layers**: Pooling layers are used to reduce the spatial dimensions of the feature maps, making the network more computationally efficient and reducing the likelihood of overfitting. Max pooling is commonly used, which selects the maximum value from each feature map region.
- **Fully Connected Layers**: After several convolutional and pooling layers, the network flattens the feature maps and passes them through fully connected layers. These layers make predictions based on the learned features, outputting the probability that an image belongs to each class (benign or malignant).
- **Softmax Layer**: The final layer is a softmax layer, which converts the output probabilities into a classification decision.

The model is trained on a large dataset of labeled mammography images, ensuring that it learns to accurately distinguish between benign and malignant cases. Continuous evaluation and retraining will be conducted to improve the model's accuracy and generalization capabilities.

2. Chatbot for Symptom-Based Risk Assessment

The chatbot is designed to interact with users, guiding them through a series of questions related to breast cancer symptoms. The questions are based on well-established medical guidelines, ensuring that the chatbot's assessments are clinically relevant.

Functionality:

- **Questionnaire Design**: The chatbot asks users about common symptoms of breast cancer, such as the presence of lumps, changes in breast size, skin alterations, and family history of the disease. The questions are designed to be easy to understand and answer, with users selecting "Yes" or "No" responses.
- **Dynamic Interaction**: The chatbot is designed to emulate a conversation, with each question appearing as a chat bubble on the left side of the screen and answer buttons on the right. This design ensures that the interaction feels natural and engaging.
- **Response Analysis**: Based on the user's answers, the chatbot calculates a risk score. The logic behind the risk assessment is based on three main criteria:
    1. If the user answers "Yes" to 60% or more of the major symptom questions, they are considered at high risk.
    2. If 75% or more of all questions are answered with "Yes," the risk level is also considered high.
    3. If all answers are "Yes," the chatbot flags the case as very high risk and advises immediate medical consultation.
- **Final Assessment**: Once all questions have been answered, the chatbot provides a summary of the risk levels (low, moderate, high) based on the responses. This summary is displayed on the screen, with recommendations for further action if necessary.

3. System Integration

The system combines these two components—a CNN for image classification and a chatbot for symptom analysis—into a single web application. The web app is developed using React for the frontend and Django for the backend, providing a seamless user experience. The integration ensures that users can easily switch between the chatbot and the image predictor, making the diagnostic process comprehensive and user-friendly.

6. Architecture Design Diagram

The architecture includes:

1. **Frontend**: A React-based user interface styled with modern CSS frameworks like Bootstrap.
2. **Backend**: A Django-based server that handles API requests, processes images, and manages the chatbot's logic.
3. **Machine Learning Model**: A TensorFlow-based CNN model trained on mammography images for accurate classification.

7. Implementation Plan

1) Technology Stack

- **Frontend**: React, Bootstrap, JavaScript, CSS
- **Backend**: Django, Python, Django REST Framework
- **Machine Learning**: TensorFlow, Keras, Python
- **Database**: SQLite (for development)

2) Timeline

- **Week 1**: Data collection and preprocessing
- **Week 2**: Model development and training
- **Week 3**: Integration of model with backend
- **Week 4**: Testing and validation
- **Week 5**: Development of chatbot and frontend interface
-
- **Week 6**: Deployment and final adjustments

3) Milestones

- **Milestone 1**: Completion of the CNN model for image classification.
- **Milestone 2**: Successful integration of the chatbot.
- **Milestone 3**: Finalization of the web application with a functional frontend.
- **Milestone 4**: Testing and Improvement of the model to get higher accuracy.
- **Milestone 5**: Deployment of the complete system.

4) Challenges and Mitigations

- **Data Quality**: Ensuring high-quality, correctly labeled data for training. Mitigation involves rigorous data cleaning and validation.

- **Model Performance**: The challenge of achieving high accuracy. Regular evaluation and tuning of hyperparameters will be employed.
- **Technical Constraints**: Managing the computational resources needed for training the CNN. This will be mitigated by using cloud resources if necessary.

8. Ethical Considerations

Given the sensitive nature of health data, strict adherence to data privacy laws and ethical standards is paramount. The system will ensure that all patient data is anonymized and that the tool complies with healthcare regulations to prevent biases in diagnosis.

9. References

- **Kaggle Datasets:**
  - ***The rsna-mammography-768-vl-perlabel***
  - ***CBIS-DDSM: Mass Case Mammograms PNG Dataset***
- **Research Articles:**
  - ***Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis***
  - ***Prediction of Breast Cancer using Machine Learning Approaches***