Project and Data Management Plan

Project Overview

Project Title: Breast Cancer Classification Using Mammogram Images for Enhanced Early Detection

Summary of Project and Background

This project explores a **hybrid model approach** that combines **DenseNet121** and **ResNet50** for classifying mammogram images to detect breast cancer early. Using a publicly available dataset, the merged model will differentiate between benign and malignant tumors. Additionally**, Grad-CAM** (Gradient-weighted Class Activation Mapping) will be used to provide visual explanations for the model's predictions.

Breast cancer is one of the leading causes of mortality among women globally. Mammography is crucial for early detection; however, image interpretation can be inconsistent. Deep learning, particularly Convolutional Neural Networks (CNNs), offers a promising approach to improve diagnostic accuracy. This project aims to leverage CNNs for mammogram classification, distinguishing between benign and malignant tumors. Studies have demonstrated the potential of AI to aid radiologists in minimizing diagnostic errors and improving early detection (Yap et al., 2018; Elter & Schulz-Wendtland, 2020; Agarwal et al., 2022).

Research Question: How can transfer learning models, combined with model fusion and explainable AI techniques like Grad-CAM, be used to improve the accuracy and interpretability of mammogram classification (benign vs malignant) for deployment in resource-constrained clinical environments?

Project Objectives:

- **Data Preprocessing & Augmentation:** Enhance performance and reduce overfitting.
- **ResNet and DenseNet Model Development:** Classify mammogram images as benign or malignant using ResNet50 and DenseNet121 architectures for deep learning.
- **Performance Evaluation:** Measure accuracy, precision, recall, and F1-score.
- **Grad-CAM Implementation**: Visualize model decisions and attention areas in mammogram images.
- **Comparison:** Benchmark against state-of-the-art classification techniques.

Project Plan Timeline

Task	Start Date	End Date	Description
Project Proposal	25th Sept 2024	2nd Oct 2024	Project selection and submitting the project proposal.
Data Collection	2nd Oct 2024	10th Oct 2024	Collection of the relevant data.
Project Development Plan	30TH Oct 2024	6th Nov 2024	Development of a detailed PDM outlining the objectives, background, timeline and dataset.
Literature Review	6th Nov 2024	11th Nov 2024	Conduct a literature review on deep learning in breast cancer detection and gather data from relevant databases .
Data Acquisition and Preprocessing	11th Nov 2024	20th Nov 2024	Preprocess the dataset through normalization, resizing, and augmentation.
Model Designing	21st Nov 2024	28th Nov 2024	Design a CNN architecture by selecting appropriate layers and setting parameters specifically for a classification task.
Model Training	29thNov 2024	5th Dec 2024	Implement the CNN architecture using frameworks like TensorFlow or PyTorch. Fine- tune hyperparameters (learning rate, batch size, epochs) to optimize model performance.
Model Evaluation and implement Grad-CAM	6th Dec 2024	20th Dec 2024	Test the trained model on validation data to evaluate performance. Implement Grad-CAM to visualize areas of attention in mammogram images and understand model predictions better.
Model Testing	21st Dec 2024	27th Dec 2024	Conduct final testing on the test dataset to ensure model reliability.
Report Writing	28th Dec 2024	6th Jan 2025	Submit the final report, summarize key findings, methodologies, and results.
Report review and changes	8th Jan 2025	24th Jan 2025	Review document with supervisor feedback, make necessary changes and finalize the report.

Data Management Plan

Overview and data collection: The Breast Mammography Dataset with Masses contains a collection of mammography images featuring both benign and malignant masses. The images were gathered from several established sources: 106 images from the INbreast dataset, 53 from the MIAS dataset, and 2,188 from the DDSM dataset.

After data augmentation, the INbreast dataset contains 7,632 images, the MIAS dataset has 3,816 images, and the DDSM dataset includes 13,128 images. Furthermore, the Inbreast, MIAS, and DDSM datasets have been combined.

The dataset used for this project is publicly available on data Mendeley, Published on 3 July 2020, Version 5. The URL is: https://data.mendeley.com/datasets/ywsbh3ndr8/5.The dataset integrates images from three major mammography databases and is organized by tumor mass type (benign or malignant) and breast density categories.

Metadata

- File Format: JPEG
- Size: Contains over 24,000 images (After Augmentation)
- Includes: Class labels (benign/malignant), breast density levels, pre- and post-augmentation images.

Document Control and Readme: The code will be managed through GitHub, where I have created a repository that includes a README file specifically for this dissertation. Here is accessible link of GitHub https://github.com/am22ajn/FINAL-PROJECT.git.

Security: Access will be restricted to authorized personnel only.

Backup: I will use a Google Drive for taking backup all my project files and share them with my supervisor.

Ethical Considerations

- The dataset complies with privacy laws like **GDPR** as it does not include any personally identifiable information (PII).
- All research using the dataset should align with the ethical guidelines of the contributing institutions.
- The dataset is licensed under **CC BY 4.0**.

References

1. Agarwal, A., Gupta, M. & Khanna, P. (2022). 'Artificial Intelligence in Breast Cancer Imaging: Early Detection and Future Directions', *Journal of Breast Cancer Research and Treatment*, 188(1), pp. 45-56. Available at: https://link.springer.com/article/10.1007/s10549-022-06609-5.

2. Elter, M. & Schulz-Wendtland, R. (2020). 'AI-Assisted Mammography: Reducing Diagnostic Errors with Deep Learning', *International Journal of Medical Imaging*, 65(3), pp. 105-113. Available at: https://www.sciencedirect.com/science/article/pii/S0899700719304758.

3. Yap, M. H., Goyal, M., Osman, S., Mart, C. & Denton, E. (2018). 'Breast Ultrasound Image Segmentation using Convolutional Neural Networks', *Journal of Imaging*, 4(12), pp. 134-145. Available at: https://www.mdpi.com/2076-3417/10/22/8298

4. Ahn, J. S., Shin, S., Yang, S.-A., Park, E. K., Kim, K. H., Cho, S. I., Ock, C.-Y., & Kim, S. (2023). 'Artificial Intelligence in Breast Cancer Diagnosis and Personalized Medicine'. Cancers, 15(1), Article 10625863. PMCID: PMC10625863. PMID: 37926067.Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10625863/