**Zeroth Review**

**On**

**Early Detection and Classification of Breast Cancer from Mammograms**

**Submitted by:**

**Laxmi Sai. G (21MIS7119)**

**Vasudha Rani. P (21MIS7121)**

**To**

**Dr. B.V. Gokulnath**

**School Of Computer Science And Engineering (SCOPE)**

****

Beside AP Secretariat, Near Vijayawada, Amaravati, Andhra Pradesh, India, 522237

|  |  |  |
| --- | --- | --- |
| **1.** | **Title** | **Early Detection and Classification of Breast Cancer from Mammograms** |
| **2.** | **Software Requirements** | **Operating System:** Windows XP or Linux  **Tools:** Google Colab or Jupyter Notebook  **Database Server**: Local Data Storage  **Deep Learning Libraries:** ViT, custom CNN, ANN classifier  **Visualization tools:** SeaBorne, Matplotlib |
| **3.** | **Hardware Requirements** | **GPU:** NVIDIA RTX 3080, T4GPU  **RAM:** 32 GB  **Storage:** 1 TB NVMe SSD  **Additional:** High-speed internet connection |

**ABSTRACT:**

Early detection of breast cancer detection is a good solution for successful treatment. The deep learning model we have uses a two-stage approach which uses Visual Transformers (ViT) and Convolution Neural Networks (CNN). In the first step we use ViT which identifies the abnormalities of the mammogram in big pictures. The important regions which have suspicion will then be zoomed on a custom CNN which extracts the detailed features specific to those areas. Combining these both creates a broad analysis allows for precise differentiation between benign and malignant masses. This leverages the strengths of both architectures.

**MODULES INVOLVED:**

* Data Collection
* Data pre-processing
* Model Training
* Cancer Cells Detection
* Visualization of the project

1. **Data Collection:**

In the first step we have a comprehensive dataset of mammogram images. The dataset includes a balanced representation of both benign and malignant cases. We are exporting the dataset from the Kaggle

<https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset>

1. **Data pre-processing:**

**Normalization:** Scaling pixel intensities to a common range for better training convergence.

**Resizing:** Resizing all images to a uniform dimension compatible with the model's input requirements.

**Data augmentation:** Artificially generating variations of existing images (e.g., rotations, flips) to increase dataset size and improve model robustness.

1. **Model Training:**

* **Network Architecture:** We'll build a two-stage deep learning model using:
  + **Pre-trained Vision Transformer (ViT):** This pre-trained model extracts global features from the entire mammogram image.
  + **Custom Convolutional Neural Network (CNN):** This custom CNN takes the output from the ViT and focuses on suspicious regions identified by potential abnormalities. It then extracts detailed features specific to those areas.
* **Training Process:** We'll split the dataset into training, validation, and testing sets. The training set trains the model, the validation set helps us fine-tune hyperparameters during training, and the testing set evaluates the model's ability to perform on unseen data.
* **Optimization:** We'll use an optimizer like Adam to update the model weights based on a loss function (e.g., cross-entropy) during training.

1. **Breast Cancer Detection:**

* **New Mammogram Analysis:** When presented with a new mammogram image:
  + The ViT analyzes the entire image, identifying potential areas of concern.
  + The CNN zooms in on these suspicious regions, extracting detailed features.
* **Classification:** Based on the combined features extracted by the ViT and CNN, the model directly classifies the masses within the image as benign or malignant.

1. **Visualization of the Project:**

Effective communication of the project's findings is crucial. Visualization techniques can be employed to:

* Display examples of correctly and incorrectly classified mammograms.
* Generate heatmaps highlighting the regions of interest identified by the model.
* Utilize techniques like Grad-CAM to understand which features in the image contribute most to the model's classification decision.