BREAST CANCER CLASSIFICATION AND DETECTION THROUGH ADVANCED PRE-PROCESSING AND DEEP LEARNING TECHNIQUES

Group ID: CSE 25-55

Guide: Mr. Gagan Thakral

Team Members:

Sajal Bhilatia

Rachit Verma

Harsh Rastogi

COURSE OUTCOMES (COS)

✓**CO2:**

Ability to implement image preprocessing techniques for enhancing and preparing medical images for deep learning models.

✓**CO3:**

Capability to develop and evaluate deep learning models for classification tasks using appropriate performance metrics.

PROGRAM OUTCOMES (POS)

✓ **PO2:**

Ability to identify, analyze, and solve complex engineering problems using first principles of mathematics and engineering sciences.

✓**PO5:**

Proficiency in using modern tools and techniques, such as machine learning frameworks and image processing libraries, for engineering applications.

SUSTAINABLE DEVELOPMENT GOALS

✓ **SDG 3:**
Good Health and Well-being – Early cancer detection improves survival.

✓ **SDG 9:**
Industry, Innovation, and Infrastructure – Promotes use of AI in healthcare.

✓ **SDG 10:**
Reduced Inequality – Supports access to diagnostic support in rural areas.

PROJECT ABSTRACTION

- Aim: To design a system for early detection and classification of breast cancer using mammographic images.
- Integrates image preprocessing with deep learning-based classification.
- Enhances diagnostic reliability, reduces false negatives.
- Supports radiologists in clinical decision-making.

OVERVIEW

- Automates classification of mammograms as benign or malignant.
- Uses preprocessing techniques like CLAHE, edge detection, pectoral muscle removal.
- Trains and compares CNN models: VGG16, ResNet50, MobileNet, and Custom CNN.
- Improves diagnostic speed, precision, and recall.

KEY OBJECTIVES

- Improve image quality and remove misleading features (e.g., pectoral muscles).
- Apply deep learning to classify tumors.
- Compare model performances to choose the most balanced one.
- Enable scalable and accurate early diagnosis.

METHODOLOGY

- **Dataset:** Publicly available mammogram datasets (DDSM, INbreast).
- **Preprocessing:** Orientation correction, CLAHE, pectoral muscle removal, morphological operations.
- **Model Training:** CNN-based models trained on binary classification task.
- **Evaluation:** Precision, recall, accuracy, F1-score, confusion matrix.

MODEL ARCHITECTURE

- **VGG16**: Deep network, good feature extraction, prone to overfitting.
- **ResNet50**: Uses residual blocks to train deeper networks.
- **MobileNet**: Lightweight and efficient for real-time use.
- **Custom CNN**: Built for control and experimentation.

EVALUATION MATRIX

MODEL	TRAINING ACCURACY	TEST ACCURACY	PRECISION (TEST)	RECALL (TEST)	F1-SCORE (TEST)
VGG16	99.78%	62.26%	63.71%	62.26%	61.50%
ResNet50	97.44%	50.94%	51.16%	50.94%	50.98%
MobileNet	89.96%	60.38%	61.75%	60.38%	59.98%
Custom CNN	99.57%	50.94%	50.74%	50.94%	50.77%

EXPECTED OUTCOMES

- Successfully developed a breast cancer classification system.
- Achieved high accuracy and model reliability.
- MobileNet identified as best performer.
- Developed a reusable, modular pipeline.

TESTING AND REFINEMENT

- 80:20 train-test split.
- Data augmentation used to prevent overfitting.
- MobileNet showed the best generalization performance.
- Refinement focused on enhancing preprocessing and model tuning.

FUTURE SCOPE

- Integrate real-time classification via web/mobile UI.
- Expand model to other medical imaging types (MRI, ultrasound).
- Use explainable AI (e.g., Grad-CAM) for visual justification.
- Train on larger datasets to enhance robustness.

KIET GROUP OF INSTITUTIONS
Connecting Life with Learning
www.kiet.edu
Delhi-NCR, Ghaziabad

THANK YOU