Breast Cancer Machine Learning

Woroma Dimkpa
Kahlel Cardona

Faculty Advisor: Dr. Zahra Nematzadeh

Table of Contents

Goals

- Train a machine learning model (i.e CNN with transfer learning) on the CBIS-DDSM dataset
- Evaluate model performance using:
  - Accuracy
  - Precision
  - Recall
  - F1-score
  - auc
- Compare different model architectures and hyperparameters
- Document results, analyze findings, and discuss limitations

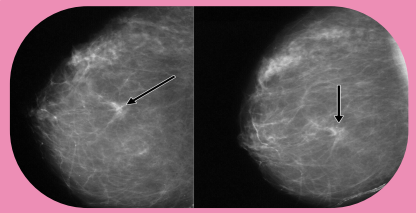
Motivation

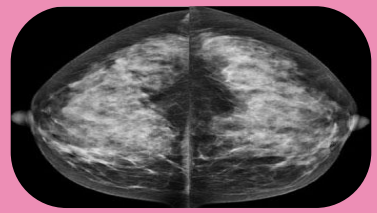
- Breast cancer is one of the most prevalent cancers worldwide and a leading cause of cancer-related deaths among women.
- Early detection can significantly improve survival rates
- Current technology includes some limitations such as
  - reliance on handcrafted features
  - limited generalization across datasets
  - sensitivity to image noise and variability

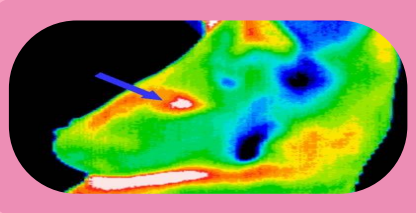
Approach (Key Features)

Feature 1: Preprocessing Pipeline

- Input raw CBIS-DDSM images
- Automatic:
  - Resizing
  - Normalization
  - Augmentation
- Consistent, reproducible preprocessing



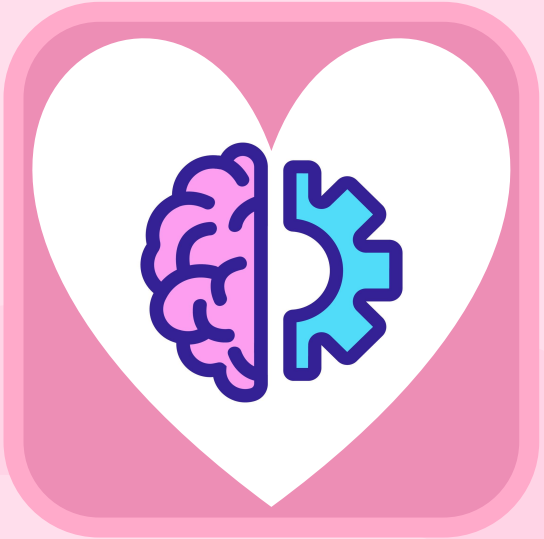




Approach (Key Features) Cont.

Feature 2: CNN-Based Classification

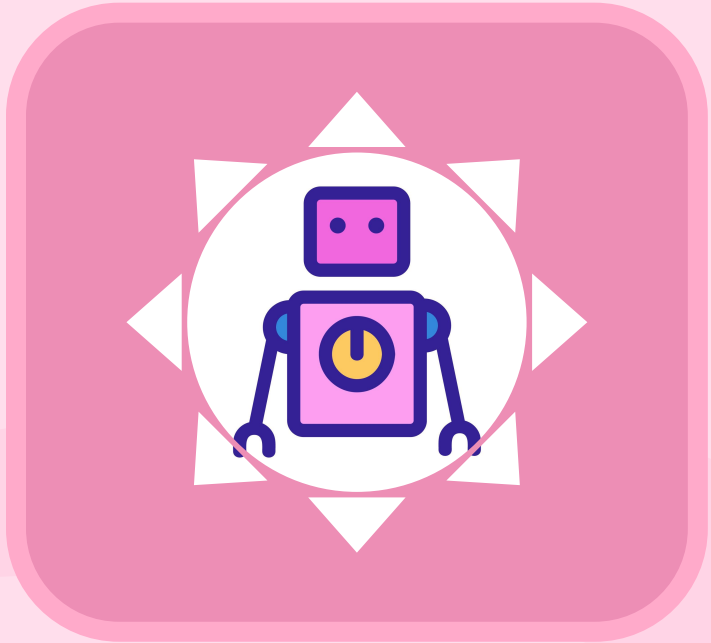
- CNN Models (e.g. ResNet, EfficientNet)
- Transfer Learning
- Output:
  - Benign / Malignant label
  - Confidence score



Approach (Key Features) Cont.

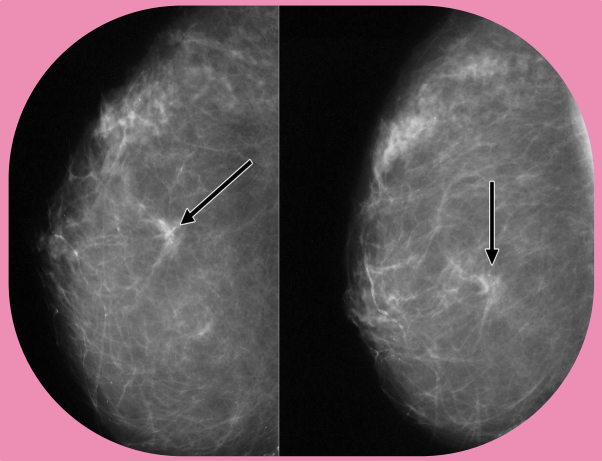
Feature 3: Evaluation & Visualization

- **Metrics:**
  - **Accuracy**
  - **Precision**
  - **Recall**
  - **F1-score**
  - **AUC**
- **Visualizations:**
  - **Confusion Matrix**
  - **ROC Curve**



Functionality

- Upload mammogram images
- Train & evaluate models
- Compare architectures
- View metrics and visual results
- Inspect misclassified cases

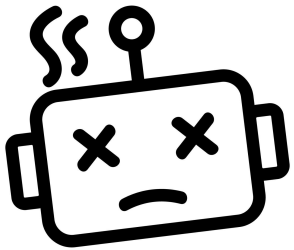


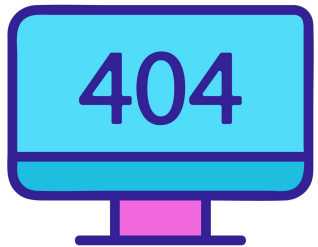
Algorithms & Tools

- **Algorithms**
  - Convolutional Neural Networks
  - Transfer Learning
  - Binary Classification
- **Tools**
  - Python
  - PyTorch or TensorFlow
  - Scikit-learn
  - Matplotlib
- **Dataset**
  - CBIS-DDSM mammogram image dataset

Technical Challenges

- **Limited Experience with Medical Image Data**
- **CNN Training & Tuning**
- **Model Evaluation & Interpretation**





Milestone 1 (Planning & Setup)

- **Compare and select technical tools**
- **Develop small demos**
- **Resolve initial technical challenges**
- **Create Necessary documents**

Milestone 2 (Core Implementation)

- **Implement and test full preprocessing pipeline**
- **Implement and train an initial CNN model**
- **Implement transfer learning using a pre trained architecture**
- **Evaluate and compare initial mode**

Milestone 3 (Refinement & Analysis)

- **Implement additional CNN architectures**
- **Fine-tune parameters and augmentation strategies**
- **Performa detailed evaluation using visuals**
- **Create a diagram which showcases each milestone and its results**

Thank you

Questions?