HOUSTON
CITY COLLEGE

Breast Ultrasound Lesion Classification (BUSI Dataset)

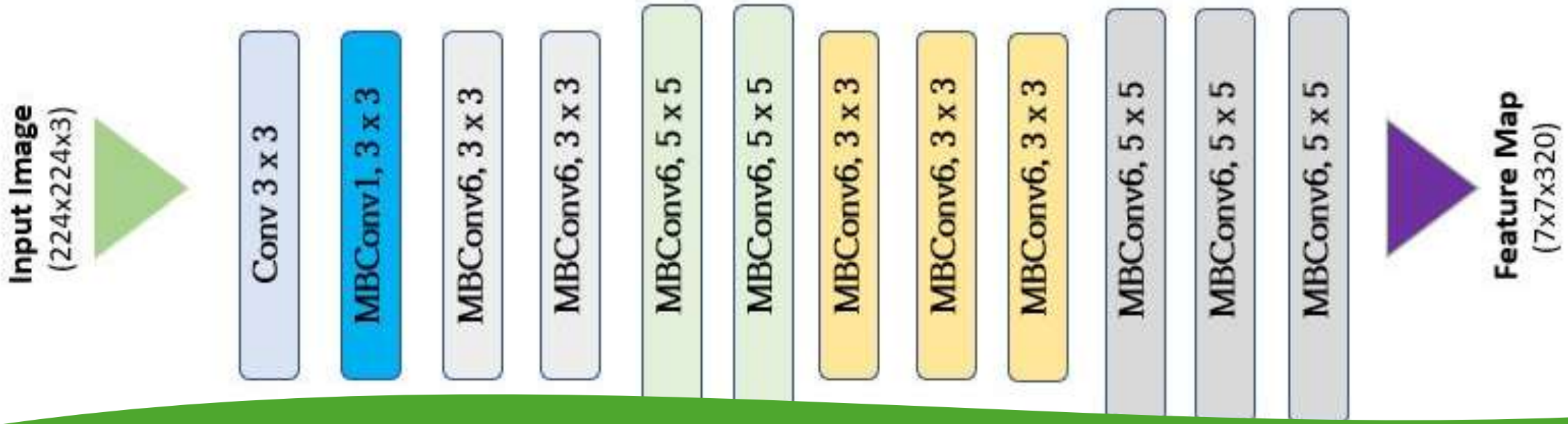
Student: **José Tuozzo**
Course: ITAI 1378 — Computer Vision & AI
Tier: 2 — Applied Medical CV Project

The Problem

- Breast cancer is one of the leading causes of death in women
- Early detection greatly improves treatment outcomes
- Ultrasound is common but interpretation varies across clinicians
- Small tumors can be easy to miss in manual review
- **Need:** A simple tool to support early triage using ultrasound images

Proposed Solution

- Develop a **deep learning model** that:
- Classifies breast ultrasound images as:
  Benign vs Malignant
- Uses **transfer learning** for efficiency
- Generates **Grad-CAM heatmaps** to show important regions
- Supports clinicians (NOT a diagnostic tool)
- → *Goal: Assist doctors, not replace them*



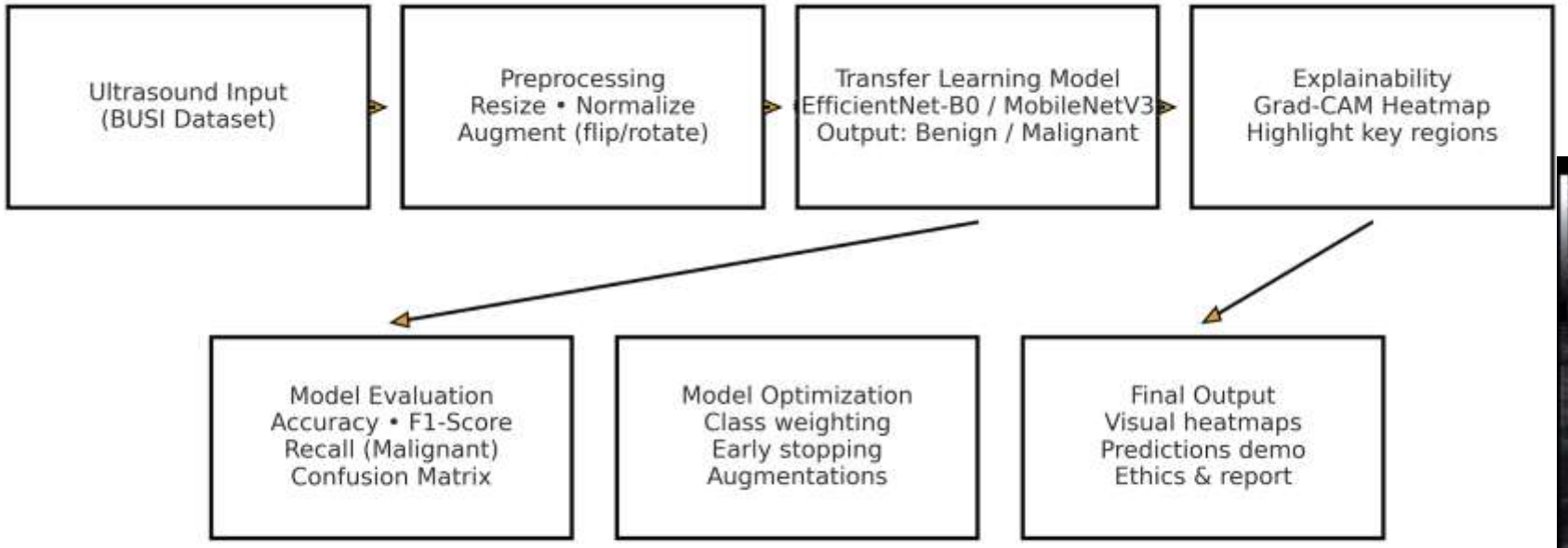
Technical Approach

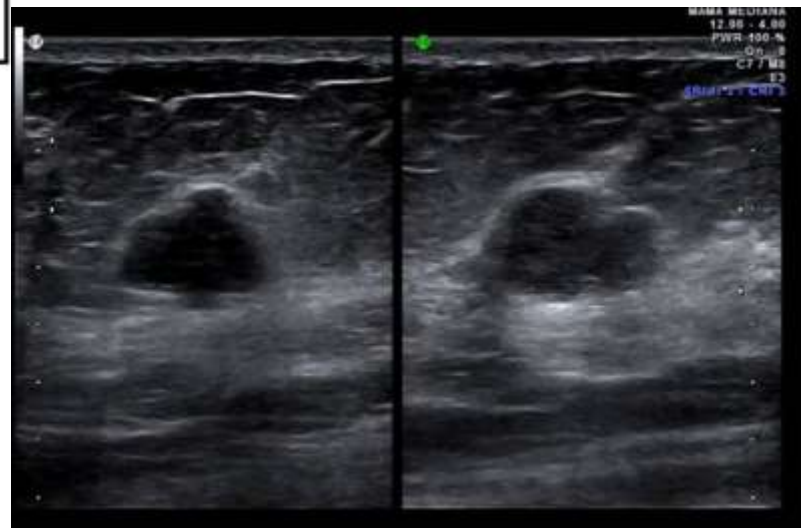
- **Method:** Transfer learning + fine-tuning
  Models: **Framework:** PyTorch + Torchvision
  Training:
- Train/Validation/Test split: **70% / 15% / 15%**
- Early stopping + data augmentation
- Class weighting for imbalance
- →□ Focus on interpretability + lightweight training

Dataset Plan

- **Dataset:** BUSI – Breast Ultrasound Images (Kaggle)
- ~780 images
- Classes: Normal / Benign / Malignant
- *For simplicity: Use Benign vs Malignant only*
- **Preprocessing:**
- Resize to 224×224
- Normalize
- Augment (flip, rotate, brightness)
- ⚠ **No patient data stored in repo**

System Diagram





Success Metrics

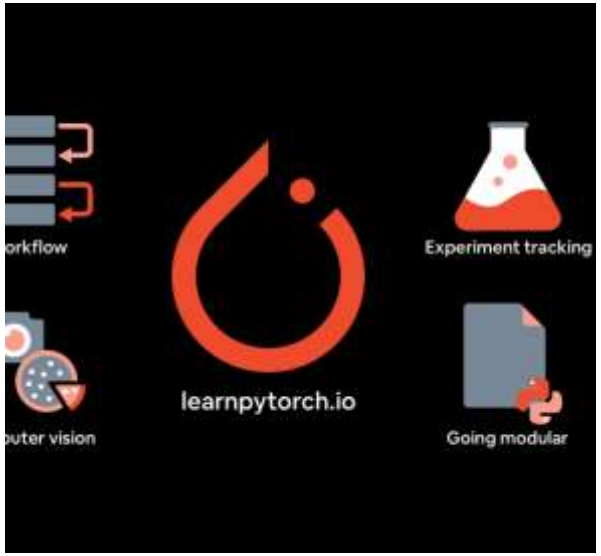
Metric	Goal
Accuracy	≥ 80%
F1-Score	≥ 0.82
Recall (Malignant)	≥ 0.85 *(priority)*
Explainability	Grad-CAM maps for predictions

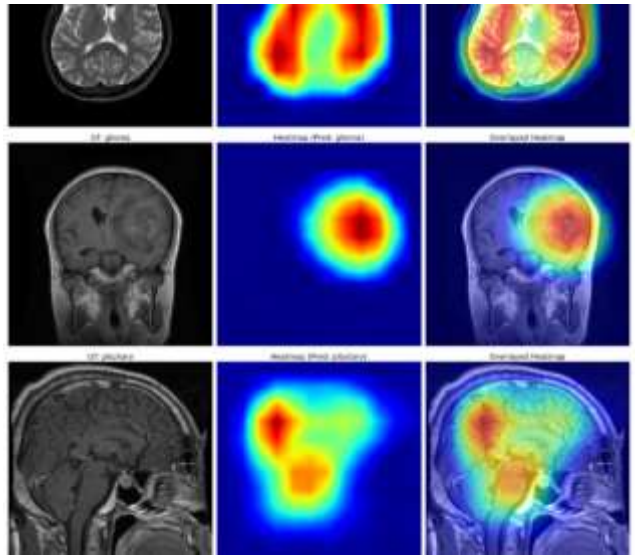
Timeline

Week	Task
Week 10	Data setup + repo structure
Week 11	Train baseline model
Week 12	Improve model + Grad-CAM
Week 13	Evaluation + confusion matrix
Week 14	Documentation + demo recording
Week 15	Final presentation

Risks & Mitigation

Risk	Solution
Small dataset	Augmentation + transfer learning
Class imbalance	Weighted loss / oversampling
Overfitting	Early stopping + dropout
Limited compute	EfficientNet-B0 on Colab GPU





Resources

- **Tools:**
- Google Colab
- PyTorch / Torchvision
- Matplotlib / OpenCV
- Grad-CAM package
- **Cost:** $0
- **Dataset:** BUSI (Kaggle)
- **Note:** Educational prototype – NOT for clinical diagnosis

References

- BUSI dataset (Al-Dhabyani et al., 2020)
- Grad-CAM (Selvaraju et al., 2017)
- Transfer learning (PyTorch, 2023)
- EfficientNet (Tan & Le, 2019)
- MobileNetV3 (Howard et al., 2019)
- American Cancer Society (2024) – background info