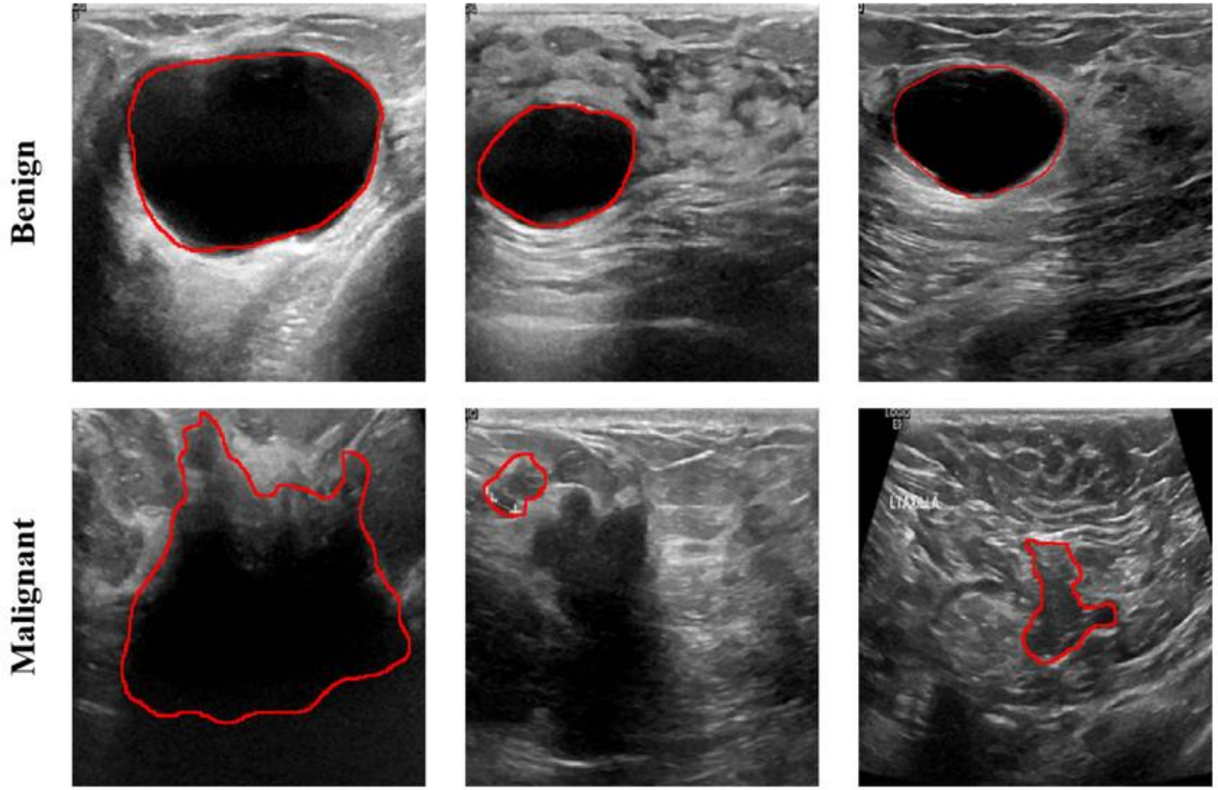
Deep Learning-Based Breast Ultrasound Classification with Focus on <u>Malignant Sensitivity</u>

Waseem Ahmad, Tajamal Hussain
Berliner Hochschule für Technik

Background

- ➢ Breast cancer is the **most diagnosed** cancer among women in the Europe and worldwide.
- ➢ Approximately **1 in 11 women in Europe** will develop breast cancer during their lifetime.
- ➢ Breast cancer represents nearly 30% of all new cancer cases in women across the EU.
- ➢ Ultrasound imaging is widely used for breast examination because it is: **Non-invasive**, **Cost-effective** and **Suitable for dense breast tissue**



Example breast ultrasound images showing benign and malignant lesions

➢ Manual interpretation of ultrasound images is operator-dependent and can vary across clinicians.

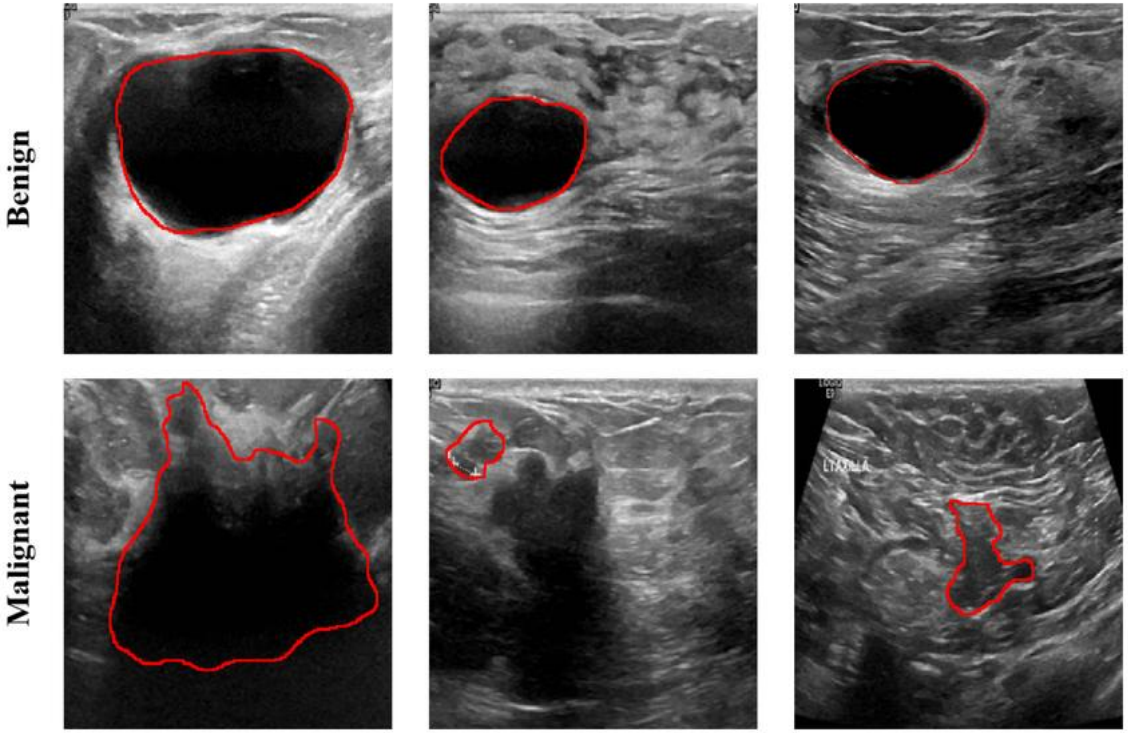
Key Challenge: Reducing false negatives in malignant diagnosis.

Clinical Risk & Problem Formulation

Breast cancer diagnosis using ultrasound imaging is challenging and time-sensitive. Importantly, **diagnostic errors do not carry equal clinical consequences.**

⚖️ Asymmetry in Clinical Risk

Clinical Finding	Clinical Consequence	Modeling Priority
Malignant tumor	False negative = missed cancer	Maximize recall (**Sensitivity**)
Benign / normal tissue	False positive = unnecessary follow-ups	Optimize under safety constraint



Example breast ultrasound images showing benign and malignant lesions

🛡️ Our work prioritizes **malignant recall over raw accuracy** to align model optimization with clinical safety.

https://www.researchgate.net/figure/Breast-ultrasound-benign-vs-malignant-comparison-image-with-the-red-solid-line-denoting_fig1_390691595

Related Work/Motivation

🎗️ **Motivation**

- ✓ Breast cancer incidence continues to rise globally, with significant impact in Europe.
- ✓ Early detection remains challenging due to variability in clinical interpretation and traditional diagnostic limitations.

📚 **Related Work and Results:**

- ✓ Several studies have applied machine learning and deep learning models to the Kaggle Breast Ultrasound dataset.
- ✓ Most approaches report high overall accuracy.
- ✓ However, limited emphasis has been placed on optimizing **malignant recall (sensitivity)** and reducing false negatives.

🎯 **Direction and Objective:**

- ✓ Evaluate multiple deep learning architectures.
- ✓ Implement strategies to address class imbalance and clinical risk asymmetry.
- ✓ **Primary Objective:** Improve recall for the malignant class while maintaining stable overall performance.

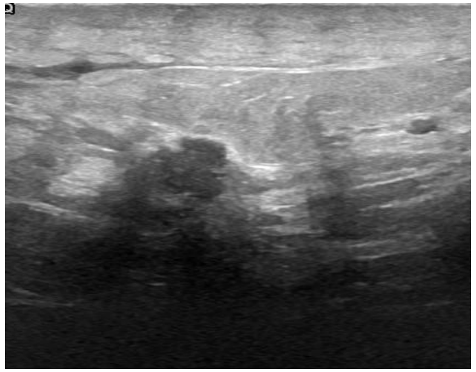
Our Solution

- We formulate breast ultrasound diagnosis as a three-class image classification task (benign, malignant, normal).
- Multiple CNN architectures are evaluated, including custom-designed models and transfer learning (AlexNet, ResNet50).
- Models are trained and compared under a consistent pipeline, with hyperparameter optimization and medical-relevant evaluation metrics.

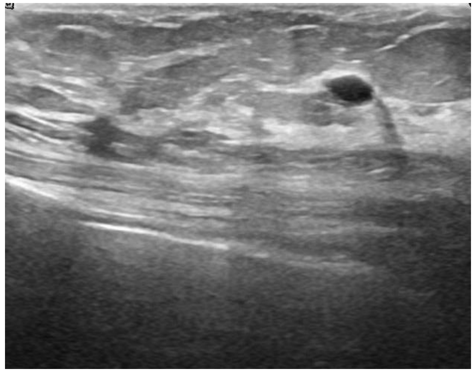
Dataset Overview

- **Source :** Breast Ultrasound Images Dataset from Kaggle
- **Classes**
  - Benign
  - Malignant
  - Normal
- **Size of dataset**
  - Benign : 437
  - Malignant : 210
  - Normal : 133
- Image size of 500*500 pixels.

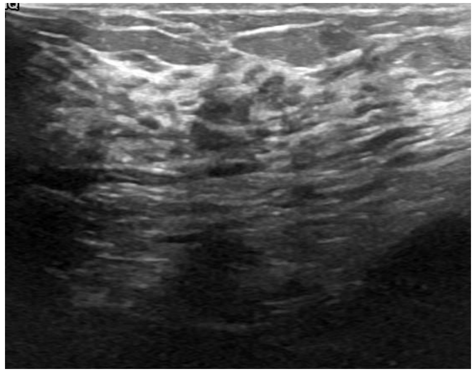
Malignant



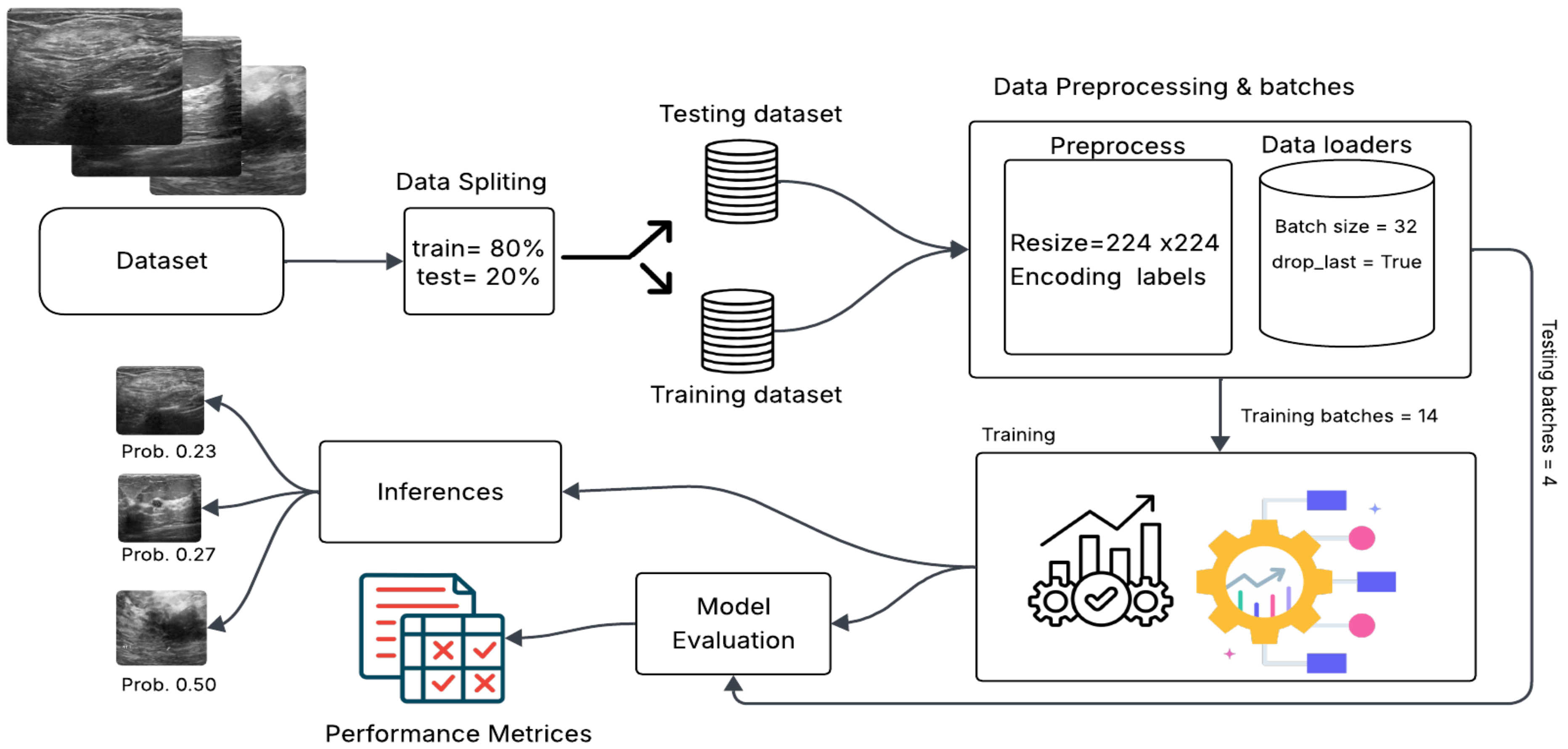
Benign



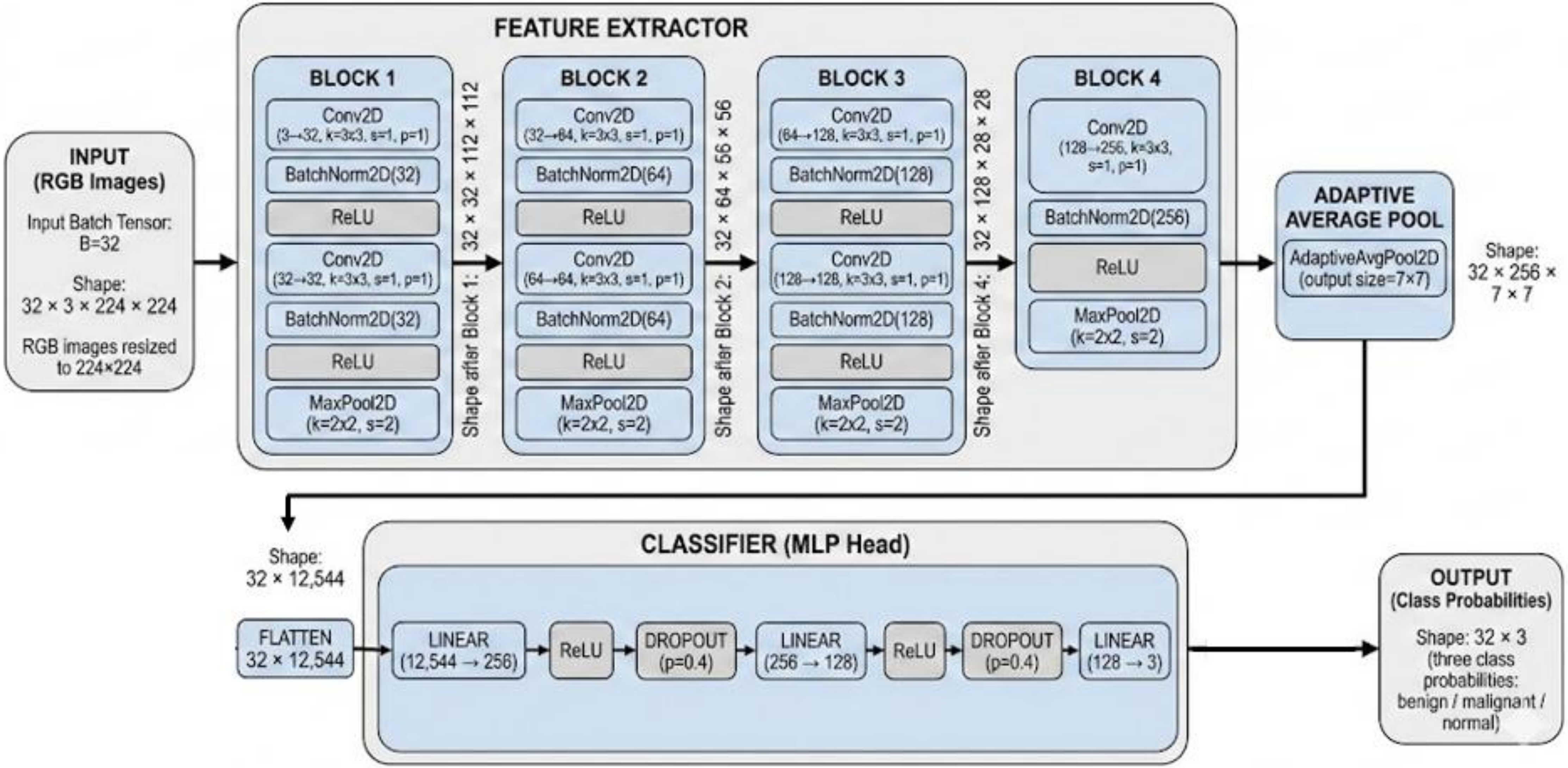
Normal



Training Pipeline



CustomCNN Model Architecture



Experimental Setup Table

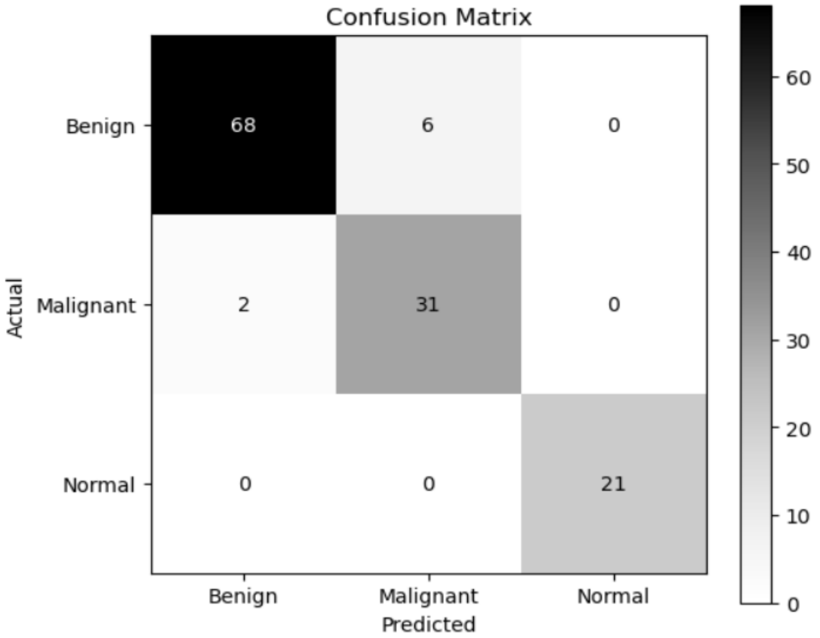
Experiment No#	Batch size	Epoch	Optimizer	LR Scheduler	Transfer Learning	Class Weights	Loss Method	Learning Rate
1	32	75	SGD	✅	❌	✅	CE Loss	0.01
2	32	50	SGD	❌	ResNet50	❌	CE Loss	0.02
3	32	75	SGD	❌	❌	❌	CE Loss	0.01
4	32	35	Adam	❌	❌	❌	CE Loss	0.01
5	32	20	SGD	❌	AlexNet	❌	CE Loss	0.02

RESULTS

Model No#	Model Name	Malignant Class			Accuracy
		Recall	Precision	F1	
1	Custom model with Exp. 1	0.94	0.84	0.89	0.87
2	Custom Model with Exp. 3 (Data Modified)	0.94	0.84	0.89	0.94
3	Custom with Exp. 4	0.91	0.88	0.90	0.82
4	AlexNet with Exp.5	0.88	0.85	0.87	0.79
5	ResNet with Exp. 2	0.64	0.70	0.67	0.74

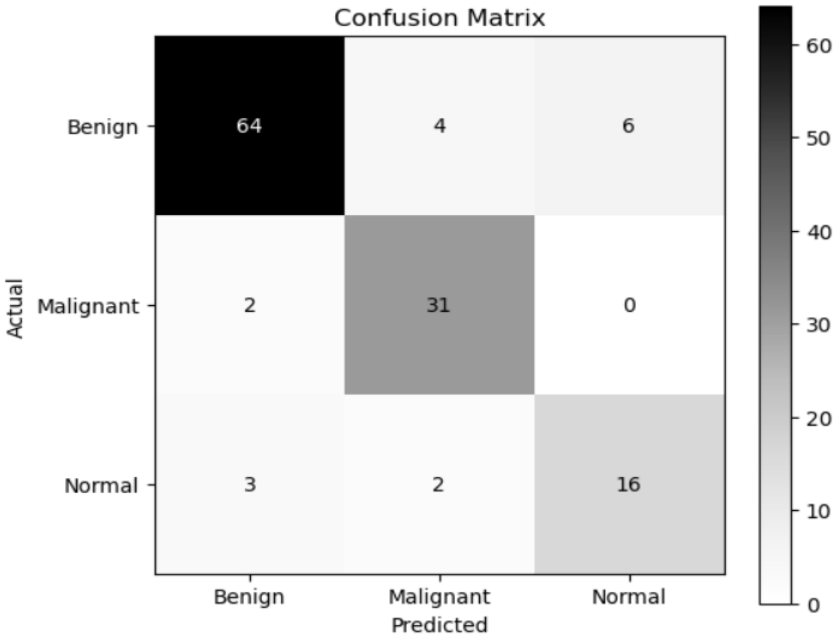
Performance

1 Custom Model with Modified Data (Masked)



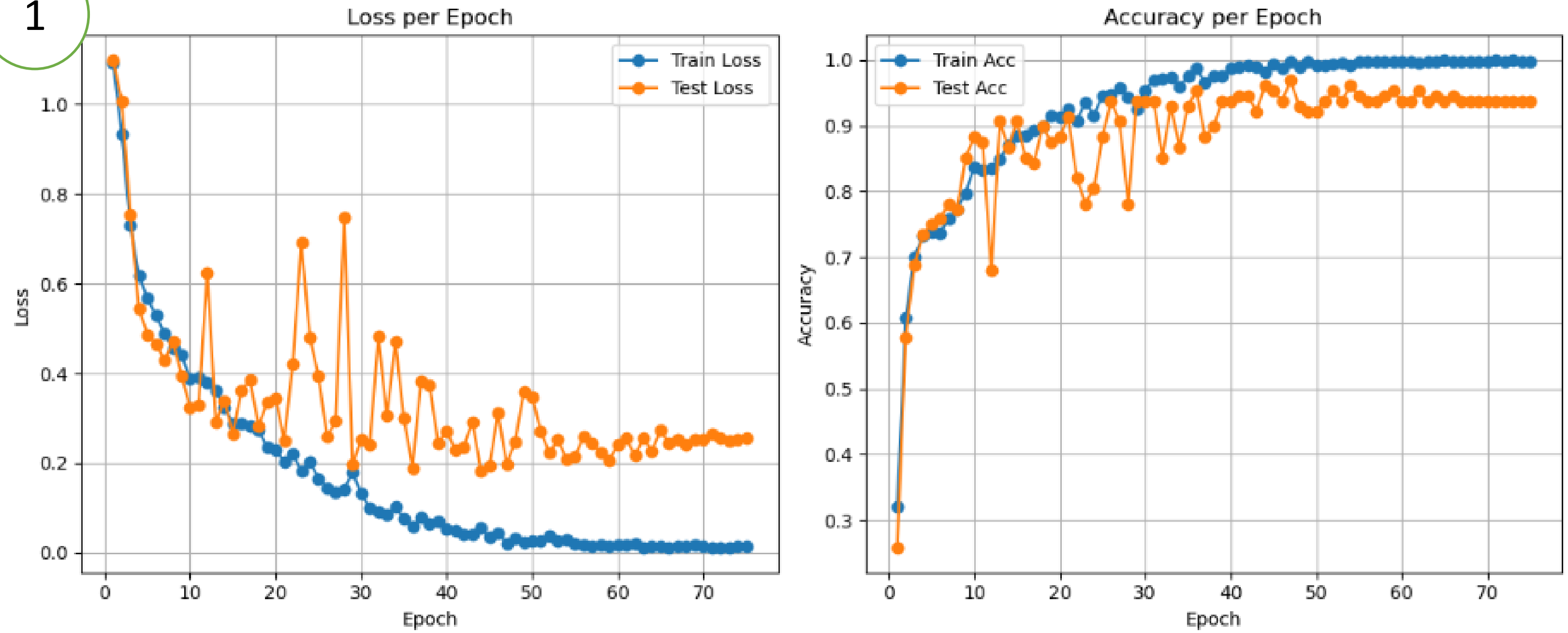
	precision	recall	f1-score	support
Benign	0.97	0.92	0.94	74
Malignant	0.84	0.94	0.89	33
Normal	1.00	1.00	1.00	21

2 Custom Model with Original Dataset



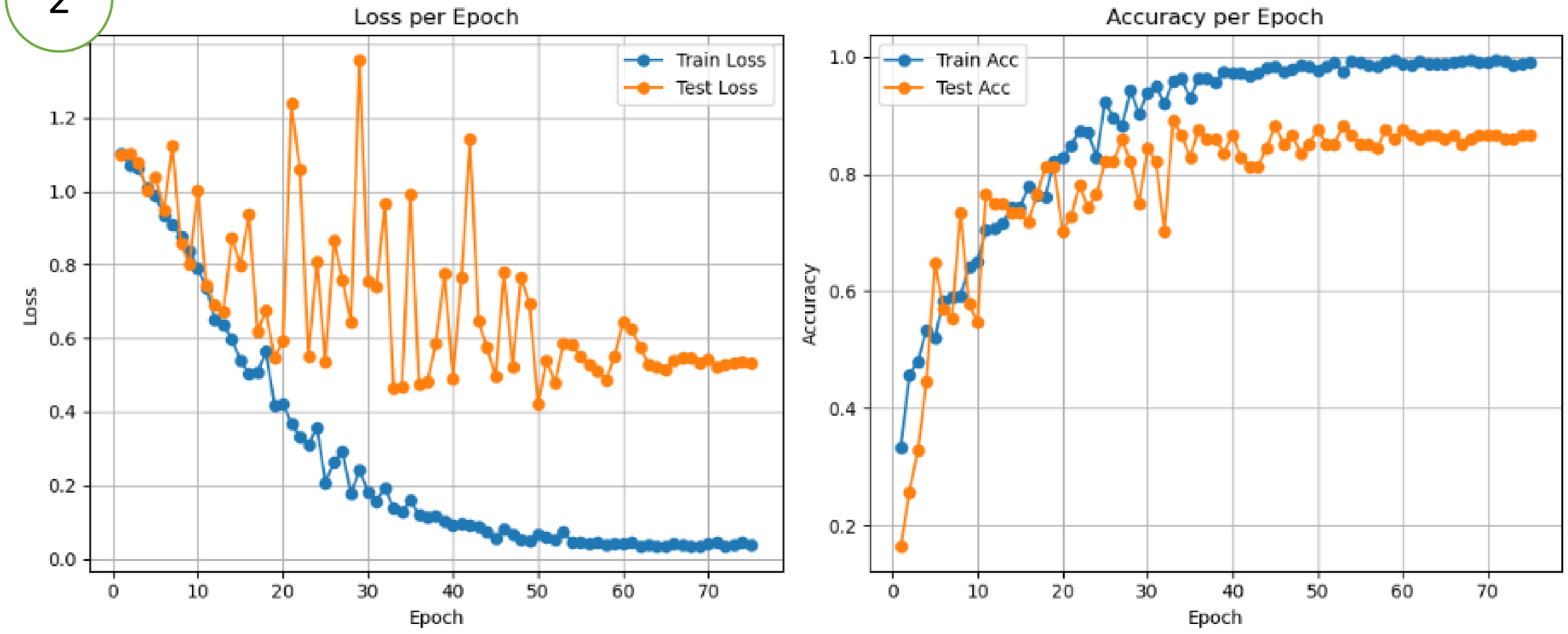
	precision	recall	f1-score	support
Benign	0.93	0.86	0.90	74
Malignant	0.84	0.94	0.89	33
Normal	0.73	0.76	0.74	21

1



Loss per Epoch
Train Loss
Test Loss
Loss
Epoch
Accuracy per Epoch
Train Acc
Test Acc
Accuracy
Epoch

2



Loss per Epoch
Train Loss
Test Loss
Loss
Epoch
Accuracy per Epoch
Train Acc
Test Acc
Accuracy
Epoch

References

https://ieeexplore.ieee.org/abstract/document/8003418

https://pmc.ncbi.nlm.nih.gov/articles/PMC6940474/