

# Breast Cancer Classification using CNNs

**DL Project Presentation**

By: Aarushi, Jayasuryan,  
Prahlad

# TABLE OF CONTENTS

**1.**

**Introduction**

**2.**

**Literature Review**

**3.**

**Dataset**

**4.**

**Methodology**

**5.**

**Training/Testing**

**6.**

**Results**

# INTRODUCTION

01.

Elaborate on what you want to discuss.

# Prevalence of Breast Cancer and the Impact of Deep Learning on Breast Cancer Detection

There were 2.3 million new cases and 685,000 deaths from breast cancer in 2020 alone. Breast cancer is the most common cancer diagnosed in women and the second largest cause of cancer-related deaths globally. Typically presenting as invasive ductal carcinoma (IDC), this type of cancer is defined by malignant cells that penetrate the ductal epithelium and infiltrate neighboring tissues. These cells can be identified by imaging techniques due to their irregular structures and cellular anomalies.

The incorporation of deep learning, specifically via convolutional neural networks (CNNs), has revolutionized the field of cancer screening by providing advanced instruments capable of deciphering and learning from intricate image data. Deep learning advances are pushing the envelope of medical imaging and breast cancer detection capabilities despite obstacles like network architecture optimization and computational demands.

# LITERATURE REVIEW

02.

**Available modalities:** X-ray, Mammography Imaging, CT

Scans, MRI, Tabular datasets etc.

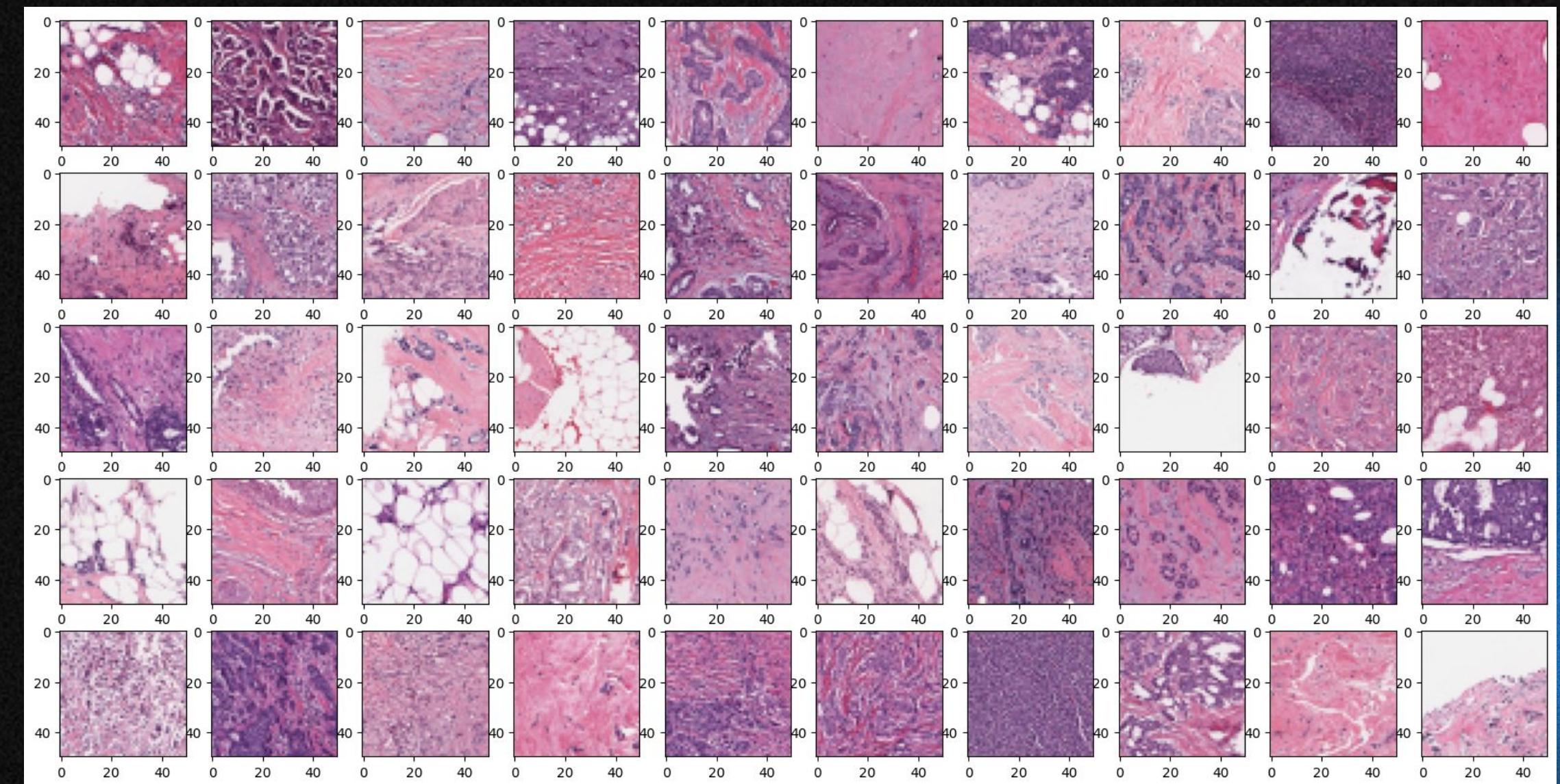
**Leading Performance Models:** CNNs (VGG-16, ResNet50, Inception Net etc.)

## Challenges with current approaches:

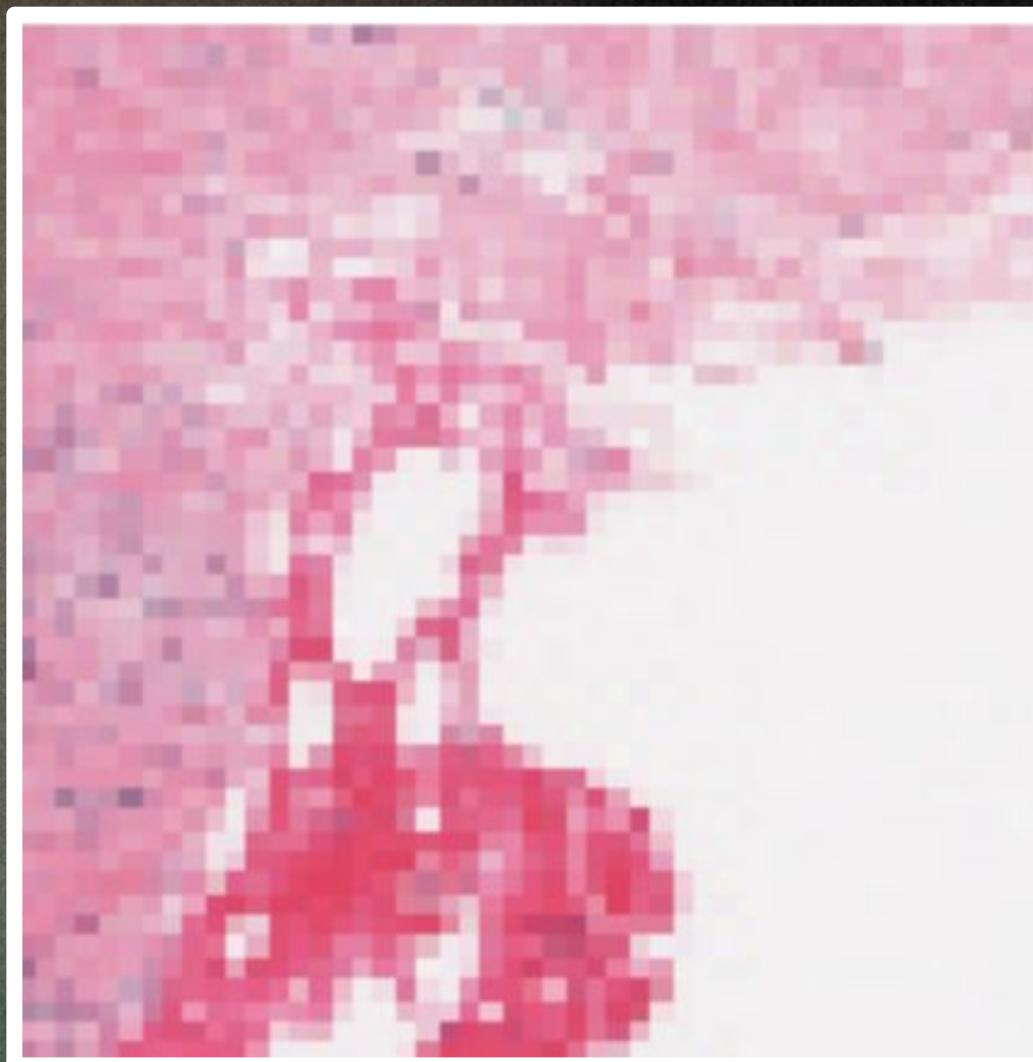
- ★ **Lack of generalization:** Models are often fine-tuned for patients of a specific region and may fail to detect visual symptoms of people from a wider range of genders, nationalities, medical history and lineage (*Convolutional neural networks for mammography mass lesion classification*)
- ★ **Limitations with certain modalities:** The authors of *Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model* mention that screen-film mammography (SFM) has several limitations in the technique of capturing the data, due to which ‘ a large number of false positives occur with this technique, with a rate between 4% and 34%’, which is a sizable proportion.
- ★ **Hyperparameter tuning:** Common issues with CNNs include choosing hyperparameter size. Variable kernel sizes make it difficult to select appropriate sizes for convolution operations, hindering the capture of all relevant features. Overfitting is another issue, particularly in deep networks, where the model memorises training data rather than generalising, leading to reduced performance on new datasets. Moreover, gradient sharing becomes challenging in deeper networks, affecting training and optimization,
- ★ **Inability to deal with class imbalance:** The datasets explored in most papers do not contain the class imbalance that real-life medical scenarios contain. One of the papers mentioned in the literature review section has 344 patients with 736 film images containing 426 benign mass lesions and 310 malign mass lesions, which is highly balanced

# DATASET

03.



The dataset consists of a total of 162 histopathological whole-slide images (WSI) of women diagnosed with invasive ductal carcinoma (IDC) at the Hospital of the University of Pennsylvania and The Cancer Institute of New Jersey.).



**Figure:** 9135\_idx5\_x1701\_y1851\_class,  
image with XY coordinates (1701, 1851) and of  
class 1 i.e. IDC positive

- ★ **Total Samples:** 277,524 patches of size 50 x 50 pixels each
- ★ **IDC Negative Sample Count:** 198,738
- ★ **IDC Positive Sample Count:** 78,786
- ★ The class counts point towards a class imbalance with a bias towards negative classification.

# METHODOLOGY

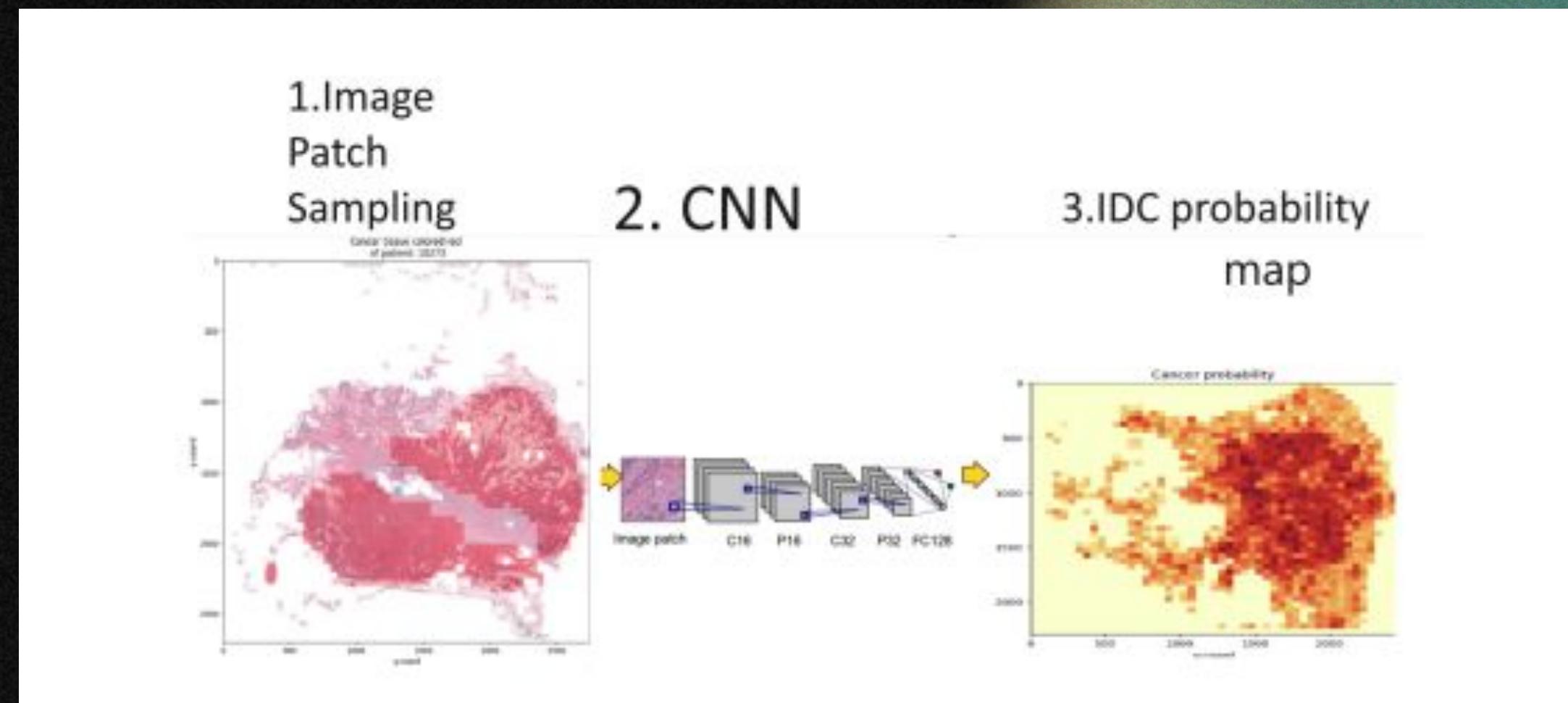
04



# Broad steps used to classify as cancer

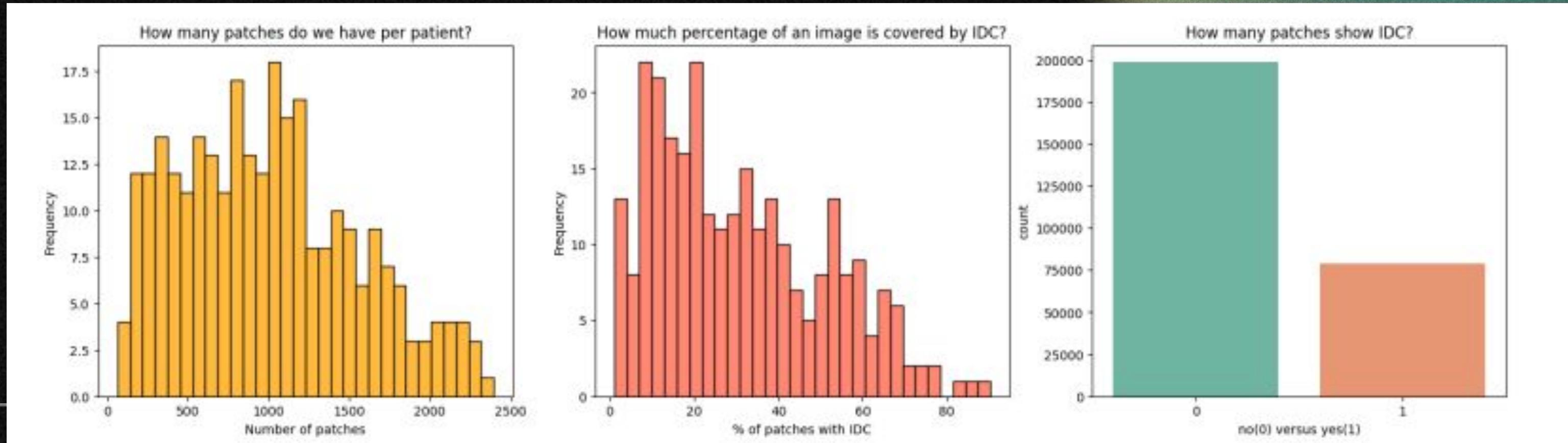
The methodology for classifying the cancer as IDC broadly follows a series of steps given below:

1. Image patch sampling
2. Convolutional Neural Networks
3. IDC probability map



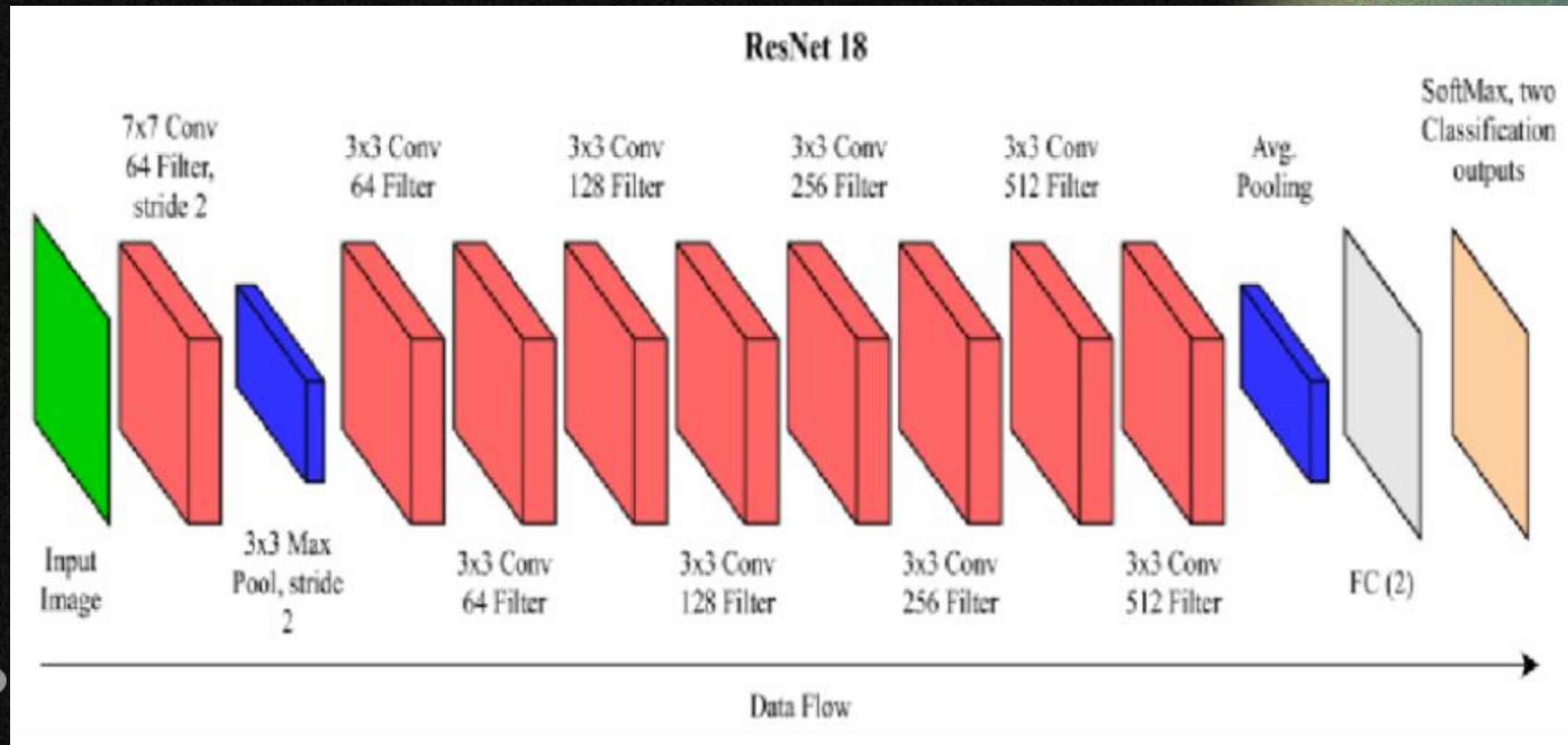
# Exploratory Data Analysis

We used a variety of data visualizations inn our project. Some of the most important ones include - Number of Patches per Patient, Percentage of Image Covered by IDC and IDC Presence in Patches. Shown Below



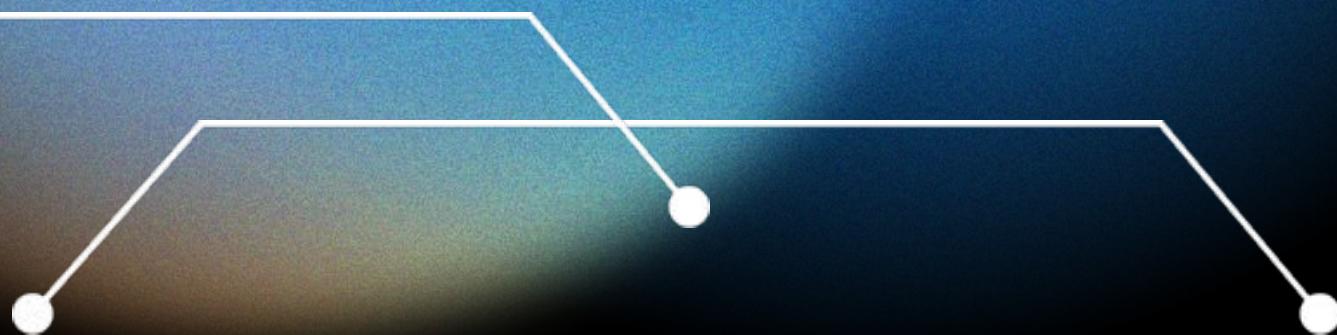
# Model Architecture

The small dataset size risks model overfitting, which can be mitigated using transfer learning with a pre-trained ResNet-18 model. This model is suitable for diagnosing ductal carcinoma from 2D tissue slides due to its fewer layers, which reduces computational demands while maintaining high performance. ResNet-18 helps prevent the vanishing gradient problem, essential for training deep networks, resulting in high precision and robust performance. This approach balances computational efficiency with diagnostic accuracy, enhancing clinical outcomes in breast cancer treatment.



# TRAINING/TESTING

05.



# Training Loop

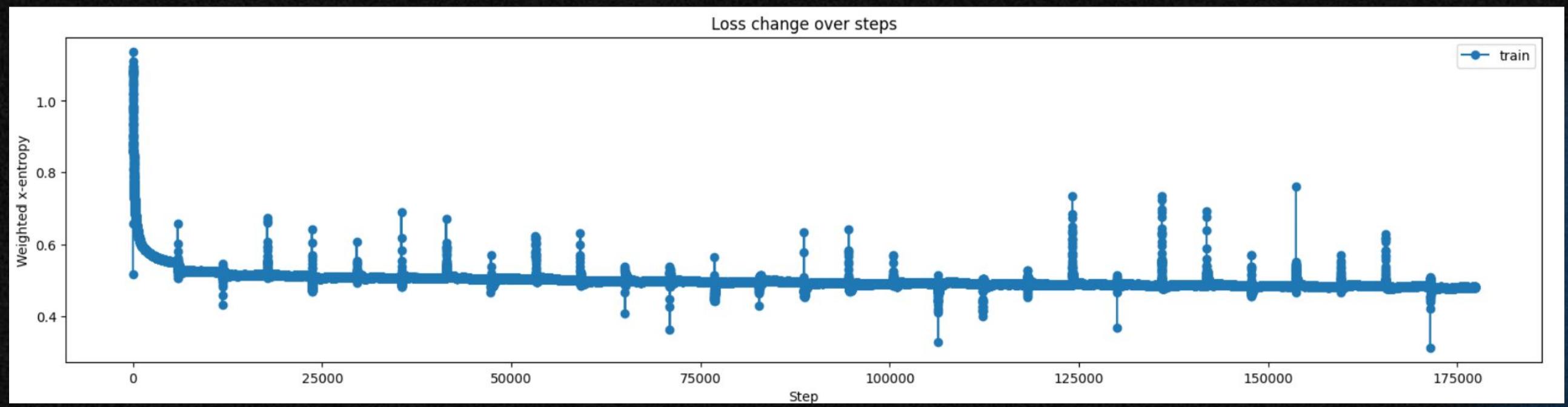
The training loop is a fundamental component of deep learning that involves iterating over a dataset for 29 epochs to optimize a model's performance. It initiates by storing the best initial weights and setting up tracking for losses and performance metrics across training, validation, and testing phases. The loop includes forward and backward passes for weight updates, with optional regularization and learning rate adjustments. Learning rates are modulated using CyclicLR from PyTorch, which applies a cyclical "triangular" policy to help quickly find an effective learning rate and enhance model convergence and performance. Improved weights are saved if accuracy increases on the development set, and the loop ends by reporting training outcomes and storing the best-performing model.

## Testing Loop

The test loop evaluates a deep learning model trained for medical image classification tasks, such as cancer detection in tissue samples, by processing images in batches without computing gradients to solely assess model performance. With the model set to evaluation mode, which disables training-only layers like dropout for consistent predictions, the loop retrieves batches of data, computes outputs, predictions, and confidence scores, and stores them along with metadata in a DataFrame. This DataFrame is then cleaned, used for detailed performance analysis and visualizations to validate the model qualitatively and quantitatively, and finally exported to CSV files for further use. This ensures the model is thoroughly evaluated for reliability before clinical deployment.

# RESULTS

06.



- ★ The model showcases proper loss convergence on the training set over the course of 20+ epochs
- ★ There are fluctuations but overall decrements in validation and testing loss, which were calculated in order to facilitate early stopping in the case of overfitting

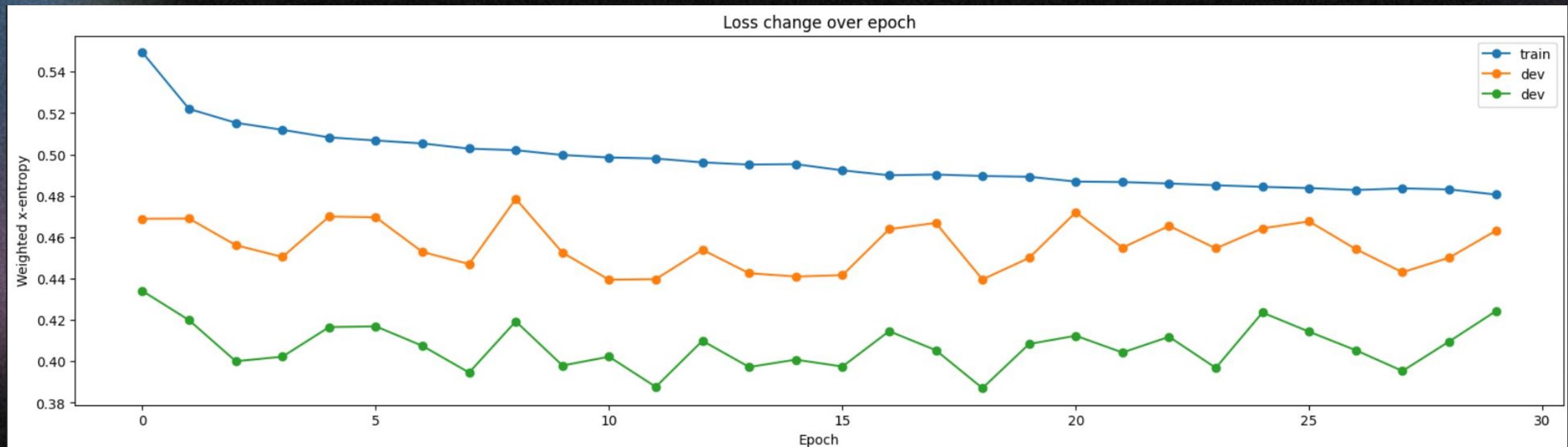


TABLE I  
PRECISION, RECALL AND ACCURACY MEASURES FOR THE TRAINING AND TESTING SET OF THE MODEL

Data segment	F1 Score $F1 = (2*P*R)/(P+R)$	Precision $P = TP/(TP + FP)$	Recall $R = TP/(TP + FN)$
Training set	0.792	0.824	<b>0.736</b>
Testing set	<b>0.806</b>	<b>0.876</b>	0.695

- ★ Comparable test and train accuracy indicate well generalized model (credited to using a pre-trained model and fine-tuning for dataset)
- ★ Recall primarily signifies the capacity to identify true positives alongside capturing false negatives which should have been classified as cancer-positive.
- ★ The training set recall is better than the testing set recall, in order to improve this, a wider distribution of training samples could be used using some level of data augmentation to generate more sub-samples or add noise to the current training set
- ★ The accuracies of both training and testing set are also comparable to the baseline paper, with an overall accuracy of 0.87

- ★ The model showcases proper loss convergence on the training set over the course of 20+ epochs
- ★ There are fluctuations but overall decrements in validation and testing loss, which were calculated in order to facilitate early stopping in the case of overfitting

# THANK YOU!