Ahmedabad University

ENR497: B.Tech. Engg. Project

School of Engineering & Applied Science

(Winter 2025)

Breast Cancer Segmentation using nnU-net

Dev Bhalodia: AU2140208

Start date of project (Jan 2024)

End date of project (May 2024)

Outline of Presentation

- Motivation
- Literature survey
- Overall objectives of the project
- Main outcomes of the project
- Methodology
- Final results
- My contribution
- Gantt chart
- Who will benefit from my work

Motivation

- Manual segmentation is **time consuming**
- Many existing models often l**ack generalization**
- **Relevance to current trends in research and development**
  - Shift from 2D to **3D volumetric segmentation** for richer spatial context.
  - Focus on **end-to-end automation** with modular pipelines.
  - Emphasis on **explainability & post-processing** for clinical trust.
- **Relevance to industry and market**
  - Rise of **AI-assisted radiology tools** in hospitals.
  - Demand for **efficient, accurate diagnostics** support.
  - Growth in **deployable AI** for low-resource healthcare settings.

Literature survey

- Traditional **3D U-Net** and **V-Net** face problem dealing with small irregular tumors.
- **Attention U-Net** helps focus on salient tumor features for 2D data.
- Recent **transformer-based** models like **Swin UNet** are **computationally expensive** as they are data hungry.
- 3D **DenseSeg** performs well with limited data.
- Pre-trained models like **MedSAM** are **not designed to handle 3D data**. It would train the 3D volume slice-by-slice, not preserving spatial context along z-axis

Objectives of the project

- **Leverage 3D nnU-Net** for volumetric tumor segmentation.
- Achieve **high performance metrics** such as Dice, IoU, PPV, when comparing with models already used for 3D segmentation of this dataset.
- **Ensure generalizability** and real-world usability.

Main outcomes

- I used nnU-net which **automated the process** of segmenting 3D volumetric images.
- Achieved **high metric scores** and **outperformed** baseline models.
- Made the project applicable for **clinical** (by radiologists) and **research use** (predictions serving as input for classifying tumors and model's performance serving as baseline for other future innovations).

Relevant Tools and Technologies

- **Language & Frameworks:**
  Python | PyTorch | nnU-Net

- **Libraries Used:**
  NumPy, Matplotlib, Nibabel, OpenCV

- **Development Platform:**
  Google Colab (with Tesla T4 GPU, CUDA)

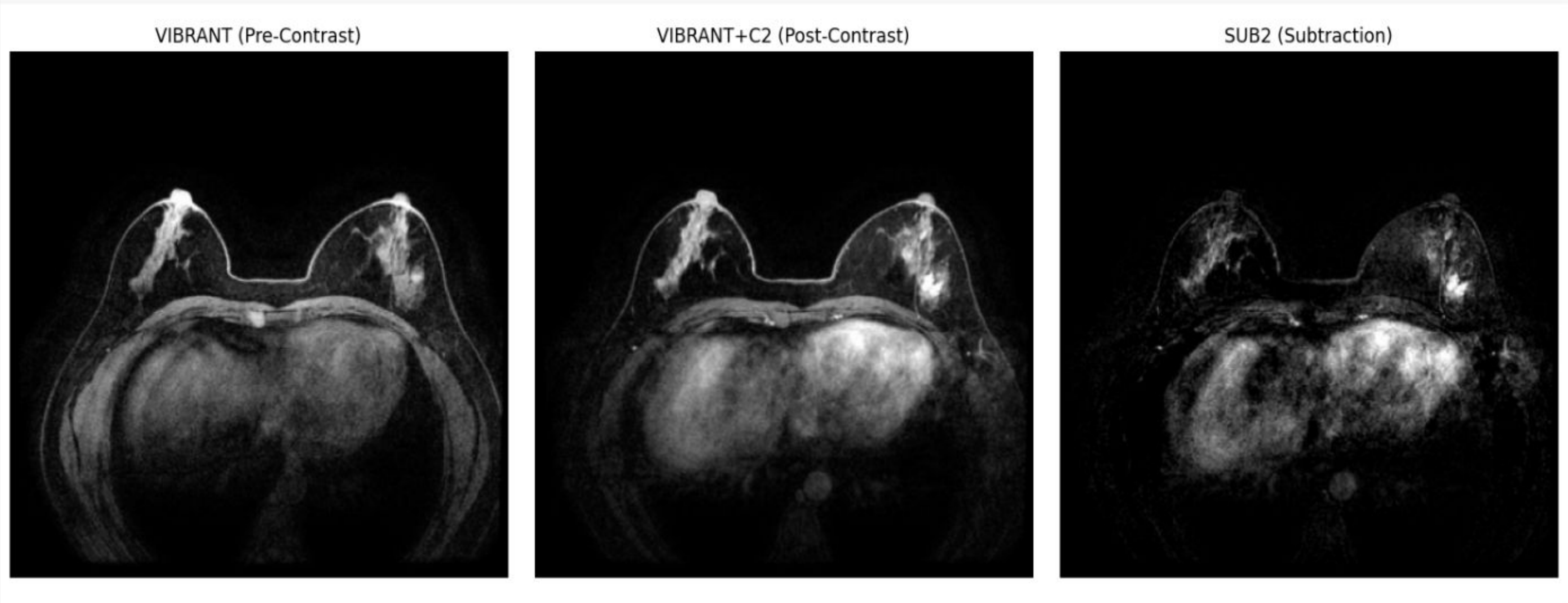
- **File Handling & Visualization:**
  NiFTI support via Nibabel, visualized using Matplotlib

Methodology

Dataset Description

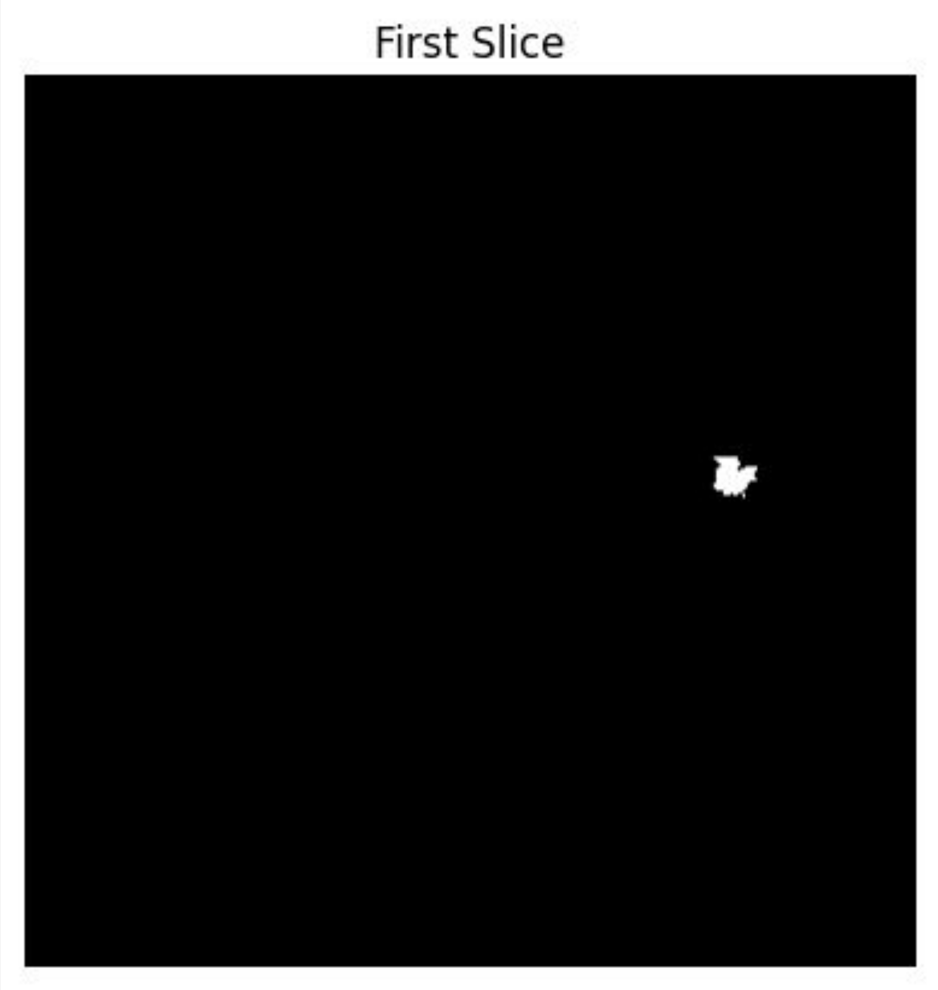
- Dataset: BreastDM
- Total Patients: 232
- Modalities per patient:
- Pre-contrast
- Post-contrast
- Subtraction image (SUB2)

Image Modalities



The figure above shows the first slice of each image volume of all 3 modalities of a patient.

Corresponding Label



The figure above shows the first slice of corresponding label (SUB2) of the previous 3 image slices.

Dataset preparation

Modality assignment

- All 3 modalities used for image input
- Only SUB2 used for ground truth labels

Slice cleaning

- Each modality had 8 slices, but many samples had blank/padded slices.
- Removed blank slices from all 3 modalities and corresponding labels in .npy format

Format conversion & structuring

- Converted **.npy → .nii.gz** using nibabel
- Followed nnU-Net **folder conventions**: imagesTr/, labelsTr/, imagesTs/
- Followed strict nnU-Net **naming convention**:
- **Images**: BC_***_@@@@.nii.gz (e.g., BC_001_0002.nii.gz: 2nd image, 3rd modality)
- **Labels**: BC_***.nii.gz (e.g., BC_001.nii.gz for corresponding label)

Label remapping

- **Remapped label value**: 255 → 1 (foreground class)
- Ensured compatibility with nnU-Net's binary segmentation expectations

Dataset.json definition

- Channel (modality) names
- Label classes (binary)
- Training sample count
- File extension: .nii.gz

Dataset preprocessing

Cropping

- Each 3D volume is **cropped to its non-zero region** to remove blank areas.
- **Reduces** input volume size and computational cost.
- A **"fingerprint"** is generated storing metadata: size, spacing, modalities, class count, and training samples.

Resampling

- Samples often have **inconsistent voxel spacing** due to scanning variations.
- nnU-Net computes the **median voxel spacing** and resamples all volumes to this target:

1. **Images** → interpolated using **3rd-order spline** for smooth intensity preservation.
2. **Masks** → interpolated using **nearest-neighbor** to preserve label values.

- Standardization ensures **consistent learning** and **prevents spatial artifacts**

Ahmedabad University

Configuration selection

One out of the 3 configurations is selected:

- **2D U-Net:** Fast, low-resource; ignores spatial (z-axis) info.
- **3D Full-Resolution U-Net:** High accuracy, preserves spatial relationships; more resource-intensive.
- **U-Net Cascade:** Two-stage refinement; used for very large datasets as it is the most computationally expensive configuration

Why 3D full res U-net was selected

- The dataset is 3D MRI with **significant z-axis information.**
- 3D full-res **captures spatial context across all three dimensions.**
- Although computationally expensive, it was automatically selected during preprocessing due to its **suitability.**

Normalization

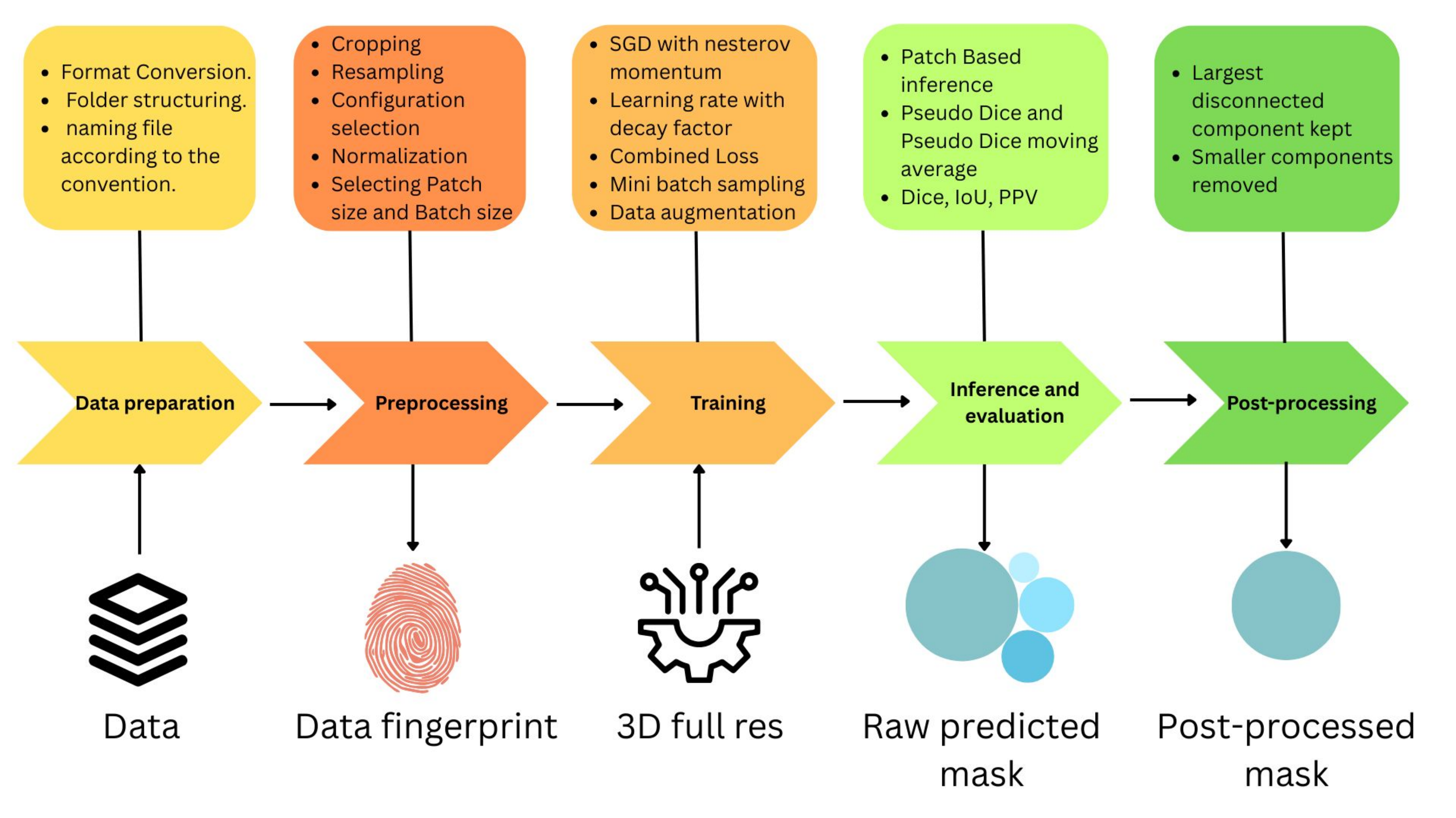
- **Z-score normalization** applied to each 3D image:

$$Z = \frac{\overbrace{x}^{\text{Score}} - \overbrace{\mu}^{\text{Mean}}}{\underbrace{\sigma}_{\text{SD}}}$$

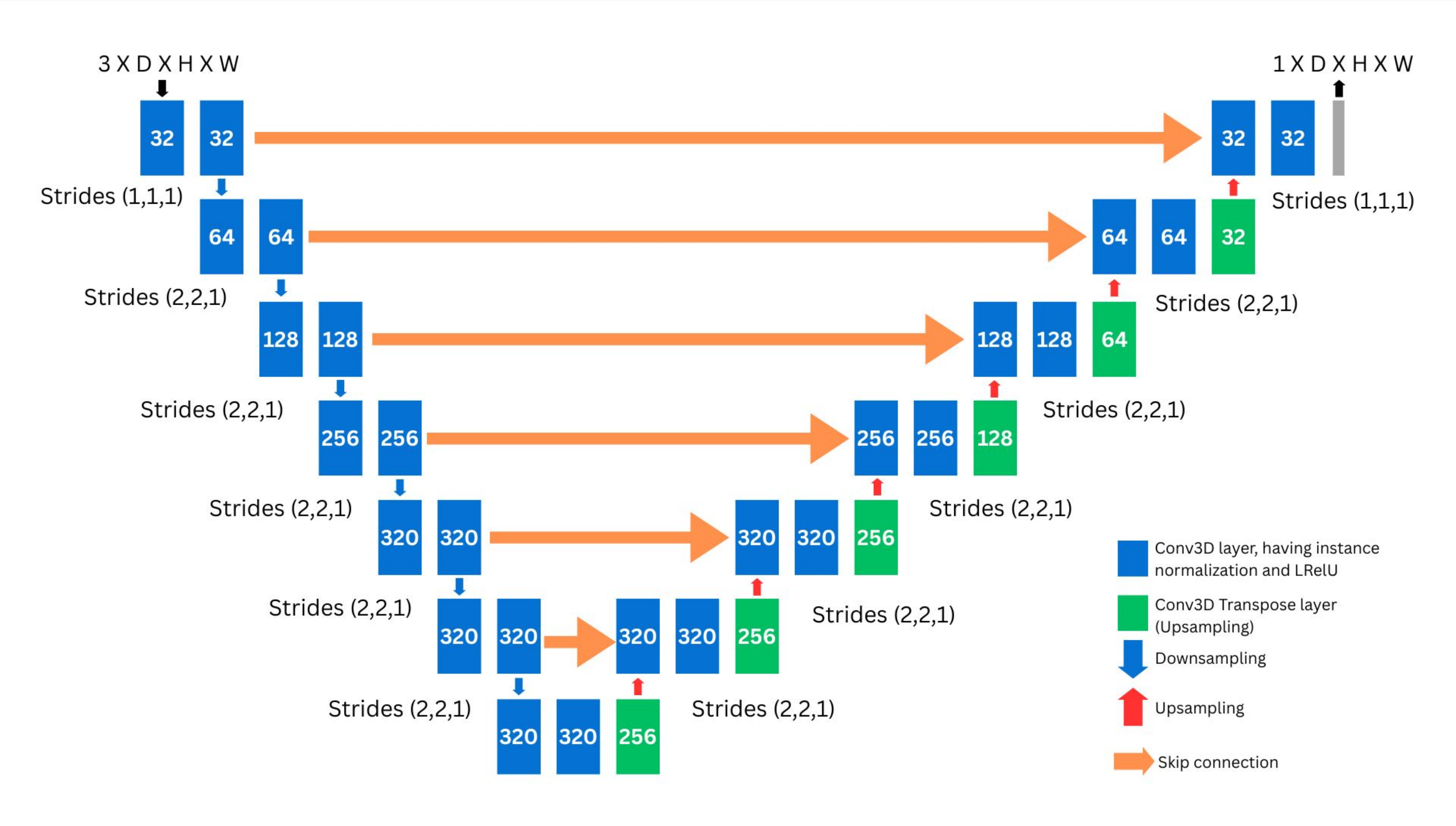
Mean and standard deviation are calculated from **non zero voxels.**

- If cropping reduces **volume by ≥25%**, normalization is performed only **within non-zero regions**, and other voxels are set to zero.

nnU-net pipeline



3D full res architecture



Key differences from Vanilla 3D U-net

Dynamic data-specific adaptation:

- Automatically sets patch size, pooling layers, depth, and batch size based on GPU memory and data.
- Vanilla 3D U-Net has fixed configurations → poor generalization

Preservation of full spatial resolution:

- Minimizes downsampling to retain spatial detail.
- Helps in detecting small, irregular tumors effectively

Dynamic Patch size-Based Training:

- Patch size is customized based on median shape and spacing.
- Ensures consistent context across patients.
- Vanilla U-Net uses arbitrarily fixed patches.

Encoder-Decoder Structure

- **Encoder:** Extracts semantic features through convolutions and pooling.
- **Decoder:** Reconstructs spatial resolution via upsampling.
- **Skip Connections:** Transfers high-res features from encoder to decoder for better localization.

Architectural Enhancements

- **Leaky ReLU (vs. ReLU):** Handles negative inputs to avoid dying ReLU issue.
- **Instance Normalization:** Each patch normalized independently instead of in batches — ideal for small batches and MRI data.
- **Patch-Based Inference:**

1. Inference done on overlapping patches.
2. Gaussian-weighted averaging used to reduce border artifacts

Why this model fits my dataset?

- **Maintains spatial resolution** — crucial for 3D MRI data.
- **Automatically adapts to dataset** — proved to generalize well.
- **Balances accuracy and memory efficiency** in 3D volume segmentation.

Training pipeline

Training progress and termination

- **Default number of epochs is 1000** but I saw no improvement after 38.
- The **best model gets saved** and training can be resumed from there.
- **1 epoch = 250 mini-batches**

Optimizer

- Stochastic Gradient Descent (**SGD**) with **Nesterov Momentum.**
- Momentum coefficient: **μ = 0.99** (carries forward 99% of previous update).
- Initial learning rate: **0.01**

Learning rate decay

- lr = lr_init × (1 − epoch / epoch_max)^0.9
- Starts with high learning rate → gradually decreases.
- Helps in **fine-tuning** and **prevents getting stuck in local minima.**

Loss function

- **Combined loss** = Dice Loss + Cross Entropy Loss
- **Dice Loss:**

  $1 - (2 \times |A \cap B|) / (|A| + |B|)$

  A = predicted mask, B = ground truth mask

- **Cross entropy loss:**

  $-\sum y_i \times \log(p_i)$

  y_i: actual label, p_i: predicted probability

Mini-batch sampling strategy

- **66.7%** of patches: **randomly selected** locations.
- **33.3%** of patches: guaranteed to **contain foreground** (label = 1)

On-the-fly Data Augmentation

- Performed during training to **avoid overfitting** and **increase variability**
- **Techniques** used:
    1. Rotation, Scaling, Gaussian noise & blur
    2. Brightness/Contrast changes
    3. Low-res simulation, Gamma correction
    4. Mirroring (flipping)

Ahmedabad University

Inference and Evaluation

Inference Strategy

- **Patch-based Inference:**
  The model does not process the full 3D volume at once. Instead, it predicts over overlapping patches to reduce memory consumption.
- **Gaussian Weighting:**
  When overlapping patches are merged, central voxels are given higher weight. This avoids border artifacts and improves segmentation quality near edges.

Post-processing

- Performed after inference to **improve segmentation output**
- **Key goals:**
    - **Remove small false positives** (e.g., noise misclassified as tumor).
    - Keep only the **largest disconnected component,** assuming tumors are continuous.

- In nnU-Net: Post-processing usually runs after 1000 epochs, but I **manually scripted** it since I trained for 53 epochs

Visualization

Purpose:

- Offers **qualitative assessment** of model's performance.
- Helps **verify prediction alignment with true mask.**

What Was Visualized:

- All **3 input modalities of image.**
- True mask, Predicted mask, and Post-processed mask
- Side-by-side layout enables easy comparison of segmentation accuracy and refinement

Evaluation

Metrics in nnU-net pipeline (default):

- **Pseudo Dice Score:** Dice score averaged over an entire mini-batch
- **Moving Average Dice:** Smoothed over epochs to track performance trend

Final Metrics for Research Comparison:

- **Dice Score** (DSC): Measures overlap between predicted and true masks.
- **IoU** (Intersection over Union): A stricter overlap metric.
- **PPV** (Positive Predictive Value): Accuracy of positive predictions — how many predicted positives are actually correct.

Final Results

Evaluation results

Best Saved Model:

- **Checkpoint:** checkpoint_best.pth
- **Pseudo Dice Score:** 0.8920
- **Pseudo Dice Moving Avg:** 0.8748

Final Metrics for Research Comparison:

- **Dice Score** (DSC) = $(2 \times |A \cap B|) / (|A| + |B|)$
- **IoU** (Intersection over Union) = $|A \cap B| / |A \cup B|$
- **PPV** (Positive Predictive Value) = $|A \cap B| / |B|$

Where:

  - A = Ground truth positive voxels
  - B = Predicted positive voxels

Before post-processing:

- **Mean Dice Score**: 0.8334
- **Mean IoU Score**: 0.7343
- **Mean PPV Score**: 0.7877

After post-processing:

- **Mean Dice Score:** 0.8698
- **Mean IoU Score:** 0.8052
- **Mean PPV Score:** 0.8760

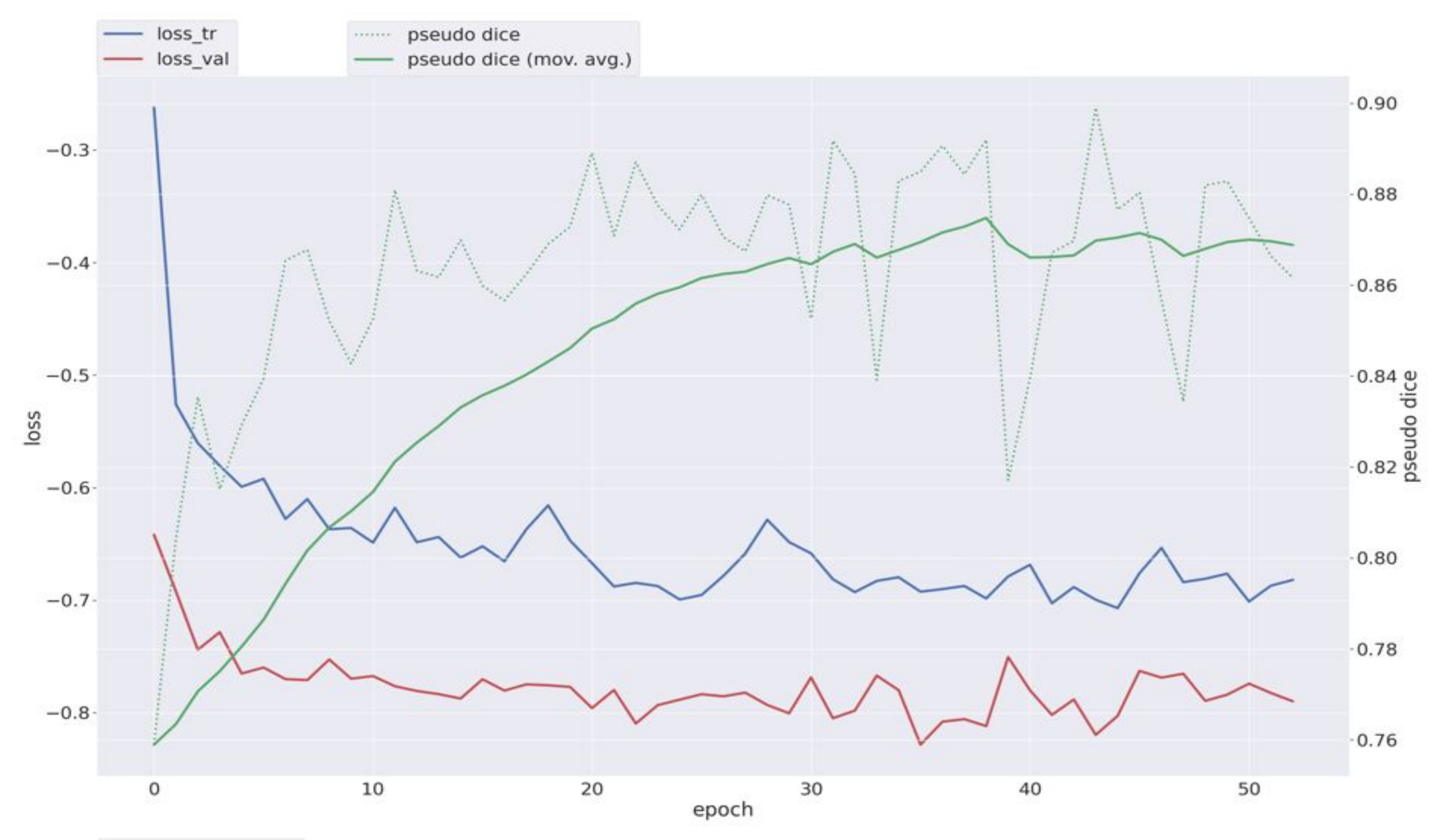
Ahmedabad University

Quantitative comparison with baseline 3D models

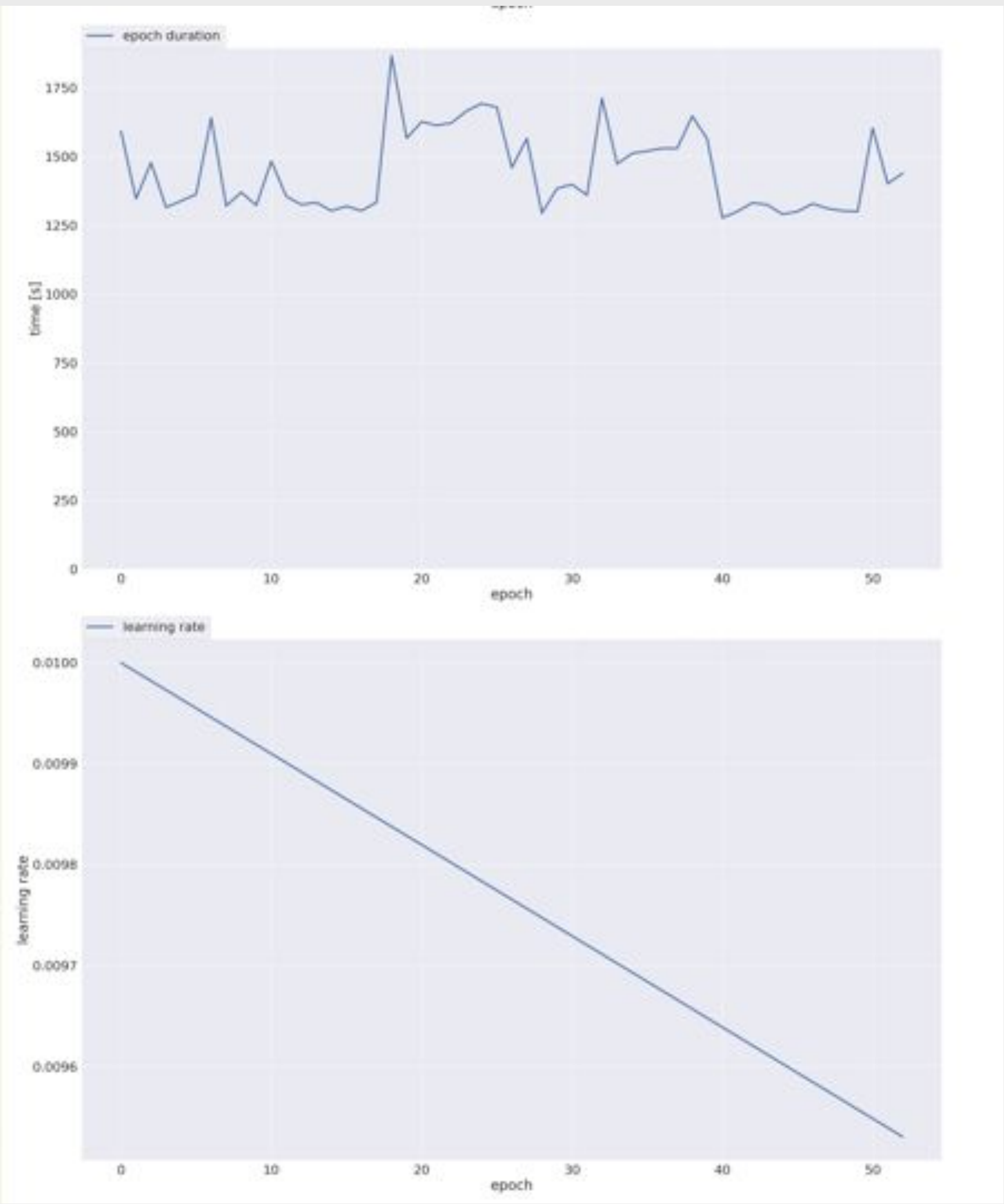
Model	DSC (%)	IoU (%)	PPV (%)
3D U-Net (Paper)	66.5	75.1	83.3
3D V-Net (Paper)	65.6	74.2	77.1
3D DenseSeg (Paper)	67.0	74.9	84.1
nnU-Net	**86.9**	**80.5**	**87.6**

Ahmedabad University

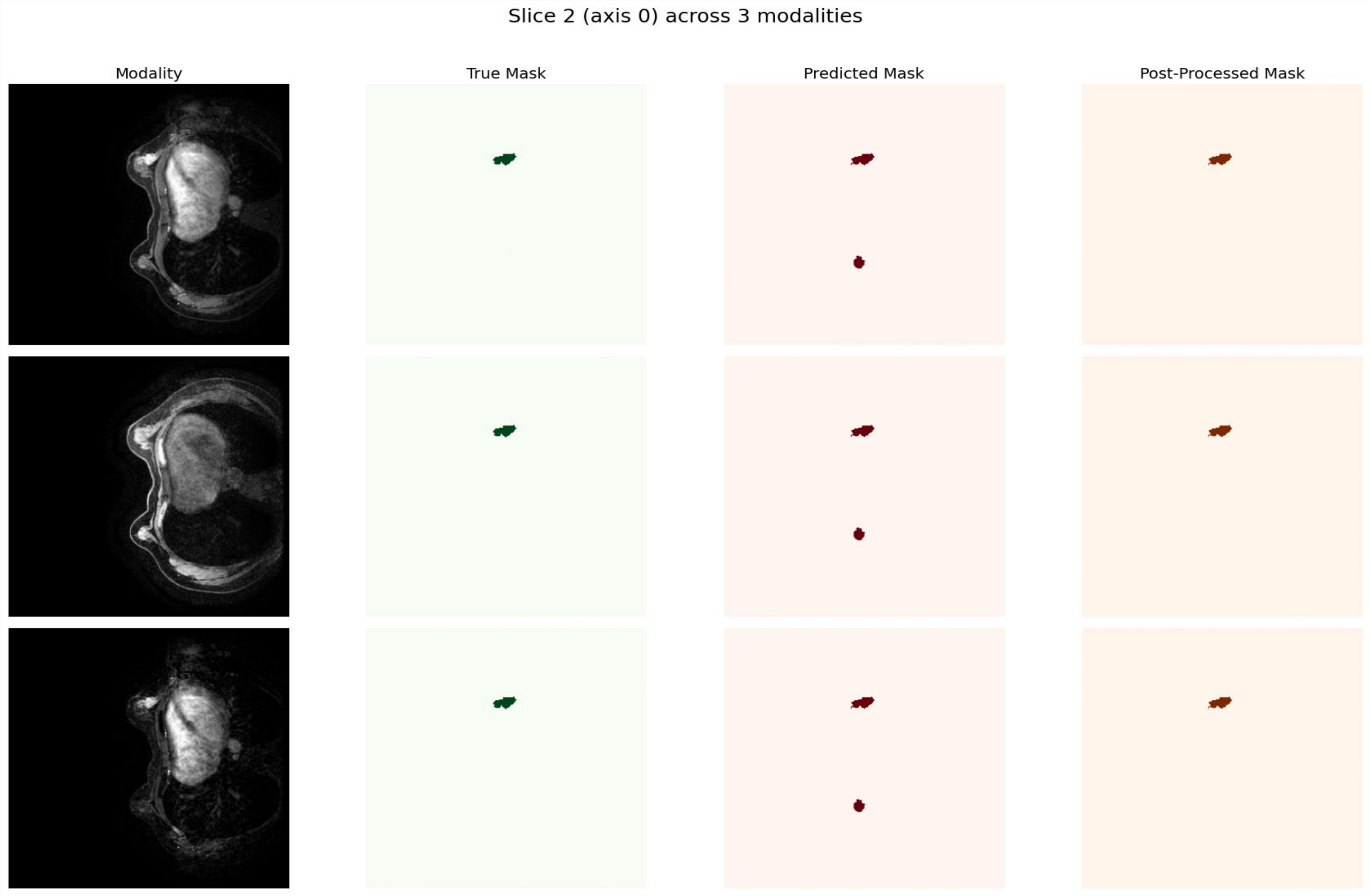
Training trends and metric progress



Time per epoch and learning rate decay:



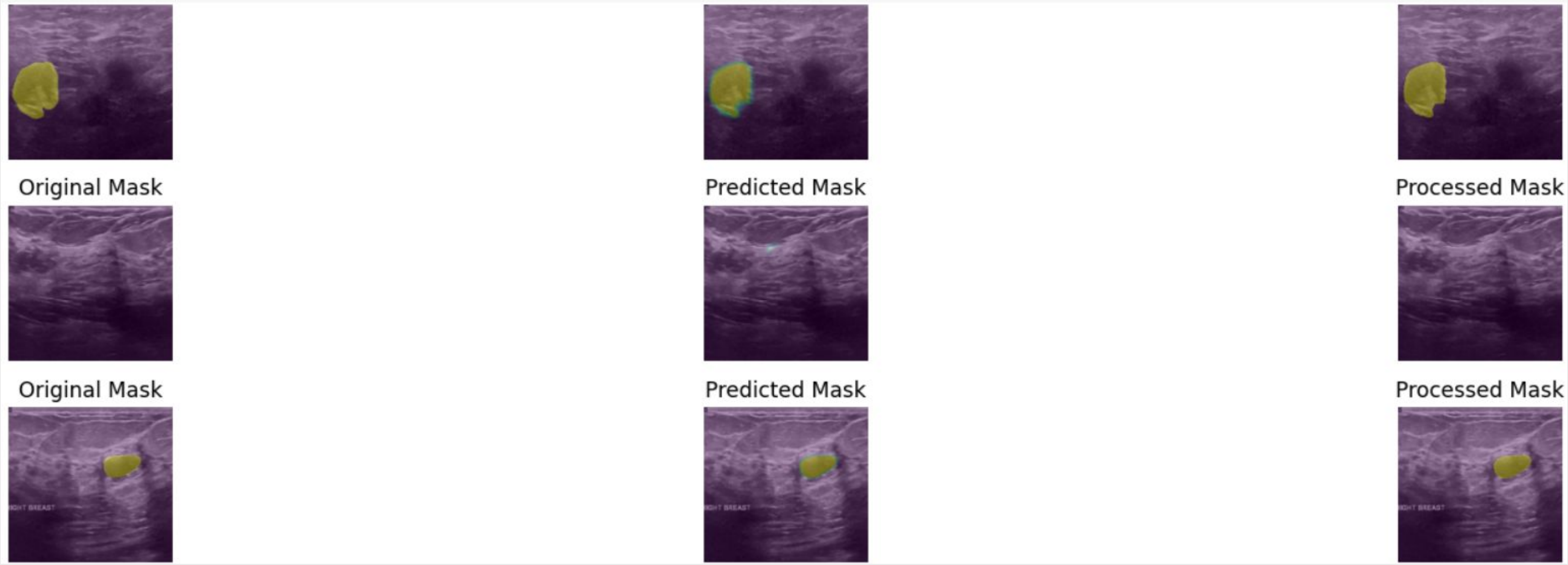
Segmentation results:



Initial Works:

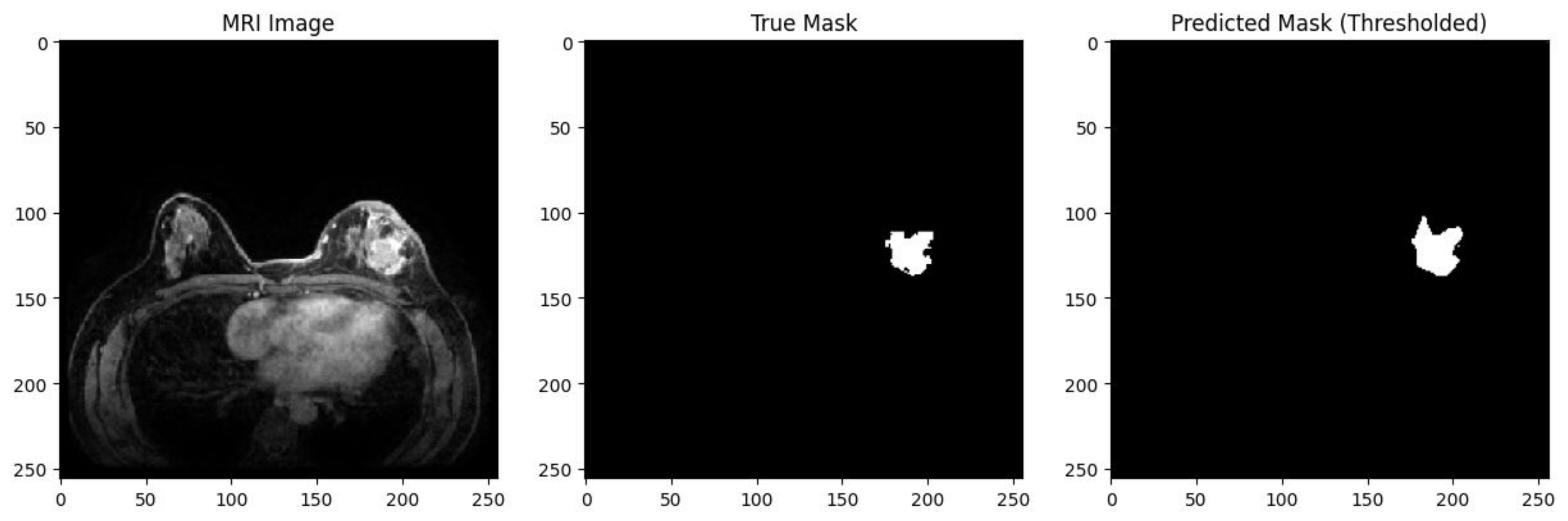
Using Kaggle 2D Ultrasound Dataset

- Chosen as it was readily available and **did not require** any extensive data preparation and preprocessing.
- **U-Net:** Achieved mean IoU of **0.46**
- **Attention U-Net:** Improved to **0.54** with better focus on tumor regions



Transition to BreastDM 2D MRI Dataset

- Precursor to 3D experimentation
- **U-Net:** mean IoU = **0.49**
- **Attention U-Net:** mean IoU = **0.928**



Ahmedabad University

Quantitative comparison with baseline 2D models

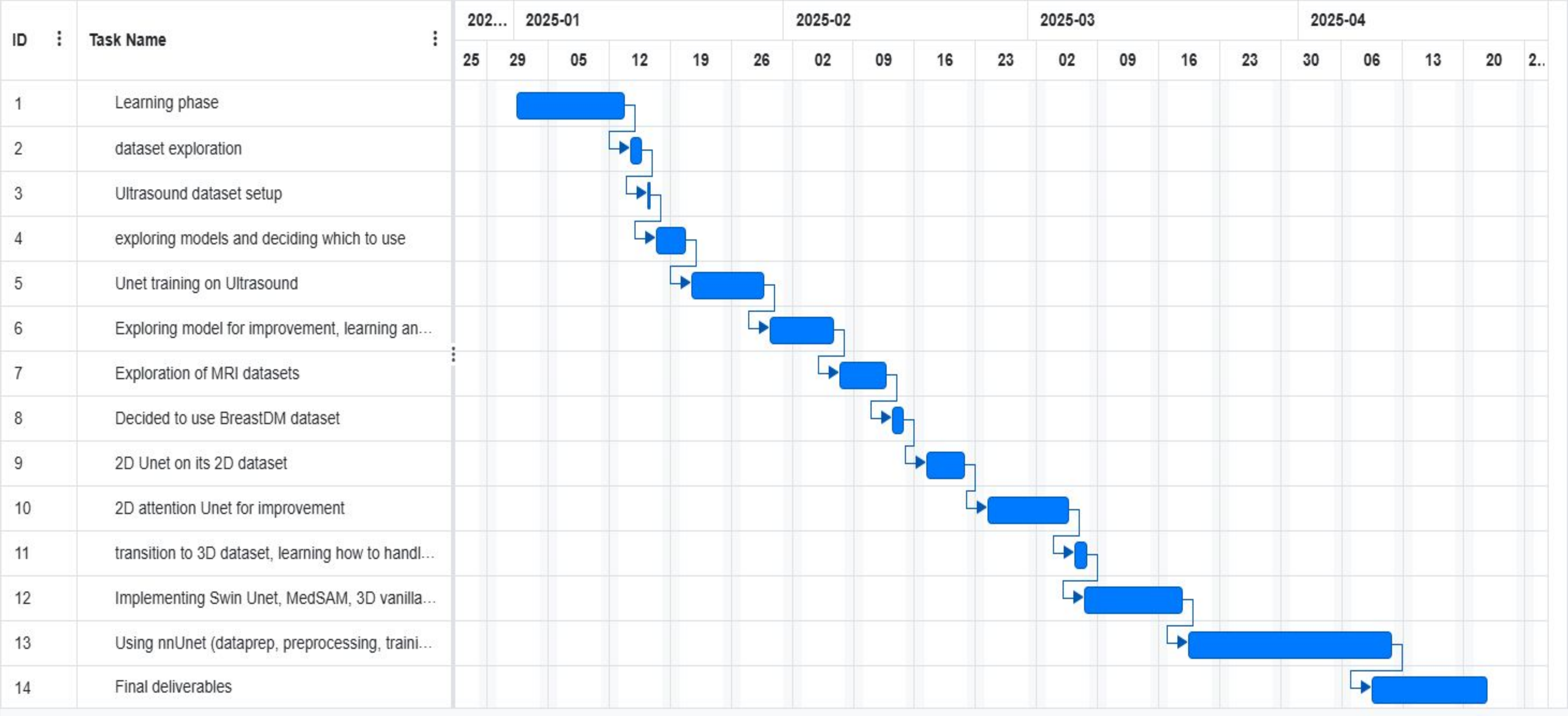
Model	DSC (%)	IoU (%)	PPV (%)
U-Net (Paper)	73.7	80.3	83.6
U-Net VGG-16 (Paper)	70.4	77.3	79.6
UNeXt (Paper)	70.1	78.5	86.3
PSPNet (Paper)	72.7	79.0	72.9
Attention U-Net	**82.5**	**92.8**	**72.1**

Ahmedabad University

My Contributions

- **Converted raw dataset into nnU-Net-compatible** format through custom scripting.
- **Checked method's applicability** across various datasets and dimensions
- **Experimented** with **multiple models** (Unet3D, Swin U-Net, MedSAM) before finalizing nnU-Net.
- **Learned** and **executed** the full nnU-Net pipeline
- Implemented **custom post-processing** to retain the largest disconnected component (due to early stopping).
- **Computed Dice, IoU, and PPV** for evaluation since pipeline used pseudo Dice.
- Developed **clear visualizations** to assess segmentation quality.

Gantt chart



ID
Task Name
202...
2025-01
2025-02
2025-03
2025-04
25
29
05
12
19
26
02
09
16
23
02
09
16
23
30
06
13
20
2..
1
Learning phase
2
dataset exploration
3
Ultrasound dataset setup
4
exploring models and deciding which to use
5
Unet training on Ultrasound
6
Exploring model for improvement, learning an...
7
Exploration of MRI datasets
8
Decided to use BreastDM dataset
9
2D Unet on its 2D dataset
10
2D attention Unet for improvement
11
transition to 3D dataset, learning how to handl...
12
Implementing Swin Unet, MedSAM, 3D vanilla...
13
Using nnUnet (dataprep, preprocessing, traini...
14
Final deliverables

Who will benefit from my work?

- **Radiologists and Clinicians**: My model reduces the **burden of manual** tumor segmentation, speeding up diagnosis and improving accuracy—especially for detecting subtle or irregular tumor structures in MRI scans.
- **Hospitals and Healthcare Systems**: This work can streamline radiology workflows and **reduce inter-observer variability**, ultimately saving time and resources. It can also be integrated with **PACS.**
- **Resource-Limited Settings**: Since my solution is lightweight and adaptable, it could be deployed in hospitals lacking specialized expertise or computational infrastructure.
- **Patients**: By enabling early and more reliable tumor detection, my system can support faster treatment planning, potentially improving survival rates.
- **Expert-limited Settings**: Automated tumor segmentation can help detect breast cancer earlier, especially in underserved areas where expert radiologists may not be available, ultimately saving more lives.