# PREDICTION OF BLADDER CANCER CURRENCE USING MACHINE LEARNING:

Bladder cancer recurrence is one of the most recurring cancers, the tumour poses a significant challenge in cancer treatment, where early prediction can guide timely clinical decisions. Medical imaging modalities like MRI and CT scans offer rich, non-invasive diagnostic information, but manual tumor annotation is labor-intensive and often unavailable. This study aims to develop an automated pipeline to predict tumor recurrence by leveraging radiomics features extracted from segmented images. Due to the absence of ground truth masks, segmentation was performed using transfer learning techniques. The extracted features were combined with clinical data to train a binary classifier that distinguishes between recurrence and non-recurrence cases.

AUTHORS

Vasireddy Satvika - 202151175

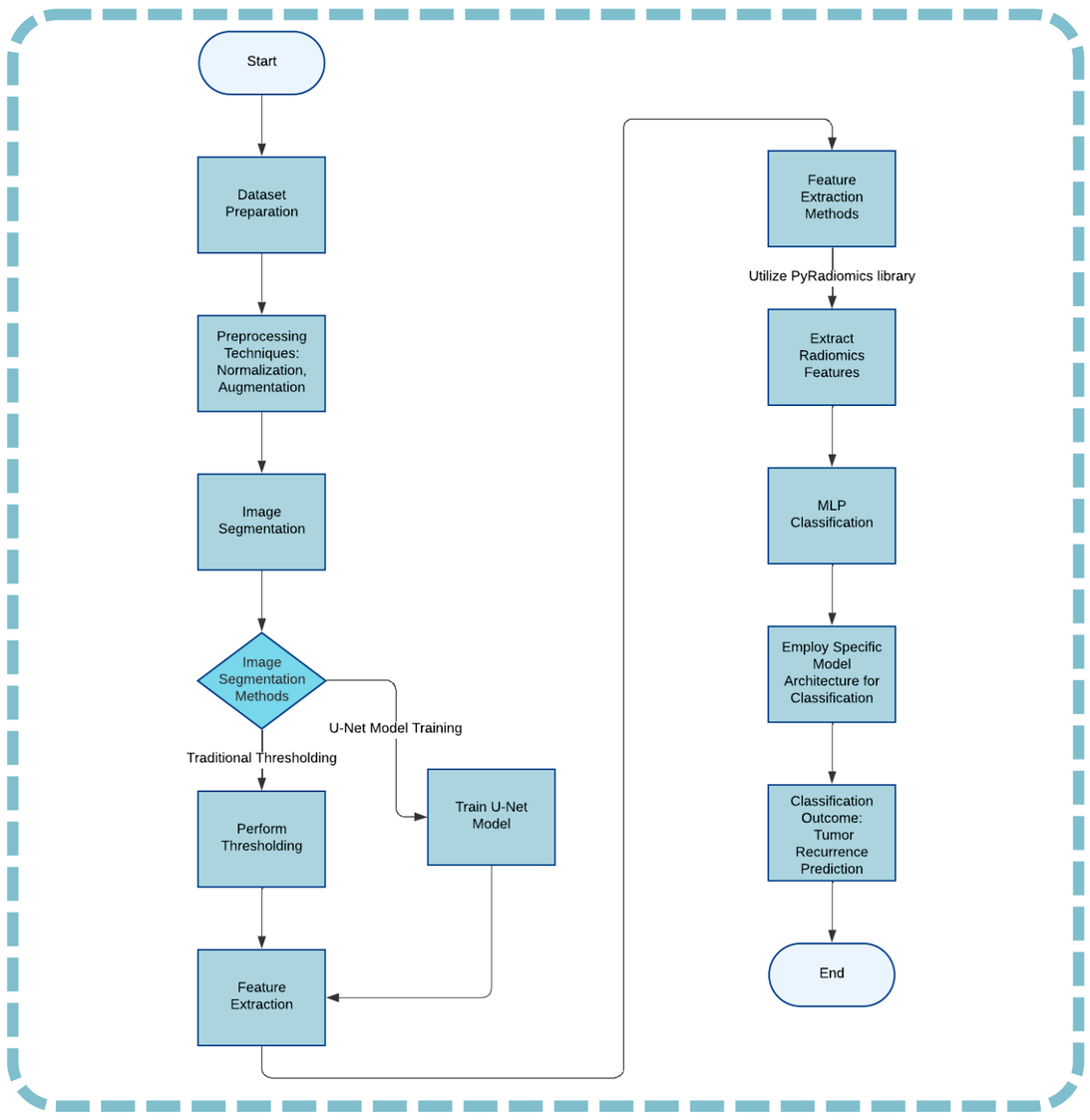
AFFILIATIONS

Research under Dr.Sanjay Saxena , IIITBHU

## METHODOLOGY

- Dataset Overview: Comprises MRI and CT scans with associated clinical data; limited size and no ground truth masks posed key challenges.
- Image Segmentation: Performed using thresholding, supervised learning, and transfer learning with UNet trained on an external dataset.
- Radiomics Extraction: Features extracted from tumor masks using PyRadiomics, including shape, intensity, and texture descriptors.
- Clinical Data Integration: Cleaned and merged relevant clinical variables with radiomics features.
- Binary Classification: Built a 3-layer MLP to predict recurrence (Yes/No) using BCE loss, Adam optimizer, and sigmoid activation.
- Model Evaluation: Used 5-fold cross-validation, tracked best fold metrics, and analyzed results via loss plots and confusion matrix.

## FLOWCHART




## RESULTS

- THE U-NET SEGMENTATION MODEL ACHIEVED APPROXIMATELY 70% ACCURACY IN TUMOR MASK GENERATION FROM MRI AND CT SCANS.
- THE 3-LAYER MLP CLASSIFIER ACHIEVED AN AVERAGE ACCURACY OF 0.67 ACROSS THE FIVE FOLDS.
- LOSS VS. ACCURACY PLOTS SHOWED CONSISTENT LEARNING.
- CONFUSION MATRIX ANALYSIS CONFIRMED BALANCED PERFORMANCE IN DETECTING RECURRENCE AND NON-RECURRENCE CASES.

## TECHSTACK

Pandas & NumPy: For data manipulation, preprocessing, and numerical computations.
OpenCV & pydicom: For processing and enhancing PNG images.
Scikit-learn: For feature selection, model training, and evaluation.
PyRadiomics: For extracting radiomics features from medical images.
TensorFlow/Keras or PyTorch: For developing the deep learning model.
Programming Language: Python
Deep Learning Model: U-Net, MLP classifier
Other Tools/Technologies:
GitHub: For accessing datasets and code
mritopng: A DICOM to PNG conversion tool