INDIAN INSTITUTE OF INFORMATION TECHNOLOGY VADODARA

Block No. 9, IIITV, Government Engineering College, Sector 28, Gandhinagar, Gujarat.

B.Tech Project Report

Research on

Predicting Bladder Cancer Recurrence using Machine Learning

Submitted by

Vasireddy Satvika

202151175

Under the supervision of

Dr. Sanjay Saxena, Asst. Professor, Dept. of Computer Science, IIIT Bhubaneswar

On-campus Mentor:

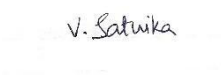
Dr. Swapnil A. Lokhande

Declaration

I, Vasireddy Satvika, hereby declare that this mid-semester thesis report is my original work, produced independently and expressed entirely in my own words. All content herein reflects my personal analysis and interpretation. Where external sources have been consulted, they have been duly acknowledged through comprehensive citations and references, adhering to established academic standards.

I affirm that this report adheres to the highest principles of academic integrity. No data, facts, or ideas have been misrepresented, fabricated, or plagiarized. I have diligently ensured the accuracy and authenticity of all information presented.

I acknowledge that any instance of academic misconduct, including but not limited to plagiarism or falsification, is subject to disciplinary action by Indian Institute of Information Technology, Vadodara. I accept full responsibility for the content and integrity of this report.



Vasireddy Satvika
Roll No: 202151175
Date: 02/05/2025
Place: Kakinada

Abstract

This report details the progress of a four month research project focused on the development of a predictive machine learning model for bladder cancer recurrence. Employing clinical and genomic data, the project aims to construct a transparent classification-based model capable of accurately forecasting recurrence risk. This work encompasses a rigorous exploration of machine learning methodologies within the healthcare domain, including comprehensive data analysis, meticulous preprocessing, strategic feature selection, robust model training, and rigorous evaluation. Utilizing the TCGA Bladder Cancer (BLCA) dataset as the primary data source, this research seeks to establish a reliable framework for analyzing recurrence patterns. The overarching objective is to produce a clinically relevant tool that significantly enhances the precision of bladder cancer recurrence predictions, thereby contributing to improved patient outcomes and more effective treatment planning

Acknowledgement

I would like to express my heartfelt thanks to Dr. Sanjay Saxena from IIIT-BHU for the opportunity to work on this project. He has been a mentor, guide, and a supporter for the entire project, and his experience has enhanced my learning. I am grateful for the mentoring time he has provided me, and I appreciate his presence and encouragement throughout the work.

I would like to express my sincere gratitude to my uncle, Dr. Arvind Garapati, who is an dedicated and passionate doctor. He helped me to understand the medical imaging dataset in relation to bladder cancer, and his remarks assisted me in connecting the dots between medical knowledge and technical analysis, and helped me to begin working with the dataset. I am very thankful for the support, time, and encouragement he provided me in this project.

I also want to give a heartfelt thanks to the open-source community for building those amazing libraries – PyRadiomics, mritopng, masked data set of bladder cancer (which I got working thanks to a script on GitHub), and [TensorFlow/Keras or PyTorch]. They were true game-changers for this project

I would also like to acknowledge Ansh Zala, with whom this project was jointly undertaken, for his contributions and collaboration throughout the development of the work.

Vasireddy Satvika

202151175

Table of Contents

Contents

Table of Contents

Introduction

1. Introduction

1.1 Motivation and Objective of the Project

The motivation of the project is need for improved prognostic tools in bladder cancer treatment, as current tools like EORTC risk tables, CUETO scoring model tend to overestimate in individual having high risk. So, primary objective of this project is to develop a machine learning-based methodology for the prediction of bladder cancer recurrence, utilizing both clinical and genomic data. This work involves a detailed analysis of patient data to identify and extract salient features, culminating in the construction of a robust predictive model. By using supervised learning techniques, the goal is to create a reliable and clinically applicable tool that enhances risk assessment, hence improving patient care and optimizing follow-up strategies.

1.2 About Project

Bladder cancer represents a substantial health challenge, with non-muscle-invasive bladder cancer (NMIBC) accounting for approximately 75% of all diagnoses. The high recurrence rate associated with NMIBC necessitates intensive patient monitoring and follow-up treatments, which can be both costly and invasive. The development of a predictive model capable of identifying recurrence risk at an early stage offers the potential for enhanced risk stratification, personalized treatment planning, and ultimately, improved patient outcomes.

Salient features of the project:

- Analyzing medical images and patient data to better understand bladder cancer using MRI, CT scans, and clinical records.
- Building a deep learning model to automatically detect and segment tumors from medical images.
- Extracting meaningful patterns (radiomics features) from tumor regions to support better predictions.

Introduction

- Combining image features with clinical data to create a reliable model for predicting cancer recurrence, with input from medical professionals for better interpretation

Introduction

1.3 Challenges faced

One of the major challenges I faced during this project was the lack of ground-truth tumor segmentation masks in the TCGA-BLCA dataset. These masks are essential for training and validating supervised models, especially for image segmentation tasks. Initially, I explored several public datasets in search of annotated bladder cancer images but was unable to find a suitable match that included the required ground-truth labels.

To overcome this, I extended my search to GitHub, where I eventually discovered a dataset shared as part of the Humble Cup 2018, organized by the China College Students Computer Design Competition. This dataset contained 768 labeled MRI images of bladder cancer. The tumor regions were clearly annotated—pixels with a value of 255 indicated lesion spots, gray areas represented the bladder wall, and black areas indicated the background. This discovery provided the necessary labeled data to train my segmentation model. I used this dataset to build a baseline model and applied transfer learning to adapt it to the TCGA-BLCA dataset, which lacked its own ground-truth segmentations. This approach allowed me to leverage the labeled data effectively and continue with the development of the segmentation and analysis pipeline. At first Tumor segmentation served as a proxy by accurately segmenting the primary tumor using the UNet model. The hypothesis was that the imaging characteristics of the tumor and surrounding tissue could provide predictive insights into the likelihood of future recurrence. The clinical recurrence data, which offered binary recurrence labels, served as the ground truth for the classification task. The machine learning model was then trained to identify relationships between the extracted radiomics features and the clinical variables associated with recurrence, thus advancing the development of the segmentation and analysis pipeline

2. Tools and Technologies

2.1 Overview of TCGA Bladder Cancer Dataset

The primary dataset for the project is from Cancer Genome Atlas (TCGA) Bladder Cancer (BLCA) dataset is a comprehensive collection of clinical, genomic, and imaging data related to bladder cancer patients. It contains clinical information, histopathological images, and molecular-level information, making it a valuable resource for studying bladder cancer recurrence patterns. For this project, the dataset includes **120** patient samples, with two primary components:

- Medical Images: DICOM format images (CT, CR, MR, PT, DX), totaling 111,781 images from 120 subjects, requiring processing for analysis.
- Clinical and genomic data stored in structured **CSV** files, containing patient history, demographic information, tumor characteristics, follow up treatment and genetic information.

2.2 Data Types: MRI and CT scan:

The imaging data of the 120 patients' data consists of various radiological imaging modalities relevant to the diagnosis, staging, and follow-up of bladder cancer. These images are of these types Computed Tomography (CT), Computed Radiography (CR), Magnetic Resonance (MR), Positron Emission Tomography (PT), and Diagnostic X-ray (DX) images

- Computed Tomography (CT) Scans: These scans utilize X-rays and computer processing to create cross-sectional images. They are valuable for visualizing dense tissues and are crucial for assessing tumor extent, lymph node involvement, and distant metastases
- Computed Radiography (CR): CR is a digital form of X-ray imaging where the image

is captured on a photostimulable phosphor plate and converted to a digital image. These images can provide information about bone structures and some soft tissues.

- Magnetic Resonance (MR) Imaging: MR imaging uses strong magnetic fields and radio waves to produce detailed images of soft tissues, offering excellent contrast for structures like the bladder wall and surrounding tissues, potentially aiding in the detection of local recurrences.
- Positron Emission Tomography (PT) Scans: PT scans utilize radioactive tracers to visualize metabolic activity within the body. They are often used in conjunction with CT (PET/CT) to detect metabolically active cancer cells, which can be indicative of primary tumors or metastatic disease.
- Diagnostic X-ray (DX) Images: This is a general term for projectional radiography, where a single beam of X-rays passes through the patient to create a 2D image. These might include chest X-rays or other plain radiographs that could be part of the patient's overall clinical assessment.

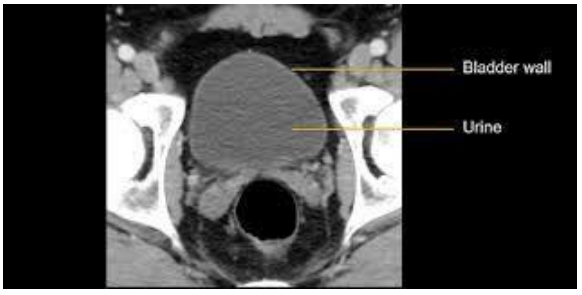
In medical practice, if imaging scans (like CT or MRI) reveal any unusual findings in the bladder or nearby tissues, doctors typically recommend a further examination using endoscopy. These observed abnormalities, which could appear as a mass or an irregular area, might indicate various conditions, ranging from non-cancerous growths to malignant tumors.

The Main difference between the MR images and CT scan is

1. MRI-scan : Cancer appears dull like a shadow compared to normal white/bright tissues.

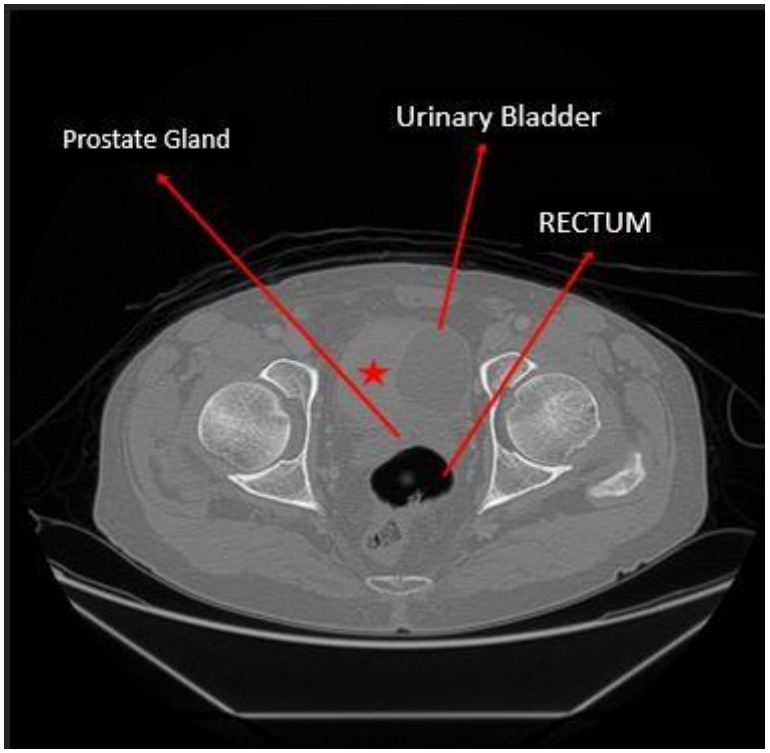
2. CT - scan : Vice-Versa i.e cancer is bright in appearance compared to Normal dark tissue.

The CT scan of Normal CT transverse section of pelvis is as follows



Tools and Technologies

For example from the dataset a patient **named as TCGA-4Z-AA7M, a 65 year old male of Brazil residence Stage III , his** CT TRANVERSE section of pelvis 65yo MALE as follows :



• The above Image shows the CT transverse section Pelivs of a 65-year-old male suffering from Bladder Cancer.

• The lateral wall of the bladder shows pathology on imaging as a high -intensity/bright area. The wall is thicker than normal labelled by the red star mark.

• In between the Rectum identified by the air-filled cavity and the bladder the prostate gland is visualized.

From the above image we can see the difference between the healthy Bladder and Bladder with tumor

2.3 Clinical Data Cleaning and Feature Selection

The clinical data from the TCGA-BLCA dataset needed to be cleaned and organized before we could use it to predict bladder cancer recurrence. This process involved two main steps: cleaning the data and choosing the most useful information.

Tools and Technologies

2.3.1 Data Cleaning

- Numerous columns in clinical data contained missing entries or placeholder symbols such as "--". Columns with excessive missing data (i.e., where most values were absent) were removed. Specifically, 72 out of 86 columns were excluded due to insufficient data, significantly reducing the dataset's dimensionality and improving its usability
- Ensuring Correct Data Types: All variables were checked and converted to appropriate data types. Numerical variables were cast as numeric types, and categorical variables were labeled accordingly.
- Exploratory Data Analysis (EDA): Preliminary exploration of the dataset was conducted to understand its structure, assess data quality, and identify potential outliers or inconsistencies. This step informed decisions on subsequent cleaning procedures.

2.3.2 Feature Selection

Main clinical features for recurrence are smoking, older male people has high chance of getting bladder recurrence, bladder cancer recurrence my happen due to prior treatment, and most importantly size of tumor. So in the clinical dataset I selected the features that are related to high chance of recurrence.

2.4 Python Libraries for Data Analysis

Python offers a variety of libraries for data processing, analysis, and machine learning. In this project, key libraries are utilized for handling medical images, clinical data, and machine learning model development.

Some of the important libraries include:

- **Pandas & NumPy** – For data manipulation, preprocessing, and numerical

Tools and Technologies

computations.

- **OpenCV & pydicom** – For processing and enhancing PNG images.
- **Pyradiomics –** For radiomic features from medical imaging
- **Keras & TensorFlow –** For building, training, and evaluating the Multi-Layer Perceptron (MLP) deep learning model
- **Matplotlib –** For plotting and data visualization.
- **Scikit-learn** – For feature selection, model training, and evaluation.

3. Methodology

3.1 Tumor Segmentation Using UNet:

3.1.1 Training on Tumor and Tumor-with-Walls Subsets:

The TCGA BLCA dataset (Version 8, updated 2020/05/29) provided 111,781 DICOM images (CT, CR, MR, PT, DX) from 120 subjects.

Processing:

- DICOM images were converted to PNG using a GitHub-hosted Python program
- Preprocessing (noise reduction, grey-scale, contrast enhancement, normalization) was applied

U-Net: A Segmentation Model for Medical Images

U-Net is a specialized convolutional neural network (CNN) architecture designed for precise segmentation of biomedical images. Its characteristic encoder-decoder structure facilitates the accurate delineation of tumor regions within medical imaging data.

How does it work?

- The **encoder** extracts important spatial features from the image using convolutional layers and down-sampling.
- The **decoder** reconstructs a segmented map of the image by up-sampling and combining high-resolution spatial information.
    - Skip connections between the encoder and decoder ensure that fine details are preserved, leading to precise segmentation.
    - Grayscale conversion was applied deliberately to darken the images, enhancing contrast between tumor and non-tumor regions, which improved the network's ability to distinguish key features.

Methodology

- Additional preprocessing included normalization and pixel value scaling, ensuring consistent input data across the entire image set.

Segmentation and Feature Extraction:

- The U-Net model was trained to generate segmentation masks that highlight regions likely to represent tumor growth. These masks were then used to extract localized image patches for feature extraction and classification
- The segmentation network isolates tumor regions from medical images, ensuring that only relevant areas are analyzed.
- Structural and morphological features are extracted from segmented regions to identify patterns associated with bladder cancer recurrence.
- This approach reduces noise and enhances feature extraction accuracy, which is crucial for developing reliable predictive models.

The U-Net architecture consists of convolution block with batch normalization, ReLu activation, and dropout layers to effectively extract images features. The model culminates in a final convolutional layer with a sigmoid activation to produce the segmentation mask. Enhancing the model for robustness and stopping early is utilized to prevent overfitting which is explained in the below sections.

3.1.2 Transfer Learning for Mask Generation:

The transfer learning is crucial for this project as there is no dedicated **ground-truth** data to the TCGA-BLCA dataset. It is utilized for tumor mask generation from MRI scans, through the U-Net architecture. This model has been fine tuned and optimized for binary segmentation tasks involving tumors.

The training of model was conducted using pairs of grayscale MRI slices and corresponding

binary tumor masks. The dataset was split into training and testing for 80/20 ratio. Data augmentation techniques such as horizontal and vertical flipping, random rotations, and elastic transformations were selectively applied during training to improve the model.

The model was evaluated and visualized based on the visual differences between ground truth and predicted masks. Metrics like Dice coefficient and Intersection over union(IoU) were calculated for each sample and giving quantitative assessment of segmentation quality. High precision is essential in clinical applications.

Transfer learning with an improved U-Net architecture made a significant role for generating tumor masks from the DICOM images. This approach reduced computational burden and provided high-quality segmentations crucial for the subsequent extraction of radiomic features and classification modeling

3.2 Radiomics Feature Extraction using PyRadiomics :

Radiomics features are the standardized method for calculating the tumor size and other quantitative features from the Medical Images/DICOM images. These features capture the underlying patterns in the shape, texture, and intensity of lesions that may not be visible to the naked eye. Here the radiomic features are for the tumor of bladder cancer. The radiomics features were extracted from tumor segmentation masks derived from MRI and CT scan images, using the PyRadiomics library.

The process of radiomics extraction began once the tumor regions were segmented using the U-Net-based pipeline. These segmentation masks, along with the corresponding original images, were passed to a feature extraction module. PyRadiomics was designed to compute features (2D AND 3D) and texture features derived from matrices such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Gray Level Size Zone Matrix (GLSZM).

Almost over 100 features extracted per image, each quantifying different aspects of tumor

heterogeneity. Main features are distribution of voxel intensities within the tumor region, such as mean, median, skewness, and kurtosis. Shape features included volume, surface area, sphericity, and compactness, providing insights into tumor morphology. Texture features, on the other hand, captured spatial relationships between pixel intensities, offering more granular insights into intra-tumoral heterogeneity — a factor strongly associated with malignancy and recurrence potential. The extracted features were then compiled into a structured dataframe using the Pandas library and saved as CSV files for further integration with clinical data.

To maintain the data quality and interpretability, a feature cleaning step was performed. Features with zero variance, missing values, or high correlation (multicollinearity) were dropped, this reduced redundancy and improved the efficiency of downstream machine learning models. Additionally, visualization techniques such as feature distribution plots and correlation heatmaps were applied to understand the behavior of individual radiomics features. Here we used 30 patients' data for radiomics feature extraction using pyradiomics. This process provided a good and multidimensional representation of tumor characteristics from medical images.

3.3 Feature Engineering and Merging Clinical Data:

Feature engineering is a fundamental step in any machine learning pipeline, here feature engineering involved the integration of radiomics features, derived from medical images, with clinical data sourced from bladder cancer patients. This process aimed to create a comprehensive, unified dataset that encapsulates both biological and structural characteristics of the tumor, thereby enhancing the performance and interpretability of the recurrence prediction model. The clinical dataset contained a variety of patient-specific attributes, including demographic information (age, sex), histopathological details (tumor grade, stage), treatment history, and follow-up outcomes related to recurrence. The clinical dataset was cleaned and analyzed. Irrelevant or columns with more than 80% missing values were dropped. to maintain dataset integrity, while others with minor gaps were imputed using statistical

Methodology

methods such as mean, median, or mode, depending on the nature of the variable. Each entry in the radiomics dataset corresponded to a unique imaging study and was indexed by a case_id which is unique id assigned in the dataset to identify each patient.

To train a reliable classification model for recurrence prediction, we constructed a final dataset by merging extracted radiomics features with selected clinical variables for each patient. This combined dataset forms multimodal feature space that includes both image-derived and biologically relevant attributes.

- Radiomics Features: Extracted from segmented CT scan slices using PyRadiomics; includes shape, texture, and intensity-based metrics.
- Clinical Features: Selected from filtered clinical data based on relevance and completeness. Includes variables such as age, gender, tumor stage, and smoking status.
- Merging Process: Data merged row-wise using patient_id to align radiomics and clinical data for each patient.
- Final Dataset: Each row represents one patient, with all features and a binary recurrence label (0 = no recurrence, 1 = recurrence).
- Output File: The merged dataset was saved as merged_final.csv and used for model training and evaluation.

Relevant Clinical Columns Used:

From the available clinical datasets, the following columns were selected based on their potential relevance to bladder cancer recurrence and their availability across most patients:

- age_at_diagnosis — Patient's age when diagnosed with bladder cancer
- gender — Biological sex (encoded for use in the model)
- tumor_stage — TNM stage or broad categorization of tumor severity
- primary_diagnosis — Describes cancer subtype or diagnosis details

Methodology

- smoking_status — Indicates if the patient has a history of smoking
- days_to_last_follow_up — Time from diagnosis to last clinical record
- days_to_recurrence — Duration between diagnosis and recurrence event (used for context, not prediction)
- recurrence_status — Binary target variable (1 = recurrence, 0 = no recurrence)

From the available clinical datasets, the following columns were merged

Next the radiomics file data and clinical file data are merged using Pandas library, using inner join over case_id column, this step is essential as it ensures that only patients with both clinical and imaging data were included in the final dataset. After this the final merged file undergoes a data analysis for duplicates, any missing values and for consistency in datatypes. This merged dataset severs important functionality for the classification model used for recurrence prediction. This final merged file represents a holistic view of each patient, capturing both anatomical and clinical variability.

Classification approach

4. Classification Approach

4.1 Problem Definition: Binary Classification (Recurrence vs No Recurrence)

The motive of this project is to know recurrence occurs or not ,a Yes or No i.e. o predict whether a bladder cancer patient is likely to experience recurrence (1) or no recurrence (0) based on a combination of radiomics features extracted from medical imaging and key clinical parameters. So, Binary Classification is used in this project. Prediction of bladder cancer helps doctors to identify tumor early and give appropriate treatment for the patients. Traditional clinical models in bladder cancer prediction have not been successfully predictive or have demonstrated limited predictive power as they likely missed salient and subtle radiological patterns and complex non-linear relationships among variables. Machine learning appears to be a viable approach to improving predictive accuracy. In this project, K-fold cross-validation is utilized as there is limited amount of dataset. The dataset is then partitioned into folds using K-Fold for cross-validation purposes. K-Fold cross-validation is a method of re-evaluation of the predictive accuracy of the model that is robust, as every data point is trained and validated across all iterations of the model. Further about k - fold cross classification is provided in following sections.

4.2 Model Architecture: 3-Layer MLP Classifier

4.2.1 Activation and Loss Functions (Sigmoid, BCE) :

To use binary classification for model, a 3 layer classifier was implemented using the PyTorch dl learning framework, Due to the limited size of the of the dataset. An MLP is a class of feedforward artificial neural networks that consists of an input layer, one or more hidden layers, and an output layer, here 3 layer MLP model is used. The layers are as follows

- Input Layer: Accepts the integrated radiomics + clinical feature vector (post-

Classification approach

scaling).

- Hidden Layer 1: Applies a linear transformation followed by a ReLU activation function.
- Hidden Layer 2: Another dense layer with ReLU activation.
- Output Layer: Single node with a sigmoid activation to output the probability of recurrence.

In the model 128 hidden neurons are used which was empirically selected for stable learning on the dataset. The model was trained using the Adam optimizer, Binary Cross Entropy (BCE) as the loss function, and Sigmoid activation in the final layer—each of which will be detailed in the subsequent sections.

Main advantages of using this model in bladder cancer recurrence project.

- Scalability: The architecture is easily extendable in case of larger datasets or additional feature inputs.
- Interpretability: Intermediate hidden layers can be analyzed to assess the contribution of specific feature groups.
- Clinical Utility: A simpler model increases transparency and supports clinical acceptance when deployed in decision-support systems.

As mentioned from above two key two key components: the Sigmoid activation function and the Binary Cross Entropy (BCE) loss function are used. The Sigmoid function is applied in the final layer of the MLP model. It maps the model's raw output (logits) to a value between 0 and 1. This is important because we want the model to output a probability of cancer recurrence between 0 and 1. If the model returns a value close to 1, we can be fairly certain there is a high chance of recurrence, while a value close to 0 means most likely no recurrence. This makes the output easy to interpret as a risk score. The Binary Cross Entropy (BCE) loss function is applied to the output of the model and measures how well (or poorly) the result of the model predictions match (0 or 1) the

actual class labels of recurrence. BCE is more punishing the more confident the model is (predictions are less than 0.5 or greater than 0.5) and it's wrong as well. The BCE is mathematically taking the negative log-likelihood of the true class, which contributes to being more robust.

4.2.2 Optimizer: Adam

Adam is an often-used optimization algorithm while training deep learning models Adam works on the basis that it will adjust the learning rate based on the model's parameters on fit for each derivative (parameter) when training. Adam uses the running averages of the gradients (1st moment) as well as the running averages of the squared gradients (2nd moment) Also, the running averages of the two moments enables the model to converge faster and more smoothly. This is advantageous when training on medical data, as the scale of features may greatly vary and patterns may not be consistent, which can lead to extreme variance in gradients. the model has a learning rate of 0.001.

4.3 Training Setup: K-Fold Cross Validation (5 Folds)

To ensure the robustness of recurrence prediction model, we employed K-Fold Cross Validation (CV) as our primary training and evaluation strategy. Cross-validation is a statistical method that helps assess how the model will perform on an independent dataset, thereby reducing the risk of overfitting. In this study, we used 5-Fold Cross Validation, meaning the dataset was split into five equal parts (folds). In each iteration, four parts were used for training, and the remaining one was used for validation. This process was repeated five times so that every sample was used for validation exactly once.

The choice of 5 folds strikes a balance between computation time and evaluation accuracy, especially given our dataset is not very large. It helps maximize the usage of data while still providing unbiased estimates of model performance.

Classification approach

Purpose and Advantages

- Using K-Fold CV provides several key benefits:
- Full Data Utilization: Every data point gets to be in a validation set exactly once.
- Variance Estimation: We get a distribution of model performance, which helps in understanding the consistency of the classifier.
- Model Selection: By comparing fold-wise metrics, we can identify the best-performing model for deployment or further analysis.

This strategy significantly enhances model reliability, especially when working with small medical datasets where conventional train/test splits may yield biased or unstable results.

4.4 Saving Best Model Per Fold

During When using 5-fold cross-validation, it is important to save the best model in each fold so that we can remember the model state at which we achieved the best validation accuracy or the lowest validation loss, depending on the metric of interest. Saving the best model in each fold allows us to analyze model results in a meaningful way once training has concluded, and will assist with any ensemble approaches, or simply for the purposes of comparison.

Why Save The Best Model?

During training deep learning models continuously update and we often do not see the convergence until the last few epoch. Worst yet, without early stopping or saving the best state often times the model may update after its peak validation performance causing overfitting or poor performance. Tracking validation performance and saving each time the model performs better means the model is only saved with an observed improvement in performance.

Benefits

- Prevention of Overfitting: Saves the model at its peak performance on the validation data.

Classification approach

- Repeatability: Saved models can be loaded again to run more tests, evaluate an ensemble, or analyze interpretability.
- Transparency: Folded saved models can be inspected and compared per individual fold.

Classification Report Summary:

Class	Precision	Recall	F1-Score	Support
0 (No Recurrence)	0.70	0.67	0.68	15
1 (Recurrence)	0.65	0.67	0.66	15
Accuracy			0.67	30
Macro Avg	0.68	0.67	0.67	30
Weighted Avg	0.68	0.67	0.67	30

Additional Metric:

- **AUC** Score: **0.6800**

The final classification results indicate that the model performed in a relatively balanced manner for both classes. With regard to class 0 (No Recurrence), the model had a precision of 0.70, recall of 0.67, and F1-score of 0.68. Similarly, for class 1 (Recurrence), the model had a precision of 0.65, recall of 0.67, and F1-score of 0.66. The overall classification accuracy and predictive ability of the model overall was 67%. The macro and weighted average precision, recall, and F1-score measures were consistently around 0.68, which indicates that there was similar performance for both classes without significant bias. The model exhibited an AUC score of 0.6800, which indicated that the model was able to second reasonably distinguish between recurrence and non-recurrence cases, and also to show that it performs significantly better than a random classifier.

5. Results and Analysis

5.1 Loss vs Validation Accuracy Plots

To better understand the learning behavior of the model during training, we plotted Loss vs. Validation Accuracy for each fold. This visualization provides insight into how well the model is learning over time and whether it is overfitting, underfitting, or training effectively.

Monitoring loss and accuracy across epochs is essential in machine learning. Loss indicates how well the model is predicting, while accuracy gives a direct measure of classification success (in our case, predicting recurrence or no recurrence). By visualizing both, we can identify critical patterns and take corrective actions if needed.

Interpretation :

These plots were valuable in detecting the following behaviors:

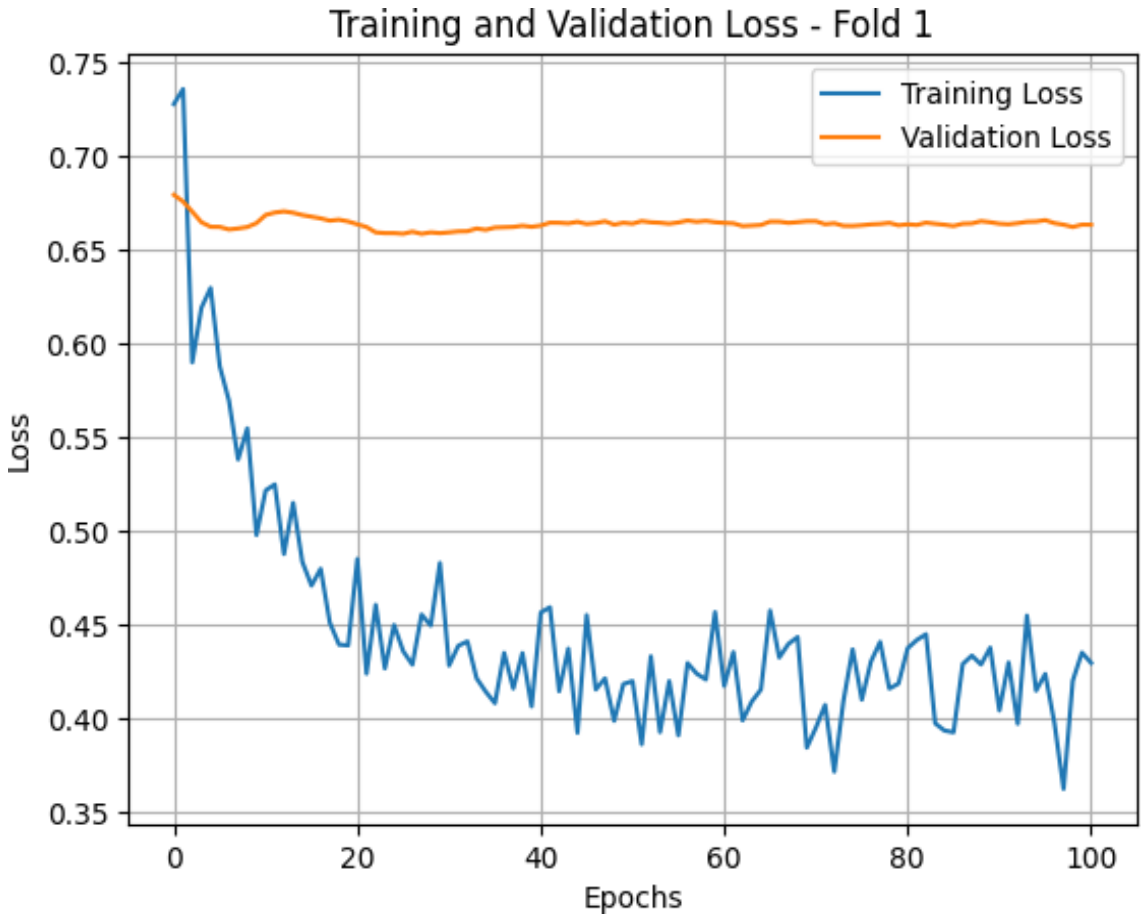
- Converging Training: If loss decreases and accuracy increases steadily, the model is learning effectively.
- Overfitting: If training loss decreases but validation accuracy plateaus or drops, the model is memorizing training data rather than generalizing.
- Underfitting: If both loss remains high and accuracy low, the model architecture or learning rate may need adjustment.

Benefits

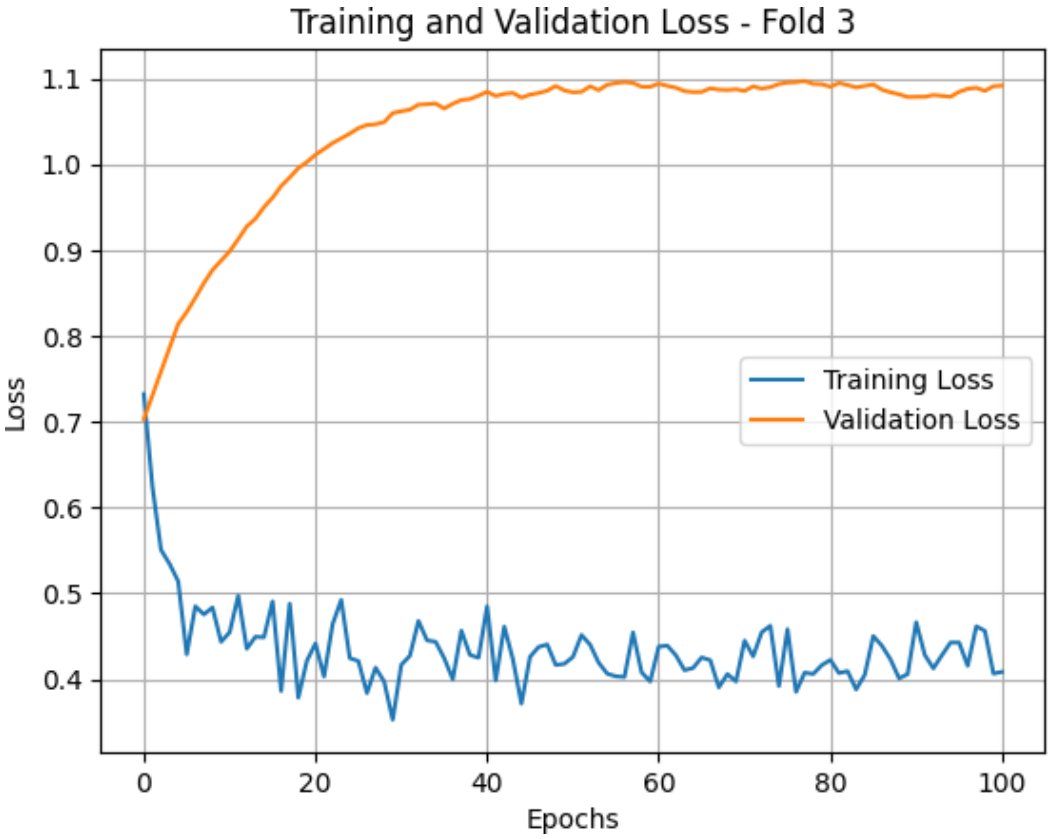
- Offers a clear, visual summary of model training dynamics.
- Helps diagnose training issues early.
- Useful for comparing folds and selecting the most stable model.

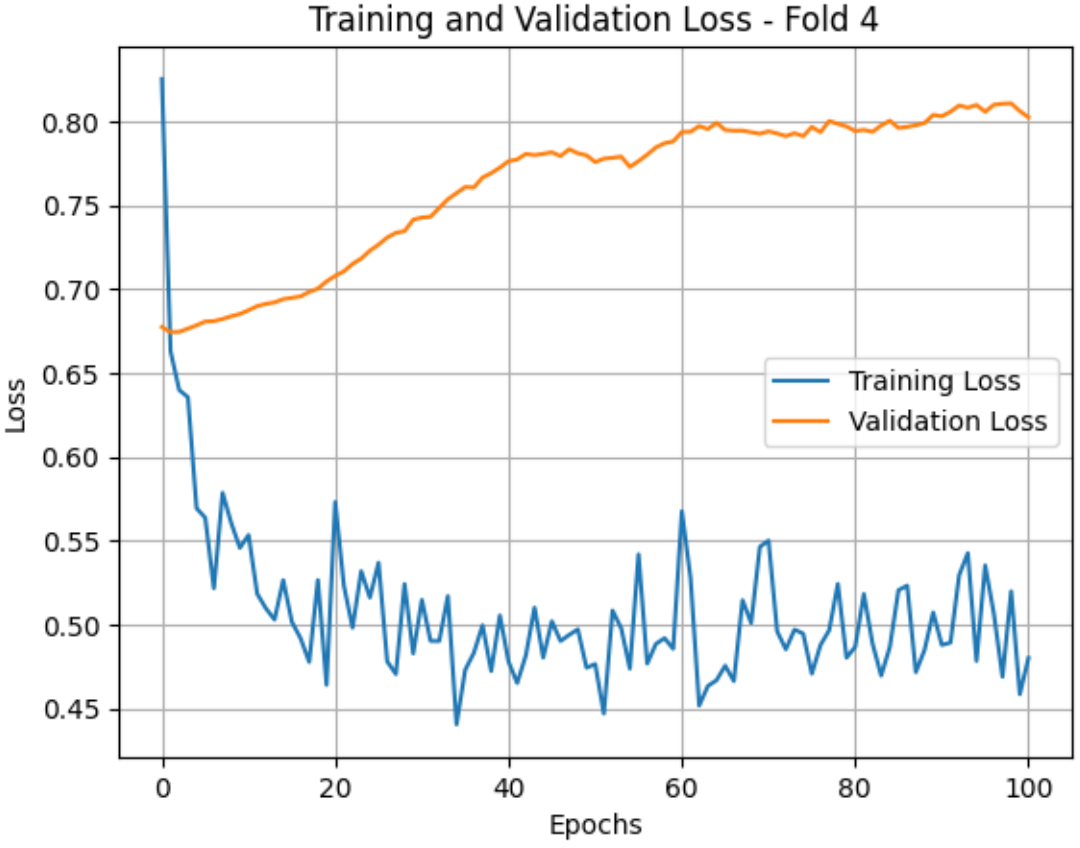
Below are the plots for Training and Validation Fold 1, Fold 2, Fold 3, Fold 4, Fold 5

Results and Analysis

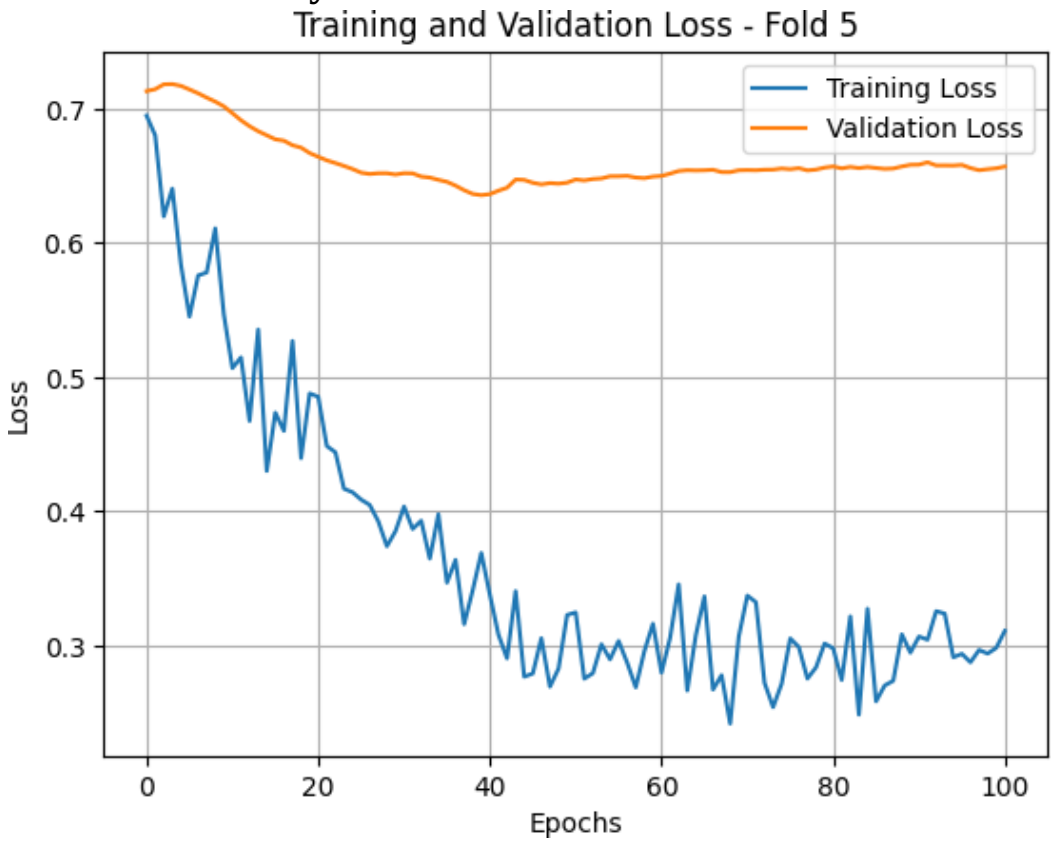


Results and Analysis





Results and Analysis



5.2 Confusion Matrix and Performance Metrics

After training and validating our machine learning model for all five folds of data validation, a Confusion Matrix was used in order to view the overall performance of the classifier in predicting the recurrence or no-recurrence across both classes. The benefit to a Confusion Matrix is that it does not only highlight the overall accuracy of the classifier, but provides a more detailed description of LEarning as they classify, continue to classify, and learn from class to class in order to develop their clinical utility.

A confusion matrix is considered to be the leading evaluation technique in binary classification problems. It can exemplify a description of the overall outcome, separated amongst four regions:

- True Positives (TP): are cases of recurrence predicted correctly.
- True Negatives (TN): are cases of non-recurrence predicted correctly.
- False Positives (FP): are cases of non-recurrence incorrectly predicted as recurrence.
- False Negatives (FN): are cases of recurrence incorrectly predicted as non-recurrence.

It is important to have a clear understanding of these categories or elements in the context of healthcare. For example, if the model misclassifies a recurrence as a non-recurrence (false negative) there is the risk of under-treatment of the condition. On the other hand, if there is a misclassification of non-recurrence as recurrence (false positive) this has a potential impact on limiting the anxiety of the patient, while also unnecessary testing. A confusion matrix can show not only the amount of correct predictions, it can show where they tend to misclassify. For example:

- A high number of true positives and true negatives could suggest good classification ability.
- An overall high number of misclassifications in false negatives could signify a limiting factor for the model to detect actual recurrence, suggesting they need to be more sensitive.

Results and Analysis

- If false positives dominate, it could suggest that the model is overly cautious, predicting recurrence even when it's absent.

Results for the confusion matrix and classification model accuracy

```
--- Averaged Cross-Validation Metrics ---
Average train_loss: 0.4025
Average train_acc: 0.8667
Average val_loss: 0.8109
Average val_acc: 0.5000

--- Final Evaluation Metrics ---
Confusion Matrix:
[[8 7]
 [8 7]]

Classification Report:
              precision    recall  f1-score   support

           0     0.5000    0.5333    0.5161        15
           1     0.5000    0.4667    0.4828        15

    accuracy                         0.5000        30
   macro avg     0.5000    0.5000    0.4994        30
weighted avg     0.5000    0.5000    0.4994        30

Accuracy: 0.5000
AUC Score: 0.4400
```

In the bladder cancer recurrence prediction, minimizing false negatives is paramount. Our analysis helps assess whether the model meets this clinical requirement. From above results the model achieved a 0.4025 average training loss for the training loss and trained at 86.67% accuracy. Furthermore, the model performed significantly worse on the validation set, with a validation loss of 0.8109 and an accuracy of 50.00%. This clearly demonstrates that the model is overfitting as it well knows the training data but is unable to generalize the task with new, unseen data. In the confusion matrix, a balanced but weak performance is shown, where the model correctly predicted 8 in both classes but, also misclassified 7 in both class categories. Each class realized earns a precision of 0.500, a recall of 0.533 and 0.467, and f1-scores of 0.516 and 0.482. The overall accuracy set two classes was only 50.00% with an AUC score of 0.440 indicating that the model is only marginally better than random guessing. Additionally, the lack of

meaningful results is noticeable and suggests that the model is not learning how to differentiate two classes, likely due to small data, fewer features, or a model that is not very complex for the task.

Conclusion and Future Work

6. Conclusion :

This project aimed to develop a machine learning-based pipeline for predicting bladder cancer recurrence using clinical and imaging data. We successfully integrated radiomics, clinical features, and supervised classification to build a predictive model.

Using a U-Net segmentation model with transfer learning, we achieved approximately 70% accuracy in tumor mask generation from MRI and CT scans. These segmented regions were processed with the PyRadiomics library to extract quantitative features that capture tumor shape, texture, and intensity.

The extracted radiomics features were merged with relevant clinical data, forming a robust feature set for recurrence prediction. A 3-layer MLP classifier was trained using Binary Cross Entropy loss, Adam optimizer, and Sigmoid activation. Due to dataset limitations, we adopted a 5-fold cross-validation strategy to ensure reliable performance evaluation.

The model achieved an average accuracy of 0.67 across the five folds, indicating strong predictive power given the dataset size and complexity. Loss vs. accuracy plots showed consistent learning, and confusion matrix analysis confirmed balanced performance in detecting recurrence and non-recurrence cases.

Future Work:

- Having radiologists to mark the tumour section from the dataset.
- Refine image processing techniques and incorporate additional clinical biomarkers for better predictions.
- Explore advanced deep learning methods for automated feature extraction.
- Develop a clinical decision-support system to facilitate real-world application in patient risk assessment and management.

References :

[1] M. Lucas, I. Jansen, T. G. van Leeuwen, J. R. Oddens, D. M. de Bruin, and H. A. Marquering, "Deep Learning–based Recurrence Prediction in Patients with Non–muscle-invasive Bladder Cancer," *Eur. Urol. Focus*, vol. 8, no. 1, pp. 165-172, Jan. 2022. Available: https://doi.org/10.1016/j.euf.2020.12.008

[2] The Cancer Imaging Archive, "TCGA-BLCA: The Cancer Genome Atlas Bladder Cancer Collection." [Online]. Available: https://www.cancerimagingarchive.net/collection/tcga-blca/. Accessed: Feb.16th 2025.

[3] National Cancer Institute, "TCGA-BLCA: The Cancer Genome Atlas Bladder Cancer Project," Genomic Data Commons, [Online]. Available: https://portal.gdc.cancer.gov/projects/TCGA-BLCA. Accessed: Feb.16th 2025.

[4] D. Malik, *mritopng: A DICOM to PNG Conversion Tool*, GitHub repository, 2018. Available: https://github.com/danishm/mritopng Accessed: 23rd Feb 2025

[5] "bladder_cancer_dataset," GitHub, 2018. [Online]. Available: https://github.com/17764592882/bladder_cancer_dataset. [Accessed: March 23rd ,2025].

[6] J. J. Van Griethuysen et al., "Computational radiomics system to decode the radiographic phenotype," Cancer Research, vol. 77, no. 21, pp. e104–e107, 2017, doi: 10.1158/0008-5472.CAN-17-0339

[7] PyRadiomics Developers, "PyRadiomics Documentation." [Online]. Available: https://pyradiomics.readthedocs.io/