Predicting Renal Function Post Nephrectomy for Renal Masses

NYU Center for Data Science

Group Members: Chen Chen, Siri Desiraju, Anna Dominic, Julia Manasson

Mentor: Dr. Madhur Nayan, MD, PhD

Abstract

Renal masses often require removal by partial or radical nephrectomy, which inevitably results in removal of healthy tissue and has the potential for kidney function impairment. In this study we introduce a predictive model that integrates the type of surgical intervention with preoperative patient characteristics to predict short-term postoperative renal function. Features for the model were extracted from structured and unstructured electronic health record (EHR) data, and extensive data cleaning and preprocessing performed to ensure quality and biological relevance. SHAP was used to select relevant features and postoperative renal function was modeled using linear and XGBoost approaches a MSE accuracy of 80%. With additional improvements, the model could be used as a tool to predict how well a patient will do several weeks after surgery, aiding in appropriate treatment selection for each individual patient, and allowing clinicians to prepare for complications at an early stage.

Introduction

A significant proportion of renal masses are malignant, necessitating their removal. In 2023 there were 81,800 new cases and 14,890 deaths from renal cancer in the United States[1](Siegel RL, et al. 2023). Two surgical approaches for resecting a kidney mass are partial nephrectomy (PN) and radical nephrectomy (RN). Both can compromise renal function by removing healthy renal tissue in addition to the mass. Since kidneys are vital organs, serving numerous functions to maintain homeostasis, predicting post-surgical renal function is essential for selecting the optimal treatment strategy for an individual patient.

Building on prior studies[2](Roussel E, et al. 2023; Bhindi B, et al. 2019; Aguilar Palacios D, et al. 2021; Schmid M, et al. 2014; Chan VW, et al. 2022), our project aims to utilize linear and XGBoost machine learning models to predict post-nephrectomy renal function. Leveraging EHR data, we extracted patient-centric features related to demographics, medical and social history, preoperative renal function, and surgical intervention. After extensive data cleaning and preprocessing, we built models that predict short-term postoperative renal function, which has the potential to provide valuable insights for clinical decision-making.

Building a Cohort

Creating a reliable cohort for our analysis involved intricate decisions and meticulous steps. One pivotal aspect was defining the postoperative timeframe for the GFR cohort. Experimenting with various periods, we settled on a two-week interval, supported by a robust sample size of 549 patients and aligning with the clinical expectation of recovery stability within this timeframe.

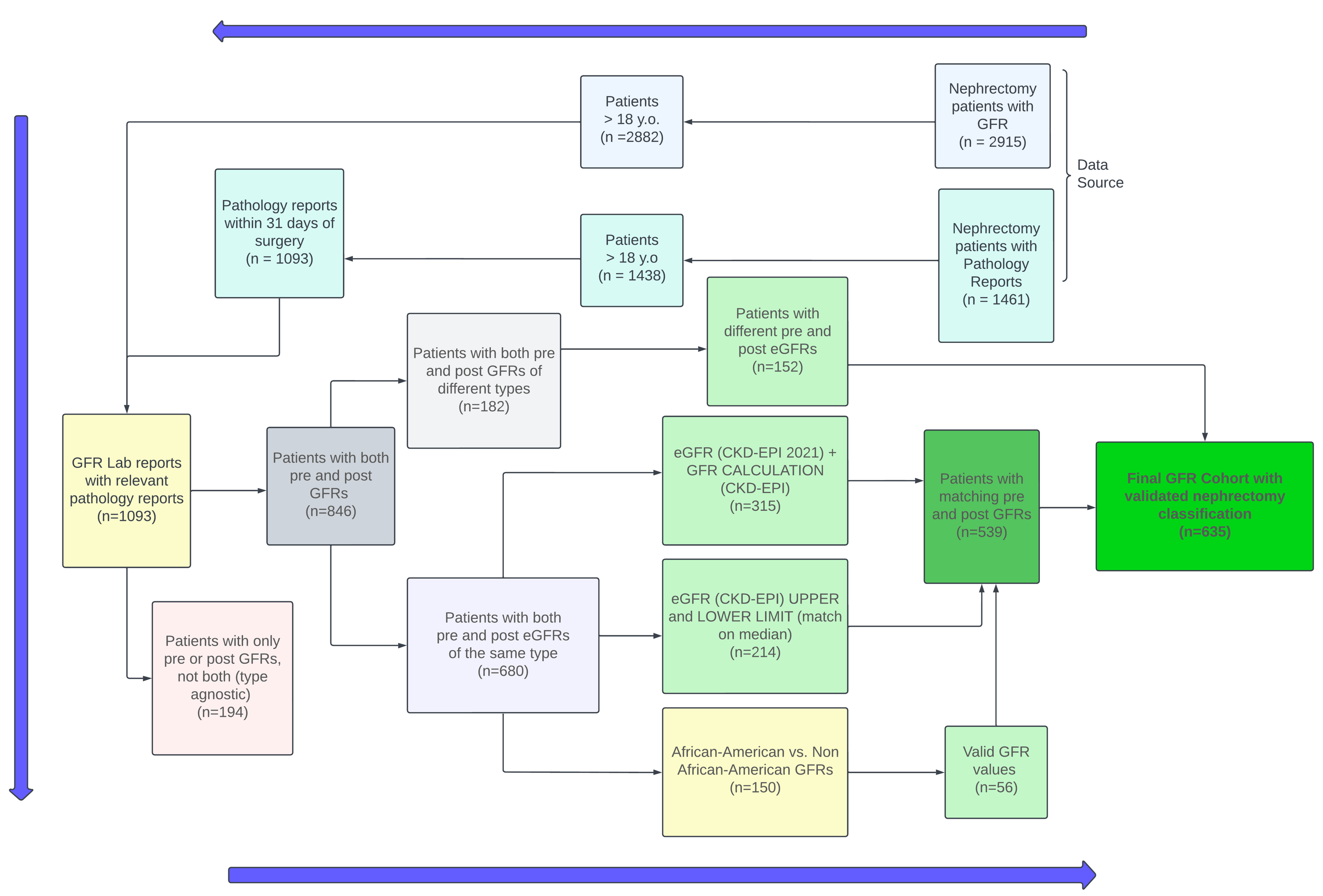


Figure 1. Flowchart for subsetting the GFR cohort

Preprocessing Methods and Challenges with EHR Data

Our preprocessing started with extracting relevant from a data lake on a remote server through SQL queries that select records relating to nephrectomy patients. This process entailed:

1. Defining the patient cohort by applying exclusion criteria: excluding individuals under 18, those diagnosed with End-Stage Renal Disease (ESRD), or missing preoperative or postoperative GFR/creatinine values.
2. Validating recorded procedure names (i.e. partial or radical nephrectomy) in EHRs with the details in pathology reports for accuracy.
3. Expand the dataset by retrieving and joining with other features(patient demographics, diagnoses of comorbidities and social history) from the data lake.

The unique challenges of EHR data included inaccuracies that warranted validation from various sources, inconsistencies across lab types complicating direct comparisons, and restricted access to all data simultaneously, requiring an iterative approach to extract the correct values.

Challenge	Solution
Dataset Construction Challenge in Aggregation and Validation	The absence of a neatly formatted, single-source dataset necessitated aggregation of features from various databases, requiring meticulous filtering, precise data joining, and thorough verification at each step to ensure accuracy.
Selecting Surgery Date from Pathology Report	Multiple dates are often associated with biopsies and lab tests for a single procedure. Accurately identifying the most relevant surgery date involved trial-and-error methods, including logic checks and investigating sources of discrepancies comparing different surgery-relevant dates.
Variances in Types of Lab Results	Informed assumptions were essential for reasonably pairing pre-operative and post-operative kidney function data. We faced several challenges, including inconsistencies between types of lab values before and after surgery, many patients presenting only with upper and lower limit values of lab results, and the need to exclude unusable values in lab results.
Classifying Nephrectomies	In many cases, procedures recorded as partial nephrectomies convert to radical nephrectomies during surgery. Often, the nephrectomy type is unspecified in records. To resolve this, we employed NLP techniques to create a rule-based classification system. This system uses random sampling for iterative enhancement, allowing us to develop a list of key terms from pathology reports that provides classification and validation of recorded procedures.
Race Categorizations	We were challenged with very granular and sometimes overlapping race and ethnicity categorizations, with 764 unique values (e.g. "Chinese", "Asian", "Not Indian", "African American and Hispanic"). To enhance clarity and consistency, we simplified these into 5 broad categories, maintaining a reflection of diverse racial and ethnic identities.

Table 1. Challenges and Solutions

Models and Metrics

Models: To predict postoperative renal function, we employed both Linear Regression and XGBoost models. These models, exploring linear and non-linear relationships in demographic and clinical variables, offered insights into feature contributions.

Evaluation Metrics: We chose MAPE (Mean Absolute Percentage Error) for its scale-independence and interpretability, as it suits the clinical context of renal failure prediction and offers a clear, understandable measure of accuracy.

Results

The baseline table for the creatinine cohort is given as follows:

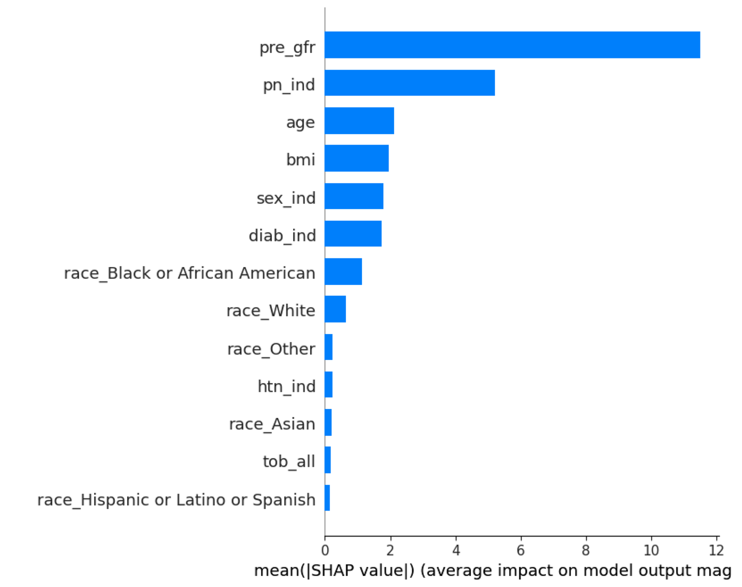
Characteristic	Partial (527)	Radical (237)	Characteristic	Partial (527)	Radical (237)
Age (mean years ± SD)	61.4 ± 12.4	62.7 ± 13.7			
Ethnicity			Gender (Male)	332 (63%)	154 (65%)
American Indian or Native	1 (0.2%)	1 (0.4%)	Gender (Female)	195 (37%)	83 (35%)
Asian	9 (1.7%)	12 (5%)	Diabetes present	143 (25.8%)	78 (27.3%)
Black or African American	67 (12.7%)	43 (18.2%)	Hypertension present	349 (63%)	180 (63%)
White	378 (71.8%)	140 (59.1%)	BMI (mean ± SD)	29.8 ± 6.4	28.4 ± 5.5
Hispanic or Latino or Spanish	17 (3.2%)	6 (2.5%)	Tobacco Use	235 (43.4%)	119 (47.4%)
Other	55 (10.4%)	35 (14.8%)			

Table 2. Baseline Characteristics of Partial and Radical Groups

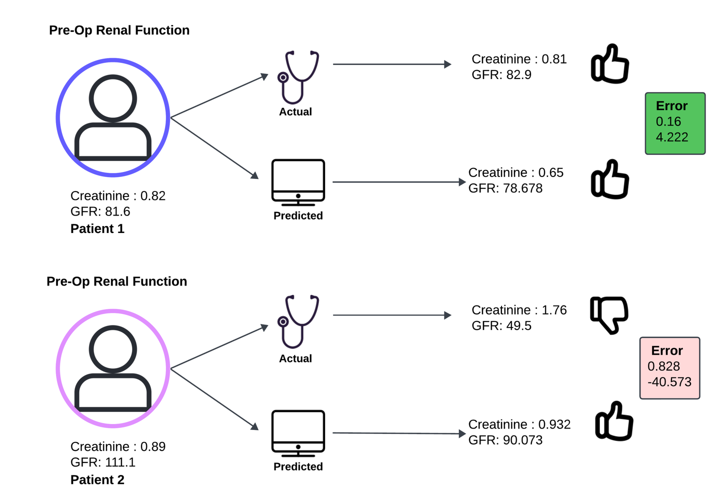
The XGBoost models outperformed the Multiple Linear Regression models, and notably, the Creatinine cohort exhibited superior results compared to the GFR cohort, possibly attributed to the larger sample size within the former. Intriguingly, following Recursive Feature Elimination, the optimal model for the Creatinine cohort identified key features, including 'bmi', 'pre creat', 'pn ind', and 'race White'. In contrast, the GFR model retained all features, excluding certain race indicators. Moreover, the models demonstrated enhanced performance within the Caucasian demographic, potentially reflecting the predominant Caucasian composition of NYU Langone patients.

XGBoost	CREATININE			GFR		
	All Patients	White Patients	POC Patients	All Patients	White Patients	POC Patients
MSE	0.805	0.147	1.073	211.68	178.33	273.49
MAPE	20.55%	18.81%	26.97%	18.81%	18.70%	19.94%

Table 3. Model Performance for the Cr and eGFR Cohorts



(a) SHAP Scores for Creatinine Cohort



(b) Sample Patients: Actual vs Predicted

Limitations

The model lacks differentiation limitations of the approach and models include: 1) exclusion of patients where the surgical intervention could not be classified as PN or RN from the procedure name or available pathology reports; 2) exclusion of patients who did not have pre- an/or postoperative renal function labs performed at NYULH; 3) approximation of procedure dates from available operative notes; 4) lack of differentiation between well-controlled and poorly-controlled hypertension and diabetes; 5) broad categorization ofracial groups; 6) focus on predicting short-term and not long-term renal function.

Future Directions

Future directions that are likely to improve model performance include incorporating pathology reports prior to 2019 to ensure the most accurate surgical intervention classification, integrating tumor size, complexity, and location from imaging and surgical operative notes, indicating whether subjects have a solitary kidney prior to surgery, and refining hypertension and diabetes features to differentiate between well-controlled and poorly-controlled states.

References

[1] Siegel RL et al. Cancer statistics, 2023. CA Cancer J Clin. 2023 Jan;73(1):17-48. doi: 10.3322/caac.21763.

[2] Roussel E et al. Predicting short- and long-term renal function following partial and radical nephrectomy. Urol Oncol. 2023 Feb;41(2):110.e1-110.e6. doi: 10.1016/j.urolonc.2022.10.006. Epub 2022 Nov 10.

[3] Bhindi B et al. Predicting Renal Function Outcomes After Partial and Radical Nephrectomy. Eur Urol. 2019 May;75(5):766-772. doi: 10.1016/j.eururo.2018.11.021. Epub 2018 Nov 23.