Predicting Kidney Stone Disease Through Models and Machine Learning

Ally Atkins*, Advisor: Jue Wang†
*Department of Neuroscience, †Department of Mathematics, Union College

Background

Kidney stone (KS) disease is the most common disease of the urinary tract and is most diagnosed in emergency departments (ED). When symptoms are present, they include acute groin pain, flank (back) pain, bloody urine, urinary tract disease, and infections.
During the past two decades, a significant increase in ED visits with stone-related symptoms have been observed, with over 1.3 million individuals per year presenting to the ED with KS in the US. With the growing prevalence, the cost of diagnosis, treatment, and prevention of KS disease is projected to increase by more than $780 million by 2030.

Introduction

Detecting kidney stones, especially in emergency departments, requires multiple lab tests (complete blood count with comprehensive metabolic panel and urine analysis) and imaging tests. However, These test results are often misinterpreted, leading to an incorrect diagnosis and prognosis. Therefore, diagnostic predictive tools are useful for calculating the likelihood of a patient developing kidney stones., such as STONE PLUS. However, significant efforts can be made to improve a KS diagnostic model, such as incorporating laboratory data and additional clinical characteristics. Creating an improved KS predictive tool would lead to earlier diagnosis and decrease the number of unnecessary imaging, thereby reducing the costs. We investigate which predictive model performs best using machine learning and statistical analysis.

Methodology

We extract information from a large database of patients. This clinical database, known as MIMIC-III Data, contains multiple tables of data relating to patients who stayed within the intensive care units at Beth Israel Deaconess Medical Center. We then analyze certain domains from MIMIC tables, such as demographics, vital signs, laboratory tests, eGFR, etc., and use various machine learning and statistical analysis to determine which domain yields the highest accuracy, sensitivity, and specificity.

Confusion Matrix

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)
Sensitivity: $\frac{TP}{TP+FN}$	Specificity: $\frac{TN}{TN+FP}$

Above is a confusion matrix that considers all TP, FP, FN, and TN when diagnosing a patient with kidney stones. From there, accuracy, sensitivity, and specificity can be calculated using the above information. Sensitivity, also known as true positive rate, measures the proportion of true positives correctly identified. Specificity, known as true negative rate, measures the proportion of true negatives that are correctly identified. Lastly, accuracy is the proportion of correct predictions among the total number of cases.

$$Acurracy = \frac{TP+FN}{TP+FN+FP+TN}$$

Estimated Glomerular Filtration Rate-eGFR

eGFR is the best way to measure a patient's level of kidney function and determines the proper state of kidney disease. A low eGFR indicates lower kidney function.

$$eGFR = 141\times\min(\frac{S_{cr}}{\kappa},1)^{a}\times max(\frac{S_{cr}}{\kappa},1)^{-1,209}\times 0.933^{Age}\times 1.018[if\ female]\times 1.159\ [if\ black]$$

where S_{cr} is serum creatinine in mg/dL, is 0.7 mg/dL for females and 0.9 mg/dL for males, α (serum creatinine is -0.329 for females and -0.411 for males, min indicates the minimum of Scr/κ or 1, and max indicates the maximum of Scr/κ or 1.

Anticipated Results

Logistic Regression

Logistic regression models are useful for determining the best cutoff value for predicting whether a new observation, or patient, is a false positive or false negative. Generally, logistic regression is used for classification and determination of a threshold that most accurately predicts someone who has kidney stones and someone without kidney stones, for example. The ROC curves in the figure below are specific to logistic regression models.

ROC and AUC

A receiver operating characteristic (ROC) curve graph shows the performance of classification models and plots two parameters: sensitivity (true positive rate) and specificity (false positive rate). For example, when analyzing demographic data across kidney stone formers (KS) and acute localized pain (ALP), sensitivity and specificity were 0.71 and 0.47, respectively. The classification model with eGFR as a domain had a sensitivity of 0.59 and specificity of 0.44.
The area under the curve (AUC) can further be calculated to tell you a model's ability to discriminate between cases (positive cases) and non-cases (negative cases).
Overall, based on the ROC curve comparisons in the above figure, the model that includes all domains (demographic data, lab tests, top ICD-9 diagnoses, eGFR, etc.) demonstrated the highest performance with a calculated AUROC of 80%, meaning that the model will correctly identify 80% of positive cases to negative cases to assign a higher risk value.

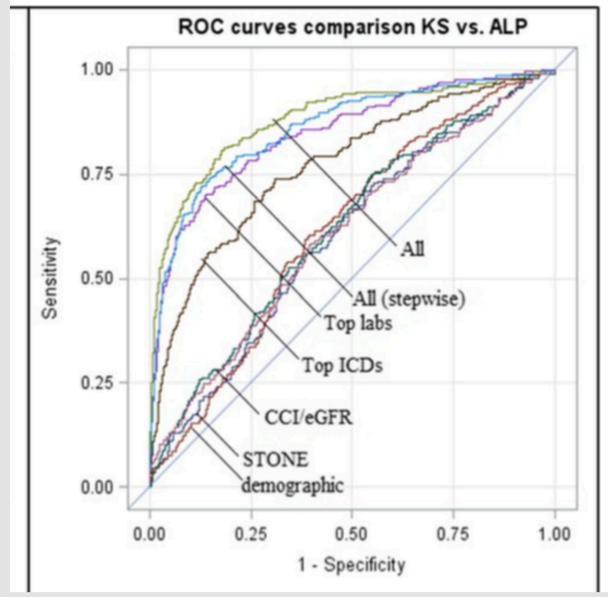

Figure 1. Model comparison via AUROC. Legend: kidney stone (KS) formers vs. acute localized pain (ALP). Logistic regression models upon stepwise feature selection, fit on selected input domains (Chen et al. 2018).

Future Work

- Continue to sort through the clinical database and create our own ROC curves
- Evaluate previous algorithms that improve US accuracy
- Make artificial kidney phantoms and take ultrasound images

Acknowledgements

Union College Undergraduate Research Fellowship

References

1. Chen, Z., Bird, V.Y., Ruchi, R. *et al.* Development of a personalized diagnostic model for kidney stone disease tailored to acute care by integrating large clinical, demographics and laboratory data: the diagnostic acute care algorithm - kidney stones (DACA-KS).*BMC Med Inform Decis Mak* 18, 72 (2018). https://doi.org/10.1186/s12911-018-0652-4
2. MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available at: http://www.nature.com/articles/sdata201635