JOURNAL OF ENDOUROLOGY
Volume 36, Number 2, February 2022

Pp. 243–250
DOI: 10.1089/end.2021.0211

Machine Learning Prediction of Kidney Stone Composition Using Electronic Health Record-Derived Features

Abin Abraham, BS,[1,*] Nicholas L. Kavoussi, MD,[2,*] Wilson Sui, MD,[2] Cosmin Bejan, PhD,[3] John A. Capra, PhD,[1,4,5,†] and Ryan Hsi, MD[2,†,i]

Abstract

Objectives: To assess the accuracy of machine learning models in predicting kidney stone composition using variables extracted from the electronic health record (EHR).

Materials and Methods: We identified kidney stone patients ($n = 1296$) with both stone composition and 24-hour (24H) urine testing. We trained machine learning models (XGBoost [XG] and logistic regression [LR]) to predict stone composition using 24H urine data and EHR-derived demographic and comorbidity data. Models predicted either binary (calcium *vs* noncalcium stone) or multiclass (calcium oxalate, uric acid, hydroxyapatite, or other) stone types. We evaluated performance using area under the receiver operating curve (ROC-AUC) and accuracy and identified predictors for each task.

Results: For discriminating binary stone composition, XG outperformed LR with higher accuracy (91% *vs* 71%) with ROC-AUC of 0.80 for both models. Top predictors used by these models were supersaturations of uric acid and calcium phosphate, and urinary ammonium. For multiclass classification, LR outperformed XG with higher accuracy (0.64 *vs* 0.56) and ROC-AUC (0.79 *vs* 0.59), and urine pH had the highest predictive utility. Overall, 24H urine analyte data contributed more to the models' predictions of stone composition than EHR-derived variables.

Conclusion: Machine learning models can predict calcium stone composition. LR outperforms XG in multiclass stone classification. Demographic and comorbidity data are predictive of stone composition; however, including 24H urine data improves performance. Further optimization of performance could lead to earlier directed medical therapy for kidney stone patients.

Keywords: kidney stone, 24H urine, machine learning

Introduction

Kidney stone composition reflects specific physiologic conditions leading to urinary calculus formation.[1,2] Stone composition is obtained from surgical extraction or captured spontaneous stone passage. Clinical guidelines recommend performing stone analysis when possible and repeat testing when additional samples can be obtained over time to guide treatment.[3] In practice, however, stone composition is known in the minority of kidney stone patients.[4] Accurate, noninvasive methods for prediction of stone composition by using demographic, clinical, and 24-hour (24H) urine data would enable targeted preventative treatment without a need for definitive stone analysis.

Previously, logistic regression (LR) has been used to predict 24H urine parameters from demographic and medical history in calcium formers.[5] However, the performance of the LR method had limited accuracy (overall 64%). Unlike LR methods, machine learning methods, such as boosted decision trees, may provide better accuracy by identifying nonlinear relationships among predictive features that better discriminate stone types. Machine learning algorithms build mathematical models for classifying new data from labeled

[1]Department of Biological Sciences, Vanderbilt Genetics Institute, and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, USA.
Departments of [2]Urology and [3]Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
[4]Bakar Computational Health Sciences Institute, University of California, San Francisco, California, USA.
[5]Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA.
*Co-First Authors.
†Cosenior.
[i]ORCID ID (https://orcid.org/0000-0003-1652-694X).

training examples by extracting predictors from raw data. Prior applications of machine learning have enabled accurate prediction of stone type from processed CT images or by performing a direct computer vision analysis of stones.[6,7] However, computational methods have yet to be developed for the prediction of stone composition from clinical parameters.

We sought to develop machine learning models for predicting stone composition based on a set of demographics, clinical, and laboratory parameters among a large cohort of kidney stone patients. We additionally sought to assess which clinical and demographic predictors were the strongest predictors of stone composition in the models. We trained both boosted decision tree (XGBoost [XG]) and LR machine learning models with a set of electronic health record (EHR)-derived clinical data and evaluated their ability to predict kidney stone composition. Our machine learning models predicted kidney stone composition and identified unique features leading to prediction from EHR-derived data.

Materials and Methods

Patient cohort

After local institutional review board approval, we performed a retrospective review of all adult patients with kidney stone disease who completed 24H urine studies at our institution between 2009 and 2019 and with kidney stone composition available ($n = 1296$). We extracted demographic and clinical information from our cohort using the Vanderbilt University Medical Center Research Derivative, an institutionally maintained database of the EHRs.[8] A specialized laboratory was used for all urine testing (Litholink Corporation, Chicago, IL, USA) and stone composition analysis (Beck Laboratories, Greenwood, IN, USA). Only one stone was analyzed per patient. If there were multiple stone analyses, we identified the temporally closest stone analysis and 24H urine test for the patient. We confirmed that all 24H urine collections were adequate based on gender-specific creatinine per kilogram (Cr24/kg) measurements. Supersaturations were calculated with EQUIL2, representing the relative supersaturation ratio.[9]

Stone composition

Our primary objective was to develop machine learning models to predict stone composition using demographic, clinical, and urine analyte data. Extracted stone compositions included ammonium hydrogen urate, calcium carbonate phosphate, calcium oxalate (monohydrate or dihydrate), calcium phosphate (brushite and hydroxyapatite), magnesium ammonium phosphate (struvite), cystine, or uric acid. If stone composition was mixed, the stones were categorized by highest percentage composition, and then classified as either calcium (i.e., calcium oxalate, hydroxyapatite, or brushite) or noncalcium stones (i.e., ammonium hydrogen urate, cystine, uric acid, and magnesium ammonium phosphate). Second, we subclassified stones as one of calcium oxalate (includes monohydrate or dihydrate), hydroxyapatite, uric acid, or other.

Predictors for machine learning models

We extracted demographic and clinical characteristics from the EHR using a semiautomated data extraction tool.[10,11] Demographic information included age at urine testing, gender, body mass index (BMI), and race. Clinical characteristics were extracted based on International Classification of Disease (ICD) coding (Supplementary Appendix Table 1). We also assessed whether patients had been prescribed an alkalinizing agent, allopurinol or a thiazide diuretic (Supplementary Appendix Table 2). Then, we extracted 24H urine laboratory values, including volume (Vol24), calcium (Ca24), oxalate (Ox24), citrate (Cit24), UA (UA24), sodium (Na24), potassium (K24), magnesium (Mg24), phosphorous (P24), chloride (Cl24), sulfate (S24), urea nitrogen (UUN24), ammonium (NH_4), creatinine (Cr24), creatinine per kilogram (Cr24/kg), calcium per kilogram (Ca24/kg), urine pH, as well as supersaturation index of calcium oxalate (SSCaOx), calcium phosphate (SSCaP), and uric acid (SSUA).

Models for predicting stone composition. We evaluated whether a gradient-boosted decision tree (XGBoost version 0.81, XG) could predict urinary stone composition. XG is a type of machine learning algorithm that combines many decision trees for classification.[12] The algorithm sequentially builds decision trees by penalizing incorrect predictions from the previous decision tree. Decision trees are well suited for EHR data and urine analytes as they are robust for the nonlinear correlation of predictors.[13]

XG models were trained using the predictors already described. Race and gender were categorically encoded, whereas clinical characteristics and medications were encoded as binary variables. All other predictors were treated as

Table 1. Cohort Demographics

Demographics	N *(%)*
Age (years ± SD)	51 ± 15
Gender, male	685 (53)
Gender, female	611 (47)
BMI (mean ± SD)	30 ± 8
Race	
Caucasian	1180 (91)
African American	56 (4)
Asian	20 (2)
Other	40 (3)
Time between stone collection and 24H urine collection, days (mean ± SD)	148 ± 365
Medical history	
Bowel disease	119 (9)
Hypertension	707 (54)
Gout	57 (4)
Diabetes	292 (22)
Cystinuria	3 (0.2)
Coronary artery disease	36 (3)
Cerebrovascular accident	36 (3)
Gastroesophageal reflux disease	469 (36)
Osteoporosis, immobility, or hyperparathyroidism	72 (5)
Medications	
Alkalinizing agent	112 (8)
Thiazide	89 (6)
Allopurinol	52 (4)

24H = 24-hour; BMI = body mass index; SD, standard deviation.

Table 2. Urinary Abnormalities and Stone Composition

24H urine abnormality	N *(%)*	*Normal range*
Low urine volume	747 (57)	>2 L
Hypercalciuria	543 (41)	<250 mg/day for male <200 mg/day for female
Hypocitraturia	588 (45)	>450 mg/day for male >550 mg/day female
Acidic urine	566 (43)	pH <5.8
Alkaline urine	407 (31)	pH >6.2
Hyperuricosuria	299 (23)	>1 g/day
High urine sodium	811 (61)	<150 mmol/day
Hyperoxaluria	474 (37)	<40 mg/day
Majority stone composition		
Calcium oxalate	880 (68)	
Hydroxyapatite	241 (18)	
Carbonate apatite	20(2)	
Uric acid	100 (8)	
Other	55 (4)	

continuous variables. Missing values (69 BMI values and 58 Cr24/kg values out of 1296 subjects) were imputed by considering each predictor as a function of other predictors.[14] Next, we randomly split the data with the equal proportion of stone composition types into a training (80%) and a validation cohort (20%).

Each stone type was weighted based on prevalence to account for the imbalanced data set. Using the training data, XG hyperparameters were optimized using Bayesian techniques. Standardized hyperparameters were used.[15] Model performance was evaluated on the validation cohort using the area under the receiver operating curve (AUC-ROC), sensitivity, and accuracy metrics. LR models were trained and evaluated using the same cohort as used for the XG model. We did not perform any hyperparameter tuning. LR model performance was also evaluated using the validation cohort.

Model interpretation

To determine predictors driving the models, we use Shapley Additive Explanation (SHAP) as implemented in the SHAP v0.35.0.[16] SHAP scores represent the marginal contribution of each predictor over all permutations of predictors

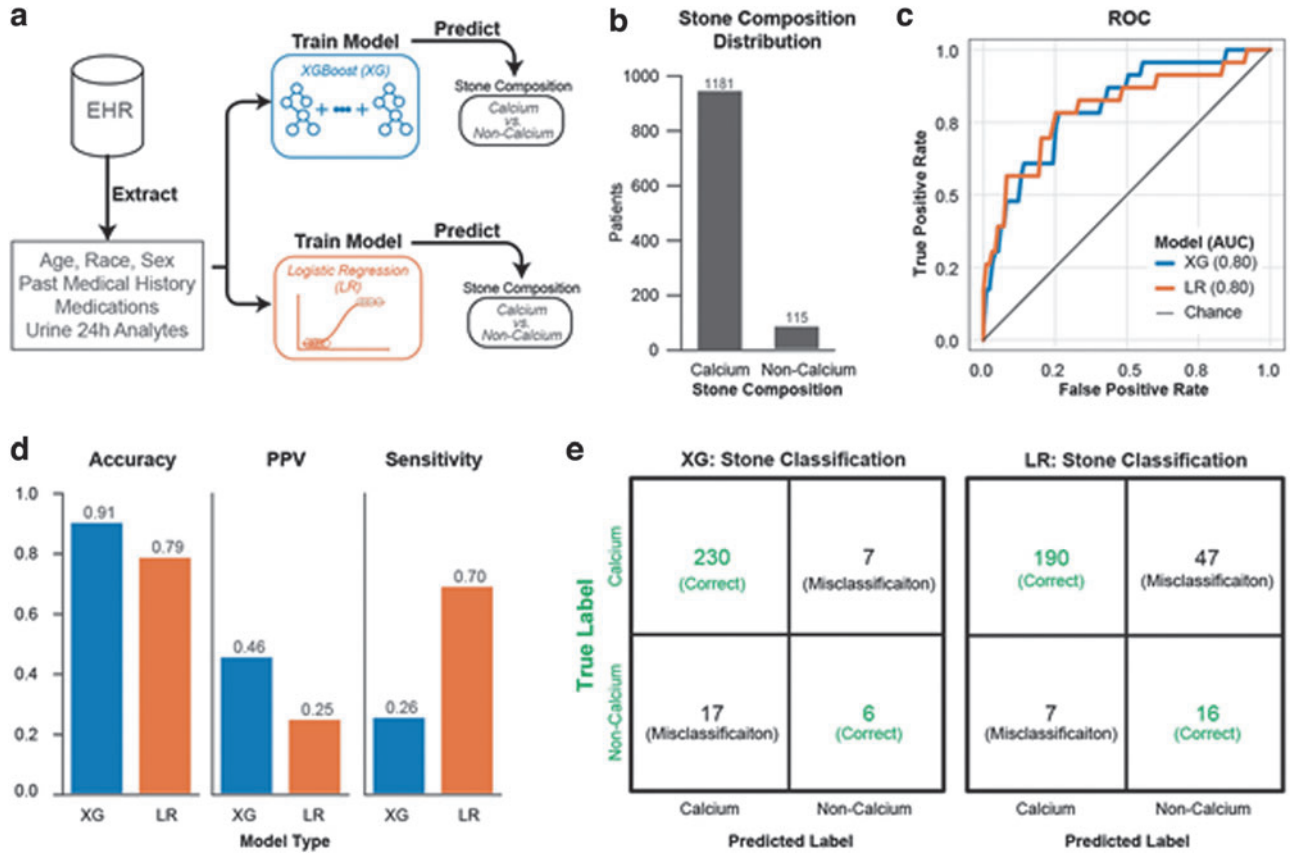

FIG. 1. LR and XG discriminate between calcium and noncalcium urinary stone composition. **(a)** Schematic workflow for training models: EHR predictors were extracted and used to train a boosted decision tree (XG, *blue*) or LR (*orange*) for the prediction of stone composition as either calcium or noncalcium. **(b)** Calcium stone composition was identified in the majority (91%) out of 1296 patients. **(c)** Receiver operator characteristic curves for XG and LR models demonstrate an AUC for both models as 0.80. Models were evaluated using the held-out test cohort of 260 patients. **(d)** Accuracy, PPV, and sensitivity for noncalcium stone composition show higher PPV for XG and higher sensitivity for LR. **(e)** Confusion matrices with number of stones correctly and incorrectly classified for XG and LR models in the validation cohort. AUC = area under the curve; EHR = electronic health record; LR = logistic regression; PPV = positive predictive value; XG = XGBoost.

used for classification. We reported the SHAP score in units of change in prediction log odds and refer to them as ''predictor importance.''

Evaluation metrics

Primary outcomes included the AUC-ROC, positive predictive value (PPV), sensitivity, and accuracy metrics for predicting stone composition with the machine learning models. Secondary outcomes included the predictor importance of each variable from the EHR-derived data used by the models for stone composition prediction. Since each urinary predictor had different numerical ranges, all urinary predictors were normalized between 0 and 1 (i.e., lowest and highest), respectively.

Results

Demographic descriptions of patients

Table 1 displays demographic information on the 1296 patients included for analysis. The patients were predominantly Caucasian (91%) with the most common comorbidities being hypertension (54%), gastroesophageal reflux disease (36%), and diabetes (22%). Only a small number of patients were on medical therapy with an alkalization agent (8%), thiazide diuretic (6%), or allopurinol (4%). Calcium stones comprised most cases (91%). The majority stone compositions included calcium oxalate (68%), hydroxyapatite (18%), uric acid (8%), and other (6%). Elevated urinary sodium (61%) and low urine volume (57%) were the most common abnormalities identified on 24H urine analysis (Table 2).

Binary classification

Using predictors extracted from EHRs, we trained both XG and LR models to classify stone composition as either calcium or noncalcium (Fig. 1a). Both XG and LR models discriminated between calcium and noncalcium stones (AUC = 0.80 for both; Fig. 1c). XG had a higher accuracy (91% *vs* 79%) for predicting noncalcium stone composition than LR as well as PPV (0.46 *vs* 0.26; Fig. 1d, e). However, LR had a higher sensitivity in identifying noncalcium stone composition than XG (0.70 *vs* 0.26; Fig. 1d, e).

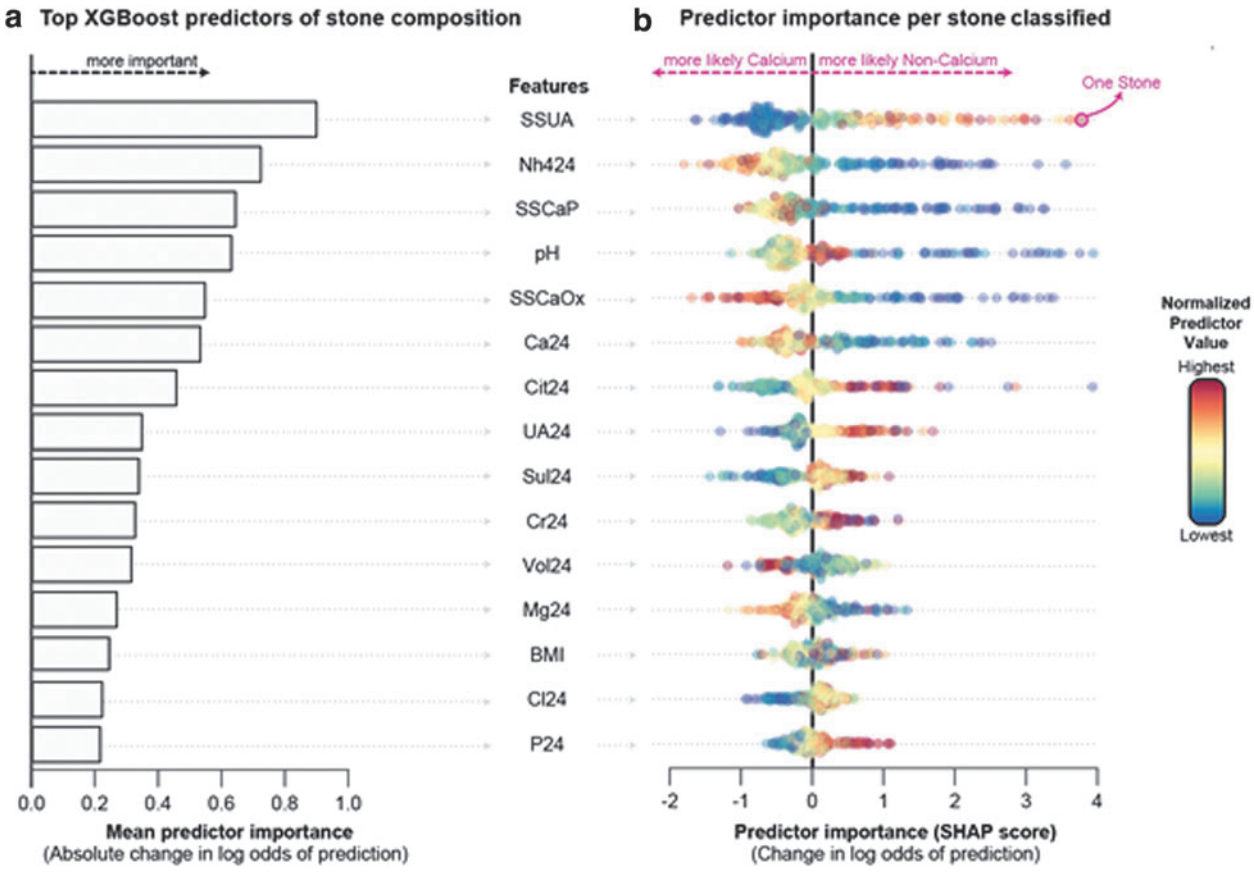

FIG. 2. Top predictors of stone composition (calcium *vs* noncalcium) for XG. **(a)** Predictor importance scores were used to quantify the contribution of each predictor in the data set to the prediction of stone composition for each patient. The top 15 predictors are shown. Predictor importance scores indicate the relative contribution of each predictor as the change in log odds of prediction. **(b)** The predictor importance scores per patient (*one dot*) for the top predictors (*y*-axis) and their impact on stone composition (*x*-axis). Positive scores indicate higher likelihood of noncalcium composition; negative scores indicate higher likelihood of calcium stone composition. Each predictor per individual is colored based on the normalized predictor value (*red* = high value, *blue* = low value). The top predictors identified by the model are known to biochemically impact kidney stone formation (abbreviations are defined in Materials and Methods section). For example, lower values of SSUA more likely predict calcium stones whereas higher values more likely predict noncalcium stones. SSUA = supersaturation of uric acid.

We identified variables most predictive for binary stone classification by the XG model by quantifying the predictor importance of each variable in the validation cohort. The top three XG predictors of stone composition were SSUA, NH_4, and SSCaP (Fig. 2a). Specifically, higher values of NH_4, SSCaox, SSCaP, Ca24, and Mg24 and lower values of SSUA, pH, citrate, UA, Sul24, Cr24, Cl24, and P24 increased likelihood of a calcium stone composition prediction. In addition, higher values of SSUA, UA24, Sul24, and P24, as well as lower values of NH_4, SSCaP, pH, SSCaox, and Mg24 were more likely to predict noncalcium stone composition (Fig. 2b).

Multiclass classification

Both XG and LR models were trained to predict stone composition as one of calcium oxalate, hydroxyapatite, uric acid, or other (Fig. 3a, b). LR outperformed XG with higher accuracy (0.64 *vs* 0.56) and AUC (0.79 *vs* 0.59; Fig. 3b). Urine pH was the most important global predictor when all four stone categories were combined. Urine pH also had the highest predictive value for calcium oxalate, uric acid, and ''other'' stone compositions. The top predictor for hydroxyapatite was Cr24 (Fig. 3c).

Urine analytes improve stone composition prediction

To quantify the impact of urine 24H analytes on prediction performance, we compared three versions of XG and LR models: a baseline model including age, race, gender, medical history, and medications; a second version adding urine 24H analytes to the baseline predictors; and a third version including only urine 24H analytes (Fig. 4a). Across all three versions, LR had higher ROC-AUC than XG (Fig. 4b). Adding urine analytes improved ROC-AUC in both XG (0.53–0.59) and LR (0.68–0.79; Fig. 4b). Models trained using only urine analytes achieved similar ROC-AUC (LR: 0.79, XG = 0.61) to those trained using both baseline and analyte predictors. Across the four stone types, different sets of predictors were prioritized for the final prediction of stone composition between XG and LR models (Fig. 4c). For example, in predicting calcium oxalate stones, the association of SSUA, Cit24, and SSCaP was strongest for predicting stone type for the XG model, and Na24, Cl24, and SSCaP were strongest for the LR model.

Discussion

There are several important findings in this study. First, machine learning models can predict stone composition

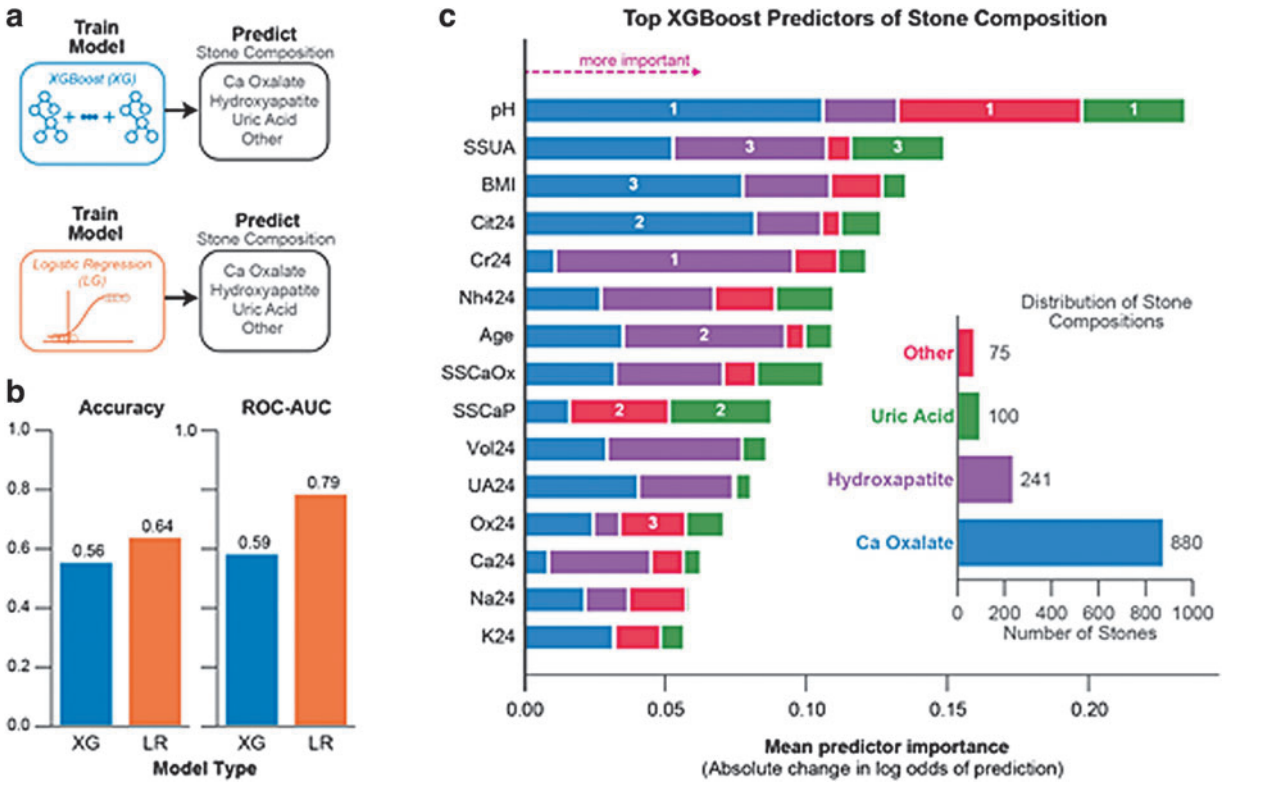

FIG. 3. LR outperforms XG models when predicting multiclass stone compositions. **(a)** XG and LR models were trained to predict one of four stone compositions: calcium oxalate (Ca oxalate), hydroxyapatite, uric acid, or other. **(b)** The LR model (*x*-axis) had higher accuracy and ROC-AUC (*y*-axis) than the XG model. Model performance was evaluated on a held-out test cohort that was not used for training or validation. **(c)** The top 15 predictors based on the highest contribution to the stone composition prediction are shown. Predictor importance was quantified based on the change in the stone composition predicted log-odds as shown in Figure 2. The contribution of each predictor was further stratified by stone composition and the top three predictors for each stone composition are annotated by its rank. The distribution of patients by stone type are shown in the inset figure. Predictor importance reflects the known pathophysiology of stone formation for the individual stone types. ROC-AUC = area under the receiver operating curve.

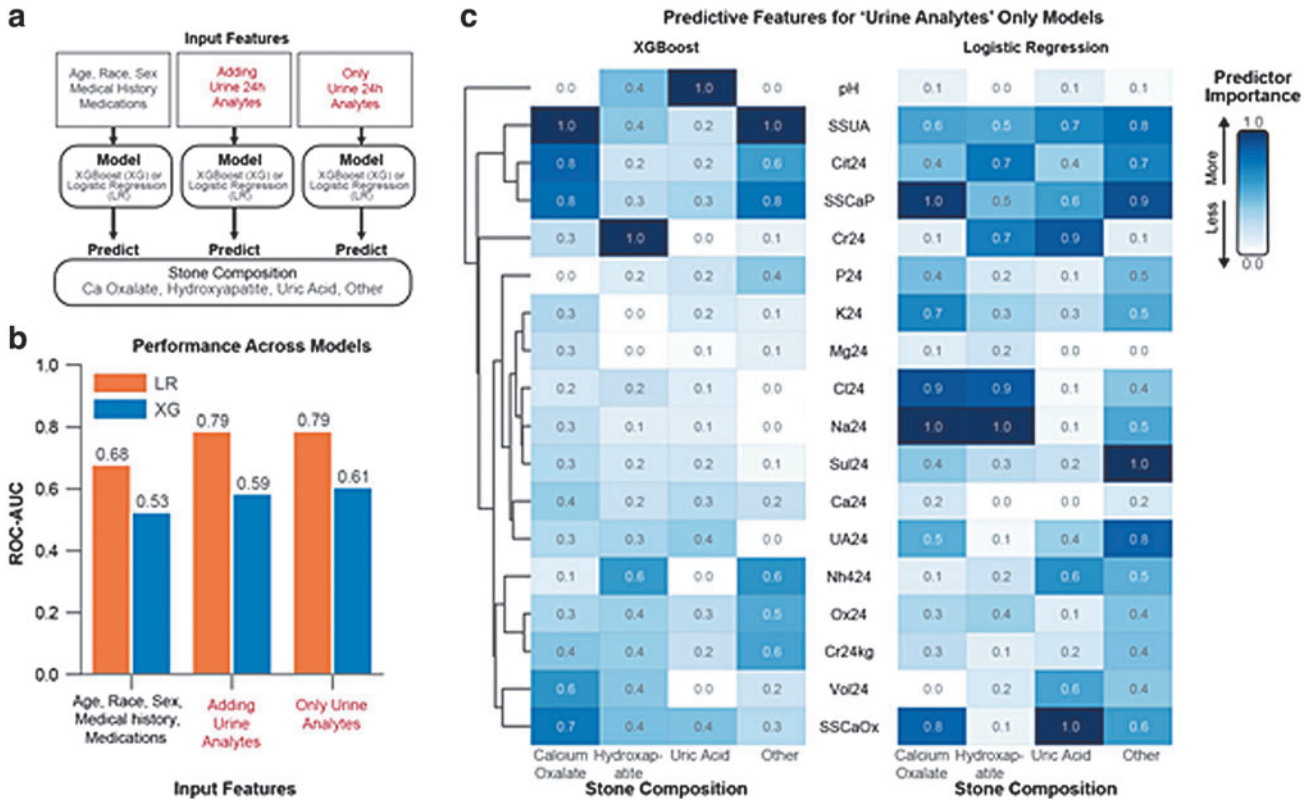

FIG. 4. Urine 24H analytes drive stone composition prediction. **(a)** Three versions of XG and LR models were compared after being trained with different inputs: (1) age, race, gender, medical history, and medications; (2) age, race, gender, medical history, and medications, as well as urine 24H analytes with the baseline predictors; (3) only urine 24H analytes. **(b)** Model performance was evaluated using ROC-AUC averaged across all stone composition categories. LR performed consistently better than XG. Including only urine analytes had the highest performance; including other nonurine analyte predictors did not improve performance. **(c)** Predictor importance score across the validation cohort normalized for each stone composition. The predictor scores allocate ''credit'' or the contribution of each predictor per individual for each predicted stone composition. More specifically, predictor scores represent the marginal contribution of each predictor over all permutations of predictors used for classification. 24H = 24 hour.

based on demographic, clinical, and urine analyte variables. In binary classification of stone prediction, XG had higher accuracy than the LR model. However, in multiclass classification, the LR model outperformed the XG model. XG models can identify more complex nonlinear patterns in large data sets, yet in this instance, weaker performance of the XG model may reflect the relatively small cohort used for training our model in the setting of a heterogeneous disease. Second, important predictors identified by the machine learning models for stone prediction support the current understanding of the pathophysiology of stone formation. Third, we demonstrate that 24H urine variables are most predictive for stone composition, even when used alone beyond demographic and clinical data in different predictive models. Conversely, we show that demographic and comorbidity data can be used to predict stone composition without 24H urine data; however, the performance is improved when 24H urine data are available. Taken together, this information shows feasibility of EHR-derived prediction tools supporting clinical practice and decision making.

Current machine learning techniques combine predictors in many layers using numerous computational transformations. By using regression and boosted decision tree approaches, we identified the most important predictors in the models for distinguishing stone composition as well as each predictor's relative impact. Factors such as urine pH, SSUA, BMI, and Cit24 were most impactful when predicting any category of stone composition, and these are known to contribute to stone pathogenesis.[17–19] For example, acidic urinary pH leads to protonation and precipitation of uric acid and cystine crystals, whereas alkaline pH facilitates crystallization of calcium phosphate stones.[20] In addition, acidic urinary pH has been associated with calcium oxalate formation through heterogeneous nucleation and a decrease in concentration of urinary buffers, such as citrate.[21,22] Some of the analytes used by our models may play a role in stone formation, but their involvement has yet to be clearly elucidated. For example, urine ammonium was the second most important predictor when predicting binary stone classification by our model and reflects individual acid–base status. Abnormalities in urine ammonium are uncommon but associated with Gastro-intestinal malabsorption and urinary tract infections.[23] Future machine learning models may help explicate the relative impact of each predictor on stone recurrence.

Previous studies have evaluated prediction models using clinical and urine analyte data for stone composition.

Moreira et al. used multinomial LR to predict stone type based on 24H urine data.[5] In their smaller cohort ($n = 508$), the 24H urine data alone predicted stone type with an accuracy of 64%. The authors concluded that utilizing urine analyte data in conjunction with other clinical features may improve stone composition prediction. By incorporating demographic and clinical data with urine analyte data when training our machine learning models, we found improved binary prediction of stone classification (accuracy = 91%). However, multiclass prediction accuracy was driven primarily by urine analyte data and comparable with the prior study (accuracy = 64%; Fig. 4b). Kidney stone patients represent a heterogeneous population, and demographic and clinical data are variable between subjects.[24] In our model, the influence of these data was small compared with 24H urine analyte data for multiclass stone classification.

Stone analysis is critical to the metabolic evaluation of stone formers and can help guide treatment in stone formers in part with medical history.[3,25] Although kidney stone composition often directs preventative therapy, stone composition testing in practice occurs in the minority of patients. By training robust machine learning algorithms to accurately predict stone composition, tools may be developed leading to earlier targeted dietary or pharmacologic therapy for stone disease. Although still investigational, the current widespread adoption of machine learning tools in daily life suggests their inevitability in clinical practice. Furthermore, these models could be optimized to identify nonlinear relationships between clinical predictors and stone composition when there are multiple compositions present and elucidate the clinical relevance of mixed stone compositions.

There are several limitations to our study. The retrospective design cannot account for unmeasured confounders and omitted variable bias. EHRs provide an incomplete summary of treatment for some individuals, and other information such as Hounsfield units or stone size from diagnostic imaging could not be confirmed. It is likely some patients were not necessarily first-time stone formers. We matched stone analyses with the temporally closest 24H urine results. However, patients could have previously been given dietary counseling or pharmacologic therapy that could have impacted 24H urine results. In addition, it is possible that collected stones had formed under previous physiologic conditions, which were no longer present in the patient at the time of the 24H urine analysis. Furthermore, all stone composition analyses were performed at a single laboratory, which may not account for variations in stone composition found across different commercial stone analysis laboratories.[26] Only one laboratory was used for urine analysis and interlaboratory variability of 24H urine analysis could limit model interpretation. Despite these limitations, our study demonstrates the potential of machine learning models to noninvasively predict kidney stone composition. Future optimization of our models in larger data sets and external validation will help achieve the clinical promise of this approach.

Conclusion

We have developed machine learning models integrating demographic and EHR data to predict kidney stone composition. The performance for predicting calcium *vs* noncalcium stones is better than for predicting specific stone subtypes including uric acid and hydroxyapatite stones. The predictors prioritized by the machine learning models for stone prediction support the current understanding of the pathophysiology of stone formation. We show that demographic and comorbidity data can be used to predict stone composition without 24H urine data; however, the performance is best when including 24H urine data. Further studies to optimize and validate the models could contribute to the creation of clinical tools used as a surrogate for stone analysis.

Disclaimer

The project contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

The project described was supported by CTSA award No. UL1TR000445 from the National Center for Advancing Translational Sciences. A.A. was supported by the American Heart Association fellowship 20PRE35080073, National Institutes of Health (NIH, T32GM007347), and the Burroughs Wellcome Fund. J.A.C. was supported by the National Institutes of Health (NIH, R35GM127087) and the Burroughs Wellcome Fund.

Supplementary Material

Supplementary Appendix Table 1
Supplementary Appendix Table 2

References

1. Zisman AL, Coe FL, Cohen AJ, et al. Racial differences in risk factors for kidney stone formation. Clin J Am Soc Nephrol 2020;15:1166–1173.
2. Youssef RF, Martin JW, Sakhaee K, et al. Rising occurrence of hypocitraturia and hyperoxaluria associated with increasing prevalence of stone disease in calcium kidney stone formers. Scand J Urol 2020;54:1–5.
3. Pearle MS, Goldfarb DS, Assimos DG, et al. Medical management of kidney stones: AUA Guideline. J Urol 2014;192:316–324.
4. Singh P, Enders FT, Vaughan LE, et al. Stone composition among first-time symptomatic kidney stone formers in the community. Mayo Clin Proc 2015;90:1356–1365.
5. Moreira DM, Friedlander JI, Hartman C, et al. Using 24-hour urinalysis to predict stone type. J Urol 2013;190: 2106–2111.
6. Black KM, Law H, Aldoukhi A, et al. Deep learning computer vision algorithm for detecting kidney stone composition. BJU Int 2020;125:920–924.
7. Hokamp NG, Lennartz S, Salem J, et al. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: An ex-vivo study. Eur Radiol 2020;30:1397–1404.
8. Danciu I, Cowan JD, Basford M, et al. Secondary use of clinical data: The Vanderbilt approach. J Biomed Inform 2014;52:28–35.

9. Werness PG, Brown CM, Smith LH, et al. Equil2: A basic computer program for the calculation of urinary saturation. J Urol 1985;134:1242–1244.
10. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208.
11. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2008;42:377–381.
12. Krishnapuram B, Shah M, Smola A, et al. XGBoost: A Scalable Tree Boosting System. Arxiv 2016: 785–794.
13. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Springer-Verlag, New York, 2008, pp. 605–624.
14. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-Learn: Machine Learning in Python. Arxiv 2012. Arxiv preprint Arxiv:1201.0490.
15. Bergstra J, Yamins D, Cox DD. Making a Science of Model Search. Arxiv 2012. Arxiv preprint Arxiv:1209.5111.
16. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2:749–760.
17. Taylor EN, Stampfer MJ, Curhan GC. Obesity, weight gain, and the risk of kidney stones. JAMA 2005;293:455–462.
18. Curhan GC, Willett WC, Speizer FE, et al. Twenty-four–hour urine chemistries and the risk of kidney stones among women and men. Kidney Int 2001;59:2290–2298.
19. Whitson JM, Cooperberg MR, Stackhouse GB, et al. Urinary citrate levels do not correlate with urinary pH in patients with urinary stone formation. Urology 2007;70: 634–637.
20. Wagner CA, Mohebbi N. Urinary pH and stone formation. J Nephrol 2010;23 Suppl 16:S165–S169.
21. Berg C, Tiselius H-G. The effect of pH on the risk of calcium oxalate crystallization in urine. Eur Urol 1986;12: 59–61.
22. Grases F, Costa-Bauzá A, Gomila I, et al. Urinary pH and renal lithiasis. Urol Res 2012;40:41–46.
23. Sui W, Hancock J, Asplin JR, et al. Nephrolithiasis and elevated urinary ammonium: A matched comparative study. Urology 2020;144:77–82.
24. Pak CYC, Poindexter JR, Adams-Huet B, et al. Predictive value of kidney stone composition in the detection of metabolic abnormalities. Am J Med 2003;115:26–32.
25. Kourambas J, Aslan P, Teh CL, et al. Role of stone analysis in metabolic evaluation and medical treatment of nephrolithiasis. J Endourol 2001;15:181–186.
26. Krambeck AE, Khan NF, Jackson ME, et al. Inaccurate reporting of mineral composition by commercial stone analysis laboratories: Implications for infection and metabolic stones. J Urol 2010;184:1543–1549.

Address correspondence to:
Nicholas L. Kavoussi, MD
Department of Urology
Vanderbilt University Medical Center
1211 Medical Center Drive
Nashville, TN 37212
USA

E-mail: nicholas.l.kavoussi@vumc.org

Abbreviations Used

24H = 24-hour
BMI = body mass index
CT = computed tomography
EHR = electronic health record
LR = logistic regression
PPV = positive predictive value
ROC-AUC = area under the receiver operating curve
SHAP = Shapley Additive Explanation
SSCaOx = supersaturation of calcium oxalate
SSCaP = supersaturation of calcium phosphate
SSUA = supersaturation of uric acid
XG = XGBoost