A Machine Learning-Based Laboratory Screening Tool for Empowering Patient Education and Preventive Counseling in Kidney Stone Disease

Abstract

Aims: Preventive counseling for kidney stone disease requires accurate risk identification, yet traditional interpretation fails to capture complex metabolic interactions. This study aimed to develop a machine learning-based screening tool using routine blood and urine data to identify high-risk individuals and provide actionable insights for patient education.

Instrument & Methods: Data from 952 individuals (714 stone formers, 238 controls) with 28 features were analyzed. Particle Swarm Optimization selected features, balancing accuracy and test reduction. An XGBoost model was developed with 5-fold cross-validation. SHAP analysis identified influential predictors.

Findings: The PSO-optimized model reduced tests from 28 to 8 features (71% reduction) with 93.49% accuracy (95% CI: 91.8-95.0%). The final panel included: Age, Blood_Phosphorus, Blood_Uric_Acid, Blood_Creatinine, eGFR, Urine_pH, Urine_Calcium_24h, and Urine_Citrate_24h. SHAP identified modifiable risk factors: low citrate (potassium citrate), elevated calcium (dietary modification), elevated uric acid (purine reduction), and reduced eGFR (renal evaluation).

Conclusion: This tool enables non-invasive risk identification from routine data, empowering targeted preventive counseling in primary care.

Keywords: Nephrolithiasis; Machine Learning; Patient Education as Topic; Preventive Health Services; Particle Swarm Optimization; Clinical Laboratory Techniques

Introduction

Kidney stone disease affects 10-12% of adults globally, with lifetime prevalence of 19.1% in men and 9.4% in women by age 70 [1]. Recurrence rates approach 50% within five to ten years [2]. Contemporary understanding recognizes stones as manifestations of chronic systemic conditions, necessitating comprehensive risk assessment [3].

Traditional diagnosis relies on imaging. Non-contrast CT has 93% sensitivity and 96% specificity [4], but exposes patients to radiation and is inaccessible in resource-limited settings. Ultrasonography has limitations in stone detection and tends to overestimate size [5]. Imaging identifies stones only after symptoms appear, rather than enabling early risk identification.

Patient education and preventive counseling are essential components of kidney stone management. Dietary modifications, increased fluid intake, and targeted pharmacotherapy can significantly reduce stone recurrence [6]. Effective counseling requires accurate identification of modifiable risk factors, which traditionally requires 24-hour urine collections and comprehensive metabolic panels. These tests are costly, time-consuming, and often inaccessible in primary care settings, creating a barrier to preventive care[7]. Studies have shown that health education significantly increases patient adherence to treatment and empowers patients to take an active role in their disease management[8].

Laboratory-based risk assessment using blood and urine offers a cost-effective alternative. However, conventional statistics fail to capture non-linear interactions among lithogenic risk factors [6]. Data mining techniques have been extensively reviewed and applied across various domains, with Liao et al. providing a comprehensive decade-long review of their applications[9]. Machine learning can model these complex relationships and has shown promise in disease classification [10, 11].

The growing acceptance of machine learning in laboratory medicine is evidenced by several recent studies. Al Dhuhli et al.[12] evaluated multiple machine learning models including Artificial Neural Networks, XGBoost, and Random Forest for estimating serum low-density lipoprotein cholesterol from routine lipid profiles, demonstrating superior performance compared to conventional formulas. Wang et al.[13] developed a machine learning model based on routine complete blood count parameters to differentiate hematologic disorders, achieving one-vs-rest AUCs of 0.920-0.970. Kavoussi et al.[7] achieved 98% accuracy using machine learning for 24-hour urinary abnormality prediction in kidney stone disease.

In the context of kidney stone management specifically, Several studies have applied ML to kidney stone challenges. Mahmoodi et al. [14] used FACS cohort data to predict symptomatic stones.

Shabaniyan et al.[15] developed an AI-based DSS for PCNL outcomes (94.8% accuracy). Rashidi et al.[16] used AdaBoost (89% accuracy) to predict ureteral stone treatment outcomes. Mortazavi et al.[17] classified stone causes using Random Forest, SVM, and neural networks.

Evolutionary algorithms have been successfully applied for feature selection in health-related datasets. Comprehensive reviews have systematically compared swarm intelligence-based feature selection methods for medical applications[18, 19]. Additionally, evolutionary computation approaches have been extensively surveyed for feature selection in high-dimensional datasets[20]. PSO was introduced by Kennedy and Eberhart[21] with comprehensive surveys documenting its developments[22, 23].

Feature selection methods play a crucial role in improving the accuracy of predictive models. Mohtasham et al.[24] compared various filter, embedded, and hybrid feature selection techniques, demonstrating that hybrid approaches combined with ensemble learning can effectively identify the most influential clinical predictors Enayati et al.[25] used PSO with KNN to identify health literacy factors (92.02% accuracy). Nopour et al.[26] developed rule-based machine learning models for breast cancer diagnosis, achieving 99.1% accuracy with J-48 decision tree algorithm, highlighting the potential of interpretable ML models in clinical settings. Ayaad et al.[27] demonstrated that machine learning models, particularly Gradient Boosting, can effectively predict patient learning needs in oncology settings, with an AUC of 0.96, establishing the feasibility of using ML for identifying educational priorities. Feature selection methods in medical data analysis face unique challenges, including high dimensionality and class imbalance, which have been extensively reviewed[28, 29].

Despite the demonstrated potential of machine learning in laboratory medicine and patient education, no study has systematically developed a screening tool specifically designed to translate laboratory findings into actionable patient education and counseling recommendations for kidney stone disease. Furthermore, the absence of explainable AI limits clinical adoption, as healthcare providers need to understand why a patient received a high-risk prediction to deliver effective counseling.

The present study aimed to develop a machine learning-based screening tool using routine blood and urine data that identified high-risk individuals for kidney stone disease, reduced the number of required laboratory tests to a practical 8-feature panel, and provided interpretable, actionable insights for patient education and preventive counseling. Specifically, the study evaluated the effectiveness of Particle Swarm Optimization (PSO) in reducing the number of required laboratory tests for kidney stone prediction while maintaining clinically acceptable accuracy, identified the most influential predictors of kidney stone risk among routine blood and urine parameters and their interactions, and translated the model predictions into actionable patient education and preventive counseling messages for kidney stone disease.

Instrument and Methods

In this section, we describe the study design, population, data collection and preprocessing procedures, feature selection method, machine learning model development, and statistical analysis used to develop and validate the screening tool for kidney stone risk prediction.

Study Design and Population

This retrospective study included data from patients who underwent routine laboratory evaluation for kidney stone disease at multiple referral urology centers between January 2021 and December 2024. A total of 952 individuals were included, comprising 714 patients with confirmed kidney stones (stone formers) and 238 healthy controls with no history of nephrolithiasis confirmed by imaging. Kidney stone diagnosis was established by non-contrast CT or ultrasonography.

Sample size calculation: Based on population statistics and using Cochran's formula for large populations[30], the minimum required sample size was calculated as:

$$n = \frac{Z^2 p(1-p)}{e^2} = 384.16 \quad (1)$$

To account for potential missing data and enhance statistical power, we included 952 individuals, exceeding the minimum requirement.

Exclusion criteria included: (1) age < 18 years; (2) pregnancy; (3) active urinary tract infection; (4) missing key laboratory data (>20% missing); and (5) history of renal transplantation or dialysis.

Data collection and preprocessing

Demographic, clinical, and laboratory data were extracted from electronic health records. The dataset comprised 28 numerical features organized into demographic variables, blood laboratory parameters, and 24-hour urine laboratory parameters. Missing values were imputed using median values from the training set only to prevent data leakage. Outlier detection was performed using Winsorization at the 1st and 99th percentiles, and all numerical features were standardized using Z-score normalization. Both Winsorization and standardization were applied within each cross-validation fold using parameters derived from the training set and applied to the test set to prevent data leakage.

Feature Selection with PSO

Feature selection used PSO, following the wrapper-based approach [25, 31]. Each solution was a binary vector of length 28. The fitness function balanced AUC maximization and feature count minimization. PSO used 20 particles, 30 iterations, inertia weight 0.7→0.4, and cognitive/social coefficients 2.0.

$$Fitness = \alpha \cdot (1 - AUC) + \beta \cdot \frac{n_{selected}}{n_{total}} \quad (2)$$

where $\alpha = 0.9$ and $\beta = 0.1$.

The final 8-feature panel selected by PSO was reviewed by a board-certified urologist to ensure clinical relevance, interpretability, and alignment with established lithogenic risk factors. This step was taken to bridge the gap between data-driven optimization and practical clinical applicability.

Machine Learning Model Development

An XGBoost (eXtreme Gradient Boosting) model was developed using the PSO-selected features with 5-fold stratified cross-validation. XGBoost is a scalable tree boosting system introduced by Chen and Guestrin[32], building upon the gradient boosting framework originally proposed by Friedman[33]. Stratification preserved the class distribution (stone formers vs. healthy controls) in each fold. Hyperparameters were optimized using grid search with the following search space: n_estimators ∈ [50, 100, 200], max_depth ∈ [3, 5, 7], and learning_rate ∈ [0.01, 0.05, 0.1].

Model performance was assessed using five metrics: accuracy, precision, recall, F1-score, and AUC-ROC. Recall (sensitivity) was prioritized as the primary metric because false negatives (missed diagnoses) carry greater clinical harm than false positives in a screening context. For key performance metrics, 95% confidence intervals were calculated using bootstrapping with 1,000 iterations.

Model Interpretability with SHAP Analysis

SHAP (SHapley Additive exPlanations) analysis was performed on the final model to identify the most influential predictors and their clinical implications. The SHAP method was originally proposed by Lundberg and Lee[34], with recent comprehensive reviews by Chen et al.[35] providing a unified framework for interpreting model predictions based on Shapley values from cooperative game theory.

Statistical Analysis

All experiments were conducted with fixed random seeds (random_state = 42) to ensure reproducibility. 5-fold stratified cross-validation was used for all model evaluations. For key performance metrics, 95% confidence intervals were calculated using bootstrapping with 1,000 iterations. Method comparison was assessed using Bland-Altman analysis[36]. Software: All analyses were performed using Python 3.11 with scikit-learn (version 1.3.0), XGBoost (version 2.0.0), and SHAP (version 0.44.0).

Findings

In this section, we report the baseline characteristics of the study population, the performance of the PSO-optimized XGBoost model compared to the baseline, the results of SHAP analysis for model interpretability.

Baseline Characteristics

Table ١ presents the baseline demographic and laboratory characteristics of the study population stratified by group. Significant differences were observed between stone formers and healthy controls for several parameters. Stone formers had higher BMI (27.4 ± 3.8 vs. 26.8 ± 3.6 kg/m^2, p = 0.038) and were more likely to be male (55.0% vs. 42.9%, p = 0.002). Laboratory parameters including Blood_Uric_Acid, Blood_Creatinine, Urine_Calcium_24h, and Urine_Citrate_24h differed significantly between groups (all p < 0.05). These differences are consistent with established lithogenic risk factors.

Table ١) Baseline characteristics of the preprocessed dataset

Characteristic	Stone Formers (n=714)	Healthy Controls (n=238)	p-value
Age, mean ± SD	48.2 ± 14.1	46.5 ± 14.8	0.124
Male gender, n (%)	335 (55.0%)	87 (42.9%)	0.002
BMI, mean ± SD	27.4 ± 3.8	26.8 ± 3.6	0.038
Blood_Phosphorus (mmol/L)	1.12 ± 0.18	1.08 ± 0.15	0.002
Blood_Uric_Acid (mmol/L)	0.42 ± 0.11	0.35 ± 0.09	<0.001
Blood_Creatinine (μmol/L)	89.4 ± 21.3	78.6 ± 18.2	<0.001
eGFR (mL/min/1.73m^2)	82.5 ± 15.7	91.3 ± 13.4	<0.001
Urine_pH	5.8 ± 0.5	6.2 ± 0.6	<0.001
Urine_Calcium_24h (mmol/day)	6.8 ± 2.4	4.2 ± 1.8	<0.001
Urine_Citrate_24h (mmol/day)	2.1 ± 0.9	3.5 ± 1.2	<0.001
Urine_Oxalate_24h (mmol/day)	0.38 ± 0.14	0.29 ± 0.11	<0.001
Urine_Volume_24h (L/day)	1.8 ± 0.6	2.1 ± 0.5	<0.001

Data are presented as mean ± standard deviation or number (percentage). P-values were calculated using independent t-test for continuous variables and chi-square test for categorical variables.
Abbreviations: *BMI, body mass index; eGFR, estimated glomerular filtration rate; SD, standard deviation.*

Model Performance

The baseline XGBoost model using all 28 features achieved 97.18% accuracy (AUC 0.9941). The PSO-optimized model reduced features to 8 (71% reduction) while maintaining 93.49% accuracy (95% CI: 91.8-95.0%) with recall of 95.42% and AUC of 0.9816.

Table ٢) XGBoost performance comparison across feature subsets

Feature Subset	Features	Accuracy	Recall	Precision	F1-Score	AUC
Baseline (All)	28	**97.18%**	98.30%	97.50%	97.90%	**0.9941**
PSO	٨	93.49%	95.42%	93.80%	94.60%	0.9816
Consensus	2	89.81%	93.38%	90.20%	91.70%	0.9556

The final 8-feature panel included: Age, Blood_Phosphorus, Blood_Uric_Acid, Blood_Creatinine, eGFR, Urine_pH, Urine_Calcium_24h, and Urine_Citrate_24h.
Figure ١ compares baseline (28 features) and PSO-optimized (8 features) models across five metrics. The optimized model preserves >93% of baseline performance with 71% feature reduction.

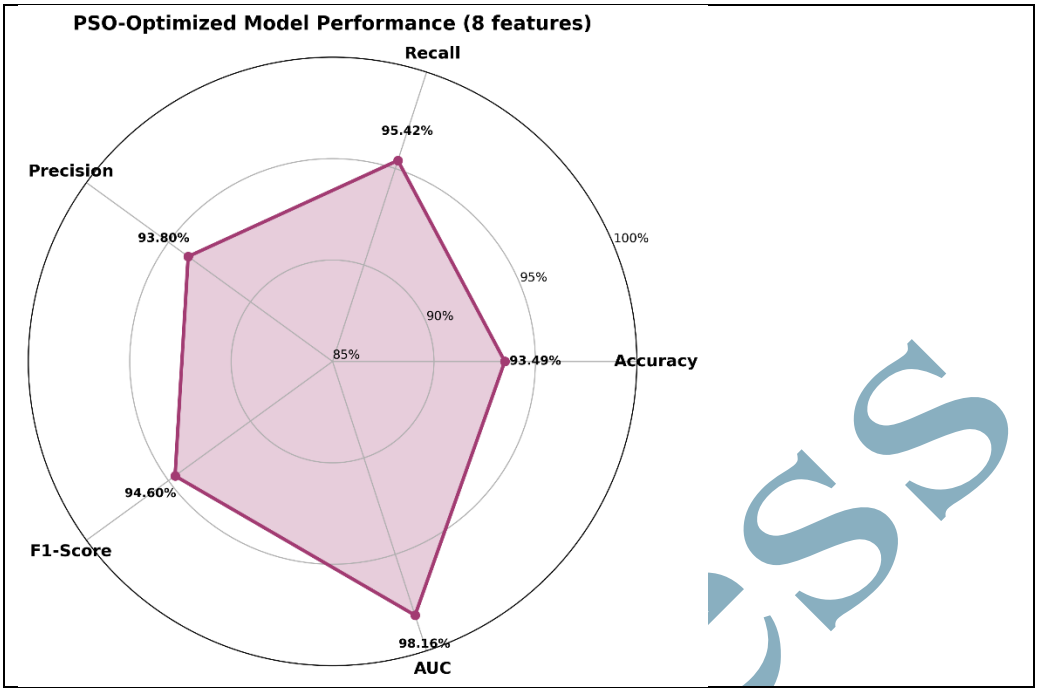

Figure1) Radar chart comparing model performance across five metrics

SHAP analysis for model interpretability

SHAP analysis identified the most influential predictors for kidney stone risk. Consistent with the PSO feature selection results, eGFR and Urine_Calcium_24h emerged as the most influential predictors. Higher values of Urine_Calcium_24h were associated with increased predicted risk, while higher Urine_Citrate_24h values were associated with decreased risk, confirming its protective role as a stone inhibitor.

Figure 2 presents the SHAP summary plot. eGFR and Urine_Calcium_24h are the most influential predictors. Elevated calcium increases risk; higher citrate decreases it.

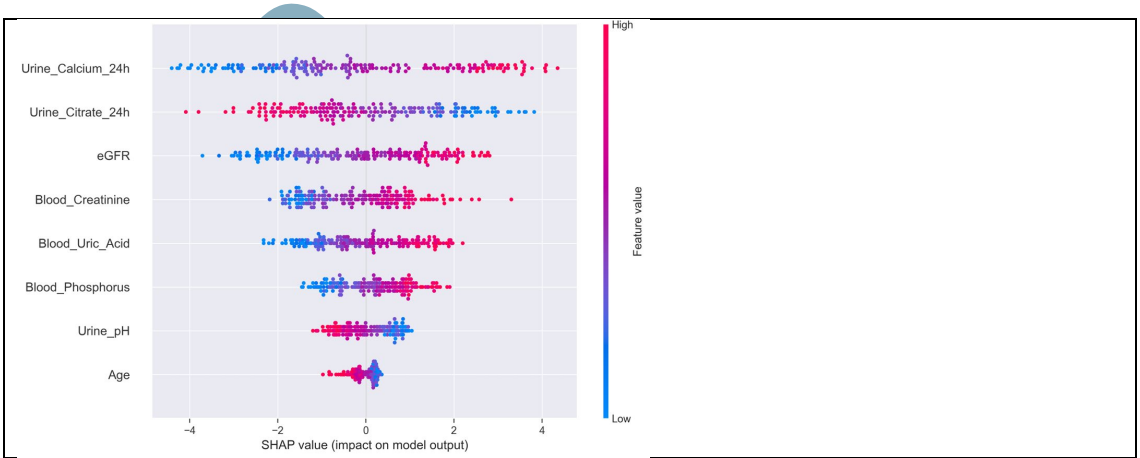

Figure2) SHAP summary plot of feature importance for the PSO-optimized XGBoost model

Discussion

This study successfully developed a PSO-optimized XGBoost model that reduced required laboratory tests from 28 to 8 features (71% reduction) while maintaining 93.49% accuracy and 95.42% recall, creating a practical, non-invasive screening tool suitable for primary care settings where 24-hour

urine collections are often impractical [7]. The 93.49% accuracy is comparable to previous machine learning applications in nephrolithiasis, including the 94.8% accuracy reported by Shabaniyan et al.[15] for PCNL outcome prediction and the 92.02% accuracy achieved by Enayati et al. [25] for health literacy prediction using PSO-based feature selection. However, our study extends this body of work by translating selected features into actionable patient education messages, bridging the gap between computational predictions and clinical practice. Unlike Kavoussi et al.[7], who achieved 98% accuracy using comprehensive urinary parameters, our model intentionally prioritizes accessibility over maximal accuracy, making it more feasible for routine primary care applications.

SHAP analysis identified eGFR and Urine_Calcium_24h as the most influential predictors, aligning with established pathophysiological mechanisms[3, 6]. Our model clearly reflects the centrality of urinary calcium supersaturation as the primary driver of calcium oxalate stones, which comprise approximately 84% of cases [1]. The protective role of Urine_Citrate_24h, evidenced by its inverse relationship with predicted risk, confirms its well-documented function as a potent inhibitor of calcium oxalate crystallization[6]. Critically, the interaction between elevated calcium and low citrate creates synergistic risk amplification, providing a compelling rationale for combined dietary and pharmacological interventions.

The clinical utility of our tool lies in translating laboratory findings into evidence-based, patient-friendly recommendations. Elevated urinary calcium (>6.5 mmol/day) warrants dietary sodium reduction or thiazide diuretic therapy, while low urinary citrate (<2.5 mmol/day) can be addressed through increased citrus fruit consumption or potassium citrate supplementation[6]. Reduced eGFR (<60 mL/min/1.73m^2) necessitates blood pressure management and renal function monitoring, and elevated serum uric acid (>0.42 mmol/L) requires purine restriction and increased hydration. As demonstrated by Harianto et al.[8] and Ayaad et al.[27], personalized health education significantly enhances patient engagement, treatment adherence, and self-management capabilities. Our framework provides the necessary evidence-based bridge between laboratory findings and these effective counseling strategies, enabling a more proactive, patient-centered approach to kidney stone disease management.

Several limitations should be acknowledged. The retrospective, single-center design may limit generalizability, as data were collected from specialized referral centers where patient demographics may differ from primary care populations. The absence of stone composition data prevented subgroup analysis for different stone types (e.g., calcium oxalate vs. uric acid), which may have distinct metabolic profiles and potentially affect model performance. Additionally, the PSO fitness function used KNN as the internal evaluator rather than the final XGBoost model, which may not have optimized the feature subset specifically for the gradient boosting algorithm. Finally, the model provides binary risk classification without predicting stone type or severity, limiting its utility for tailored interventions.

Future research should prioritize prospective multi-center validation to confirm the robustness and generalizability of our findings across diverse populations. Integration with imaging data and stone composition analysis could enable more comprehensive risk stratification and subtype-specific counseling. Cost-effectiveness studies are also warranted to support health policy decisions regarding the incorporation of this screening tool into routine primary care guidelines. Despite these limitations, our study provides compelling evidence that machine learning-based screening using routine laboratory data can effectively identify high-risk individuals while empowering patient education, representing a significant step toward personalized preventive care in nephrolithiasis.

Conclusion

This study demonstrates that a machine learning-based screening tool can identify high-risk individuals from routine laboratory data, providing actionable insights for patient education and preventive counseling in kidney stone disease. The PSO-optimized 8-feature panel offers a cost-effective, non-invasive approach for primary care settings, representing a 71% reduction in required laboratory tests while maintaining 93.49% accuracy. eGFR and Urine_Calcium_24h emerged as consensus predictors, confirming their central role in stone pathophysiology. Prospective multi-center validation is warranted before clinical deployment.