Mahmoodi *et al. BMC Research Notes* (2024) 17:318
https://doi.org/10.1186/s13104-024-06979-2

BMC Research Notes

RESEARCH NOTE

Predicting symptomatic kidney stones using machine learning algorithms: insights from the Fasa adults cohort study (FACS)

Fatemeh Mahmoodi[1], Aref Andishgar[2], Eisa Mahmoudi[3], Alireza Monsef[4], Sina Bazmi[1,2] and Reza Tabrizi[4,5*]

Abstract

Objectives To enhance the identification of individuals at risk of developing clinically significant kidney stones.

Methods In this study, data from the Fasa Adults Cohort Study were analyzed to explore factors linked to symptomatic and clinically significant kidney stone disease. After cleaning, 10,128 participants with 103 variables were studied. One outcome variable (presence of symptomatic kidney stones) and 102 predictor variables from surveys and tests were assessed. Five Machine learning (ML) algorithms (SVM, RF, KNN, GBM, XGB) were applied to examine kidney stone factors, with performance comparisons made. Data balancing was done using SMOTE, and metrics like accuracy, precision, sensitivity, specificity, F1 score, and AUC were evaluated for each algorithm.

Results The XGB model outperformed others with AUC of 0.60, while RF, GBM, SVC, and KNN had AUC values of 0.58, 0.57, 0.54, and 0.52. RF, GBM, and XGB showed good accuracy at 0.81, 0.81, and 0.77. Top predictors for kidney stones were serum creatinine, salt intake, hospitalization history, sleep duration, and BUN levels.

Conclusions ML models show promise in evaluating an individual's risk of developing painful kidney stones and recommending early lifestyle changes to reduce this risk. Further research can enhance predictive accuracy and tailor interventions for better prevention/management.

Keywords Kidney stones, Machine learning, Fasa Adult Cohort Study, Prediction

Introduction

Kidney stone disease is a significant global healthcare concern [1], with a growing number of patients affected by kidney stones and their associated complications each year [2]. This condition not only affects individual health but also imposes substantial financial burdens on healthcare systems [3]. The recurring nature of kidney stones presents challenges for both physical suffering and community well-being [4]. Despite advancements in the medical and technological fields [5], accurately predicting kidney stone disease remains a challenge for healthcare professionals and researchers [6].

The complexity of kidney stone disease, influenced by factors such as genetics, diet, lifestyle, environment, and

*Correspondence:
Reza Tabrizi
kmsrc89@gmail.com
[1]Student Research Committee, Fasa University of Medical Sciences, Fasa, Iran
[2]USERN Office, Fasa University of Medical Sciences, Fasa, Iran
[3]Department of Statistics, Yazd University, Yazd, Iran
[4]Clinical Research Development Unit, Valiasr Hospital, Fasa University of Medical Sciences, Fasa, Iran
[5]Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa 74616-86688, Iran

demographics, requires a comprehensive understanding to develop effective prevention strategies and early interventions [1]. Machine Learning (ML), a branch of artificial intelligence involving processing large datasets through statistical analysis, offers a potential solution to enhance predictive capabilities for kidney stone disease [7].

ML algorithms can identify intricate patterns and relationships within diverse datasets, aiding in the identification of individuals at risk for kidney stones and enabling targeted interventions [8].

Review of literature and previous studies

Studies by Abraham et al. [9], Chmiel et al. [10], and Kazemi et al. [11], have demonstrated the effectiveness of machine learning models in predicting kidney stone composition and guiding treatment based on influential parameters. Specifically, Abraham's study investigated the accuracy of machine learning models in predicting kidney stone composition using 24-hour urine data and demographic and comorbidity information extracted from electronic health records (EHR). The study found that the XGBoost model performed better with higher accuracy (91% vs. 71% for logistic regression) in predicting binary stone types (calcium vs. non-calcium stones). Additionally, for multiclass prediction, logistic regression outperformed XGBoost, indicating that the use of more comprehensive data can enhance prediction accuracy. This study highlights the importance of 24-hour urine data in predicting kidney stone composition and selecting appropriate treatment.

Furthermore, Zhu et al. investigated the use of CT-based radiomics and machine learning for screening individuals at high risk of kidney stones. This study focuses on employing advanced imaging techniques and analyzing imaging data to aid in more accurate and timely identification of individuals at risk for kidney stones. By utilizing CT radiomics, which involves extracting complex features and patterns from CT scan images, this research contributes to new advancements in screening and predicting kidney stones and highlights ongoing progress in this field [12].

It is essential to note that not all kidney stones cause pain and clinical problems [13], and the major psychological, physical, and financial burden of kidney stones is related to the severe pain they cause [3]. Although some large asymptomatic stones can lead to long-term impairment of kidney function [14], many asymptomatic stones are small and pass spontaneously without causing pain [15]. Considering these factors and the lack of imaging data for kidney stones, our study focuses solely on symptomatic kidney stones reported by individuals. These individuals were diagnosed by a physician based on radiology after presenting with severe pain, as indicated in our questionnaire.

Thus, this study utilized data from the Fasa Adults Cohort Study (FACS) in a high prevalence community to explore the relationship between various variables and predict clinically significant kidney stone formation. By applying advanced machine learning techniques and meticulous data analysis, the study aims to provide a comprehensive understanding of the effectiveness and limitations of machine learning algorithms in predicting symptomatic kidney stones. This research seeks to reduce the burden of kidney stone disease and improve healthcare strategies by uncovering the intricate relationship between predictive variables and symptomatic kidney stone formation.

Materials and methods

Data description

This cross-sectional study utilized data from the Fasa Adults Cohort Study (FACS), which focused on analyzing risk factors for non-communicable diseases (NCDs) in Fasa's rural population. Specifically, the study targeted the rural community of "Sheshdeh", consisting of 41,000 inhabitants, with a focus on individuals aged 35 to 70 (11,097 people) Inclusion of data required comprehensive information from participants. The study design, as outlined in Fig. 1, provides a visual representation of the research methodology and procedures adopted in the present study.

Data preprocessing

Data preprocessing involved three main steps: data cleaning, feature scaling, and one-hot encoding. Outliers were removed, variables with more than 10% missing data were eliminated, and multiple imputation was performed for variables with missing data below the threshold. The final dataset consisted of 10,128 individuals with 103 variables. Continuous variables were scaled, and variables with more than two categories were converted to dummy variables using one-hot encoding.

Outcome variable

The outcome variable in this study was the presence or absence of kidney stones, obtained through questionnaires completed by participants in the cohort study. The outcome variable was treated as a binary variable (with kidney stone/without kidney stone). Previous studies have shown that self-reporting of kidney stone status is reliable [16].

Predictors

The study included 102 independent variables (Table S1) covering a wide range of factors obtained through various methods detailed in the FACS. Both laboratory

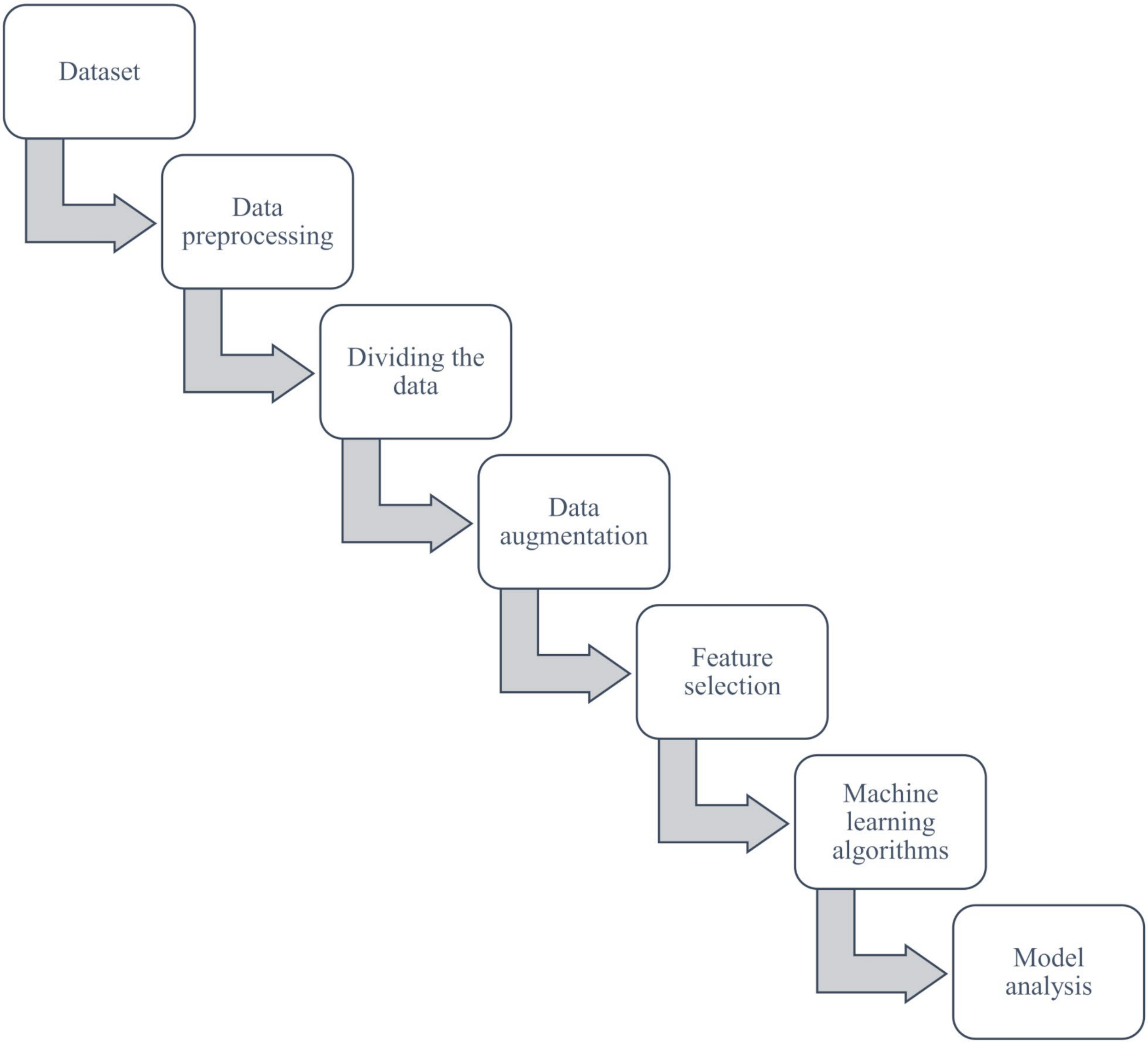

Fig. 1 Flowchart of the study

and non-laboratory variables were incorporated, with blood samples collected from fasting individuals and non-laboratory characteristics gathered through standard questionnaires. Blood pressure and anthropometric measurements were also taken by a professional nurse (Table 1s).

Data division

The preprocessed data were divided into a training set (80% of the data) and a test set (20% of the data). The training set was utilized for feature selection, hyperparameter tuning, 5-fold cross-validation, and training the ML algorithms.

Data augmentation

To address the imbalance in the training dataset, where there were 8,371 participants without kidney stones and 1,757 patients with kidney stones, the synthetic minority over-sampling technique (SMOTE) was employed. SMOTE generates synthetic instances of the minority class by creating samples along the line segment connecting the minority class with its k nearest neighbors [17]. Through this process, 6,614 instances with kidney stones were generated, resulting in a balanced dataset with 8,371 patients with kidney stones and 8,371 without kidney stones.

Feature selection

Feature selection methods were applied to reduce data dimensions and identify the most relevant predictors for kidney stone prediction. Permutation importance and a tree-based ML model (random forest) were employed to identify the top 20 features as the best predictors. Descriptive analysis of the selected characteristics was conducted using statistical tests such as independent samples T-Test, Chi-Square Test, and Mann-Whitney Test, with statistical significance defined as *P*-values below 0.05. Data analysis was performed using SPSS version 18 (IBM Corp., Armonk, N.Y., USA).

Machine learning algorithms

Five supervised ML algorithms - Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting Machine (GBM), and XGBoost (XGB) - were employed using Anaconda (version 4.12.0) on Jupyter Notebook Platform (version 3.3.2), and Scikit-Learn Module (version 1.1.3). Hyperparameter tuning was conducted using grid search with 5-fold cross-validation on the training data. The ML models were then trained using the optimal parameters identified through this process.

Model analysis

The trained models were evaluated using the test data, with accuracy, precision, sensitivity, specificity, F1-score, and area under the curve (AUC) used to assess and compare the models. The confusion matrices of the best model and feature ranking of the top predictors were presented for analysis.

Results

Descriptive analysis of selected variables

Table 1 presents a comparison of various characteristics between participants with kidney stones ($N=1,757$) and those without kidney stones ($N=8,371$). Significant differences were observed between the two groups. Participants with kidney stones were found to be significantly older (49.42 vs. 48.47 years, $p<0.001$) and had a higher proportion of men (49.1% vs. 44.3%, $p<0.001$). Socio-economic status also differed significantly ($p=0.005$), with fewer individuals in the kidney stone group belonging to

Table 1 General and selected characteristics of the participants according to prediction of kidney stone (total number of participants = 10128)

Variables		Kidney stone		*P*-value
		No (*N* = 8371)	**Yes (*N* = 1757)**	
Age, years*		48.47 ± 9.55	49.42 ± 9.64	≤ 0.001[a]
Sex*	Man	3712 (44.3)	862 (49.1)	≤ 0.001[b]
	Woman	4659 (55.7)	895 (50.9)	
Socio-economic status	Low-income	2130(24.3)	402(22.8)	0.005[b]
	Middle-income	4197(50.1)	867(49.3)	
	High-income	2044(23.4)	488(27.7)	
BMI, kg/m^2		25.5 ± 4.86	26.23 ± 4.80	≤ 0.001[a]
White blood cell count, ×10^3/mm^3		6.64 ± 1.72	6.54 ± 1.67	≤ 0.073[a]
Hematocrit, %		41.76 ± 4.56	42.49 ± 4.49	≤ 0.001[a]
MCV, femtoliter		84.747 ± 8.31	84.90 ± 7.66	0.474[a]
GGT, IU/L		22.67 ± 22.23	23.38 ± 16.10	0.117[a]
Alkaline phosphatase level, U/L		209.38 ± 72.91	211.16 ± 62.66	0.294[a]
BUN, mg/dL		12.82 ± 4.00	13.31 ± 3.85	≤ 0.001[a]
Serum creatinine level, mg/dL		0.97 ± 0.19	1.00 ± 0.19	≤ 0.001[a]
Cholesterol, mg/dL		184.54 ± 39.49	185.94 ± 39.29	0.174[a]
Triglyceride, mg/ dL		130.42 ± 82.82	138.09 ± 79.75	≤ 0.001[a]
HDL-C, mg/ dL		5103 ± 15.73	51.02 ± 16.49	0.978[a]
Sugar products consumption, gr/day		56.59 ± 58.76	50.73 ± 54.57	≤ 0.001[a]
Grain products consumption, gr/day		666.26 ± 322.42	673.35 ± 330.24	0.412[a]
Salt consumption, gr/day		4.08 ± 2.72	3.92 ± 2.48	0.025[a]
Dairy products consumption, gr/day		213.91 ± 180.57	205.79 ± 180.18	0.086[a]
Meats consumption, gr/day		98.93 ± 69.10	94.92 ± 64.26	0.025[a]
Fruits consumption, gr/day		393.30 ± 321.56	411.19 ± 334.32	0.040[a]
Past medical history of hospitalization, yes		2563(30.6)	344(19.6)	≤ 0.001[b]
Sleep duration, hours/day		6.99 ± 1.59	6.90 ± 1.60	0.052[a]

Data was presented as Mean ± SD, and Number (%). Statistical analyses such as a: Independent Samples Test and b: Chi-Square were used. BMI; Blood Mass Index, BUN; Blood Urea Nitrogen, MCV; Mean Corpuscula Volume, GGT; Gamma Glutamyl Transferase, HDL-C; High Density Lipoprotein-Cholesterol

*Age and sex are not one of top-20 variables and they are only presented as the demographic variables of the study the high-income category. Furthermore, the kidney stone group had higher values in various variables compared to the non-kidney stone group, including BMI (26.23 vs. 25.54 kg/m², $p<0.001$), hematocrit (42.49% vs. 41.76%, $p<0.001$), BUN (13.31 vs. 12.82 mg/dL, $p<0.001$), serum creatinine (1.00 vs. 0.97 mg/dL, $p<0.001$), triglycerides (138.09 vs. 130.42 mg/dL, $p=0.001$), sugar product consumption (70.73 vs. 56.59 g/day, $p<0.001$), salt consumption (3.92 vs. 4.08 g/day, $p=0.025$), meat consumption (94.92 vs. 98.93 g/day, $p=0.025$), fruit consumption (411.19 vs. 393.30 g/day, $p=0.040$), and past medical history of hospitalization ($p<0.001$).

However, no significant differences were observed in some variables such in white blood cell count ($p=0.073$), MCV ($p=0.474$), GGT levels ($p=0.117$), alkaline phosphatase levels ($p=0.294$), cholesterol levels ($p=0.174$), HDL-C levels ($p=0.978$), grain product consumption ($p=0.412$), dairy product consumption ($p=0.086$), and sleep duration ($p=0.052$).

ML algorithms performance

Among the ML algorithms used, XGB exhibited the highest AUC of 0.60, with RF, GBM, SVC, and KNN achieving lower AUC values. RF, GBM, and XGB models demonstrated acceptable accuracy levels of 0.81, 0.81, and 0.77, respectively (Fig. 2.**A**). AUC of XGB model for training and test dataset is indicated in Fig. 2.**B.**, and confusion matrix of XGB model for predicting symptomatic kidney stone is demonstrated in Fig. 2.**C.** Additional details on precision, specificity, sensitivity, and F1-score can be found in Table 2.

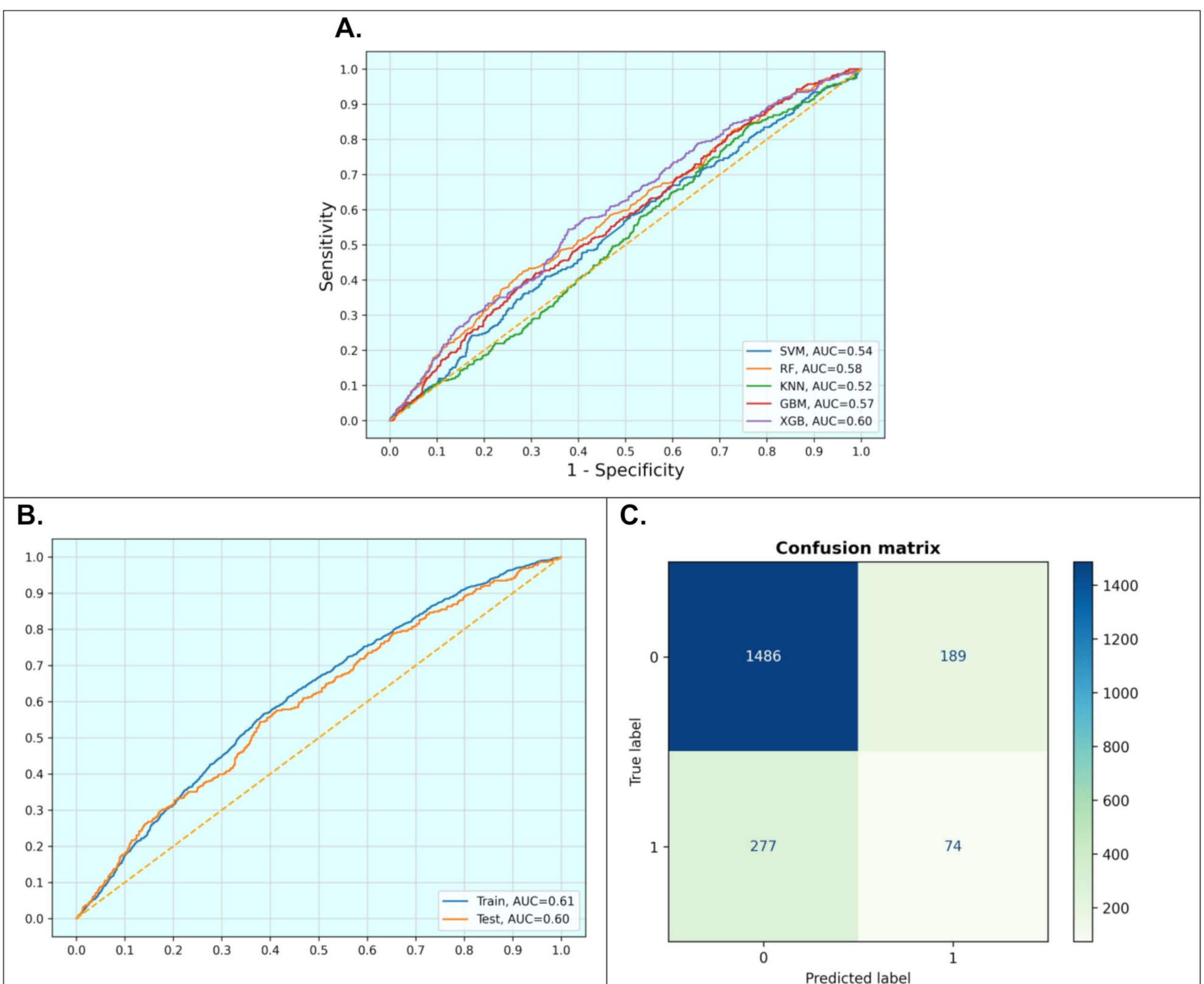

Fig. 2 (**A**) Area under curve (AUC) of all machine learning algorithms. (**B**) AUC of XGB model for training and test dataset. (**C**) Confusion matrix of XGB model for predicting kidney stone

Table 2 Performance of machine learning algorithms

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	AUC
SVC	0.45 (0.42–0.48)	0.19 (0.17–0.21)	0.66 (0.63–0.69)	0.4 (0.37–0.43)	0.27 (0.24–0.30)	0.54 (0.51–0.57)
RF	0.81 (0.78–0.84)	0.26 (0.24–0.28)	0.06 (0.05–0.07)	0.96 (0.95–0.97)	0.1 (0.09–0.11)	0.58 (0.55–0.62)
KNN	0.48 (0.45–0.51)	0.18 (0.16–0.2)	0.58 (0.55–0.61)	0.46 (0.43–0.49)	0.28 (0.25–0.31)	0.52 (0.48–0.55)
GBM	0.81 (0.78–0.84)	0.22 (0.2–0.24)	0.04 (0.03–0.05)	0.97 (0.96–0.98)	0.07 (0.06–0.08)	0.57 (0.54–0.6)
XGB	0.77 (0.74–0.8)	0.28 (0.25–0.31)	0.21 (0.19–0.23)	0.89 (0.87–0.91)	0.26 (0.24–0.28)	0.6 (0.56–0.63)

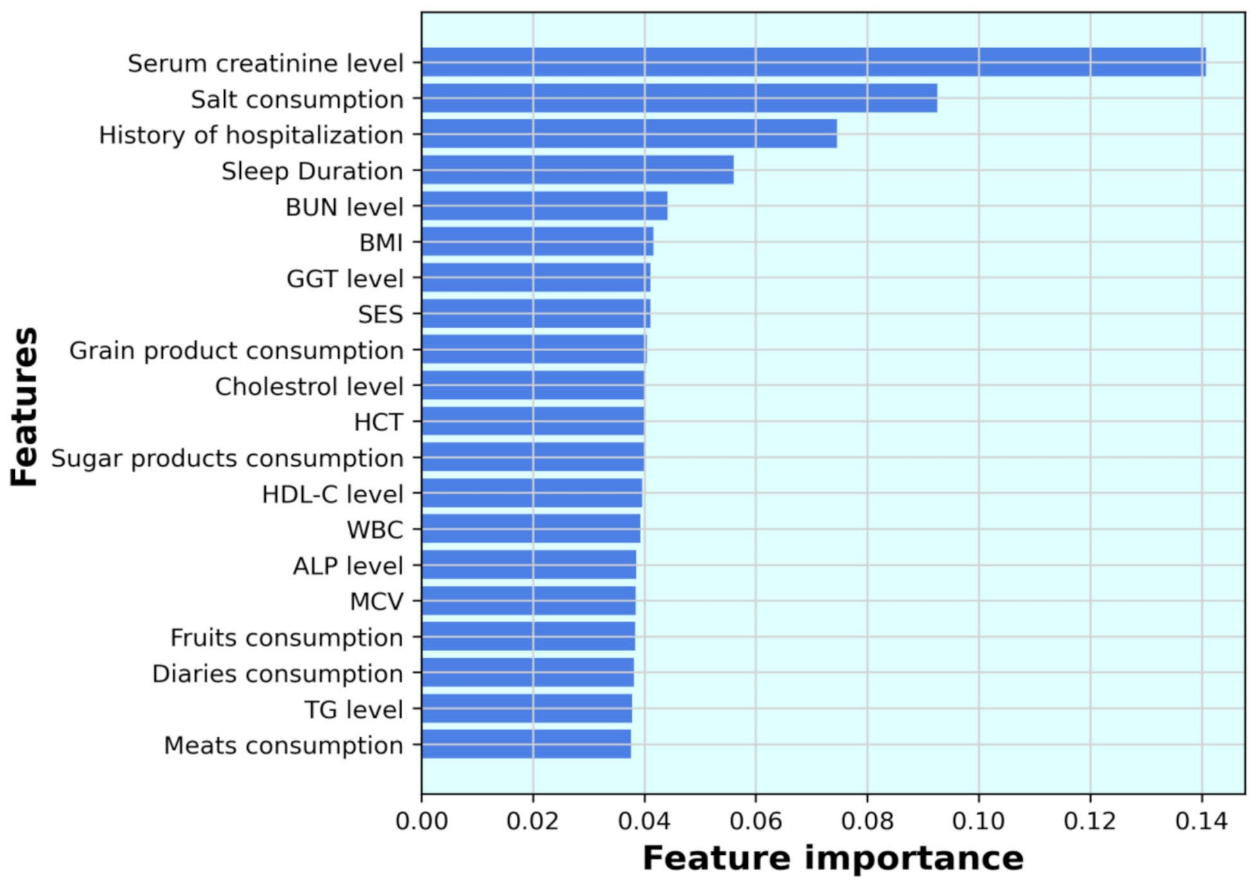

Fig. 3 Feature importance of top-20 predictors

Feature importance

The importance of predictors in predicting kidney stones was assessed using the XGB model and the permutation importance technique, resulting in a weighted ranking of the variables (Fig. 3). The top five most influential variables, in descending order, were serum creatinine level, salt consumption, and history of hospitalization, sleep duration, and BUN level.

Discussion

The present study aimed to predict the risk of kidney stone formation in a rural population using ML techniques, based on data from the FACS involving 11,097 participants aged 35 to 70 years. This research is the first of its kind to investigate the factors associated with kidney stones in the Iranian population and evaluate the performance of ML models using cross-sectional data. Our study results, building on previous research, provide new insights into the risk of kidney stone formation.

After data cleansing and removal of low-information variables, a total of 10,128 individuals with 103 variables were prepared for analysis. The study utilized five different ML algorithms, with the GBM model achieving the best performance, identifying the top twenty features as important predictors. The model's accuracy, with an AUC of 0.60, was highly noteworthy.

In previous studies, decision trees and multi-layer neural networks were identified as effective ML models for predicting kidney stone risk. However, in present study, the GBM model emerged as the optimal choice. The study conducted by Kaladhar et al., data related to kidney stones were analyzed using a dataset of 10,000 cases, consisting of 5,000 positive and negative instances. The researchers employed 42 features to develop a predictive model for kidney stone risk. The decision tree algorithm achieved an impressive performance of 93%, while the support vector machine reached 91.9%. Several significant risk factors were identified in the study, including marital status, tea and milk consumption, blood type B+, walking, consumption of sweet water, bathing with hot water, and consuming non-green foods as breakfast and dinner [18]. Similarly, another study conducted in Nigeria utilized data mining techniques to explore kidney stone risk among Nigerians. Initially, endocrinologists and questionnaires were used to identify the risk factors associated with kidney stones. The data were then processed using three different algorithms: decision tree, multilayer perceptron neural network, and genetic algorithm. The results demonstrated that the multilayer perceptron neural network achieved remarkable accuracy of 100% using 33 initial variables identified by the endocrinologists, making it the best-performing algorithm. This study emphasized the superiority of the multilayer perceptron neural network for developing predictive models of kidney stone risk compared to other algorithms [19].

Our best model confirmed that five primary factors, namely serum creatinine level, salt intake, hospitalization history, sleep duration, and BUN level, were significantly associated with the risk of kidney stone formation.

Serum creatinine levels are a fundamental indicator for assessing kidney function and damage. Serum creatinine, a byproduct of muscle metabolism, is excreted by the kidneys and its levels are directly related to the kidneys' filtering performance [20] .Elevated serum creatinine levels indicate a reduction the kidneys' ability to filter toxins and waste products from the blood, which can lead to the accumulation of substances that contribute to kidney stone formation [21]. The positive correlation between high serum creatinine levels and urinary crystals such as calcium, magnesium, phosphate, and uric acid highlights its significance in diagnosing both kidney function decline and kidney stone formation. Increased levels of serum creatinine are associated with elevated concentrations of these substances in the urine, which are key factors in the formation of kidney stones, particularly calcium stones [22]. A study by Xudong et al. demonstrated that elevated serum creatinine levels pose a moderate risk of kidney stone formation, especially among adults, middle-aged individuals, the elderly, and individuals of Caucasian descent [23]. This finding underscores the importance of serum creatinine as a key marker in assessing the risk of kidney stones and chronic kidney disease. Monitoring and managing serum creatinine levels can significantly aid in the early detection and prevention of kidney stones and other related kidney dysfunctions.

Additionally, similar research has shown that high serum creatinine levels are associated with reduced glomerular filtration rate (GFR) and disturbances in urinary electrolyte and metabolite regulation, conditions that further contribute to kidney stone formation [24]. Elevated serum creatinine is also linked to increased urine acidity and decreased urine volume, both of which are important factors in the formation of kidney stones, particularly calcium stones. This comprehensive understanding highlights the role of serum creatinine in both diagnosing and managing kidney stone disease and underscores the need for careful monitoring of this indicator in clinical settings.

One of the unexpected and intriguing findings in this study was the inverse relationship between salt consumption and the risk of kidney stone formation. Contrary to the common expectation that increased salt intake may lead to a higher risk of stone formation, this study found that salt consumption was associated with a reduced risk of kidney stones [25]. This discrepancy may be attributed to the specific characteristics of the study and the behavioral changes of patients after experiencing disease symptoms. Typically, high salt intake is associated with increased urinary calcium excretion, which is linked to a higher risk of kidney stones. Numerous studies have shown that excessive salt consumption can lead to elevated calcium levels in the urine and, consequently, an increased risk of forming calcium stones. However, the inverse relationship observed in this study might be related to behavioral changes in patients following the onset of symptoms. For instance, patients may spontaneously reduce their salt intake after being diagnosed with kidney stones or experiencing symptoms, in an attempt to prevent similar issues. This dietary behavior change could result in a lower observed salt consumption in those who have experienced kidney stones than what would be expected under normal conditions. This issue may contribute to the contradictory results observed in cross-sectional studies.

These findings underscore that cross-sectional studies may not fully reveal causal relationships and highlight the importance of longitudinal study designs to better understand the complex interactions between dietary behaviors and chronic disease incidence. Longitudinal studies can more accurately track changes in salt consumption and other dietary behaviors and examine the long-term effects of these changes on kidney stone risk [26]. Such studies are better equipped to identify true patterns and causal relationships between salt intake and kidney stone formation and will be more effective in informing preventive strategies and disease management.

This study also highlighted the significant impact of sleep disturbances and circadian rhythm disruptions on the risk of kidney stone formation. Disruptions in sleep and changes in sleep patterns can have extensive effects on overall health, particularly in relation to increased risk of chronic diseases, including kidney stones. Variations in sleep quality and duration can lead to hormonal imbalances and alterations in the body's biological regulation, which may contribute to an increased risk of kidney stone formation.

Various studies have demonstrated that sleep disturbances can affect metabolic processes and lead to changes in the secretion of hormones such as calcitonin and parathyroid hormone, which in turn can impact urinary calcium levels and other components, creating a conducive environment for kidney stone formation. For example, research by Si-ke He and colleagues indicated that sleep disturbances can directly affect calcium metabolism and increase the risk of kidney stones [27].

These findings clearly indicate that assessing and optimizing sleep patterns can play a crucial role in preventing kidney diseases. Improving sleep quality and implementing targeted interventions to manage sleep disorders may help reduce the risk of kidney stones and enhance overall kidney health. Utilizing modern techniques for monitoring and improving sleep, including behavioral and psychological interventions, could be effective. Future research should further investigate the relationship between sleep quality and kidney stone formation and develop effective strategies for managing these disturbances to reduce kidney disease risk. Ultimately, these studies emphasize that a more detailed examination of sleep disturbances and circadian rhythm disruptions, along with the development of preventive strategies based on improving sleep quality, could contribute significantly to reducing the risk of kidney diseases and enhancing overall health.

Blood Urea Nitrogen (BUN) is a key biomarker used to assess renal function and is influenced by various metabolic processes. Elevated BUN levels can reflect impaired kidney function, which is often associated with chronic kidney disease and other renal disorders. Recent studies have investigated the role of BUN in the formation of kidney stones, providing insights into its potential as a predictor for stone formation [28]. BUN is a byproduct of protein metabolism and is normally excreted by the kidneys. Increased levels of BUN in the blood indicate reduced renal clearance and potential kidney dysfunction. Elevated BUN levels have been linked to a higher risk of kidney stone formation through several mechanisms. Firstly, impaired renal function can lead to decreased urine output and concentrated urine, which may increase the risk of stone formation due to higher concentrations of stone-forming substances such as calcium, oxalate, and uric acid [29].

While several study have explored the association between kidney function and stone formation, specific study directly linking elevated BUN levels to kidney stone development remain limited.

The available literature generally supports the notion that impaired renal function, which can be reflected by increased BUN levels, correlates with higher incidence of nephrolithiasis [30].

This combined text presents an overview of the association between BUN levels and kidney stone formation, highlighting the current understanding and the need for more focused research in this area.

ML serves as a vital tool in transforming health-related data into actionable knowledge, offering significant improvements in clinical performance. It assists physicians and researchers in identifying hidden patterns and relationships within data, optimizing their use in diagnosis, prediction, and disease management [31]. In addition to evaluating the performance of ML models, assessing the financial benefits associated with each model is also crucial and should be investigated in future research.

Limitations

- Cross-sectional design restricts the ability to provide long-term follow-up, which can be particularly problematic in accurately assessing the timing of kidney stone risk in future patient [24].
- Due to constraints in accessing comprehensive datasets from various cohorts, we were unable to conduct external validation of our models. Therefore, it is important to exercise caution when interpreting the findings of the current study, taking into account its limitations [25].
- Interpreting the clinical significance of some of the prediction variable differences is challenging, as in many cases, the overall differences between individuals with kidney stones and those without were minimal, which needs to be taken into account.
- Patients' self-reported history of stones may not align accurately with the timepoints of their self-reported diets and lifestyle. This issue could potentially be addressed in future studies by incorporating survival analyses to examine the temporal relationship between diet, lifestyle, and the incidence of kidney stones.
- There is a possibility of a significant time gap between the onset of kidney stones and the date of each factor measurement, emphasizing the need for caution when interpreting our results.

Future research should incorporate survival analyses to explore temporal relationships between diet, lifestyle, and kidney stones and investigate the financial benefits of different models. Using objective imaging studies like CT scans and ultrasounds could enhance diagnosis and prediction of asymptomatic stones. CT scans offer precise information about stone size, location, and characteristics, while ultrasounds provide non-invasive, radiation-free assessments, especially useful for routine follow-ups and conditions where radiation is unsuitable.

Combining these imaging methods could improve diagnosis, monitoring, and management of kidney stones. Continued research and application of machine learning techniques, considering these limitations, are essential for improving accuracy and reducing healthcare costs.

Conclusion

In conclusion, the predictive models developed in this study using machine learning techniques have provided valuable insights into the risk factors associated with kidney stone formation in the Iranian population. The findings underscore the significance of serum creatinine levels, salt intake, history of hospitalization, sleep duration, and BUN levels as key predictors of kidney stone risk. These results contribute to the existing knowledge base and highlight the potential of machine learning in predicting and managing kidney stone disease. By leveraging machine learning techniques for predictive modeling and risk assessment, healthcare professionals can enhance diagnostic accuracy, facilitate early intervention, and ultimately reduce the burden of kidney stone disease on healthcare systems and individuals alike. Continued research and advancements in this field are crucial to improving patient outcomes and healthcare strategies for kidney stone management.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-024-06979-2.

Supplementary Table S1: All characteristics and clinical features of participants

Acknowledgements

This study was supported by the Deputy of Research and Technology of Fasa University of Medical Sciences, Fasa, Iran (No. 401206).

Author contributions

Conceptualization: R.T. Methodology: A.A. Software: A.A. Validation: R.T., A.M., S.B.Formal analysis: A.A. Investigation: F.M., E.M. Resources: R.T. Writing (original draft preparation): F.M., E.M., and A.A.; Writing (review and editing): S.B., A.M. Visualization: A.A. Supervision: S.B., R.T. and A.M. Project administration: R.T.; All authors have read and approved the final version.

Funding

This research received no specific grant from any funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

The Fasa University of Medical Sciences Research Council and Ethics Committee approved this study protocol (Ethical code: IR.FUMS.REC.1401.196). The research procedures were conducted in accordance with all relevant guidelines and regulations for human research (Helinski). Before participating in FACS, individuals were asked to give written informed consent for their information to be used in and future studies, which was reviewed and approved by the research ethics committee. Participant personal information was collected from the system, ensuring that the identities of all individuals remained anonymous.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 July 2024 / Accepted: 15 October 2024
Published online: 24 October 2024

References

1. Caglayan A, et al. Deep learning model-assisted detection of kidney stones on computed tomography. Int braz j urol. 2022;48:830–9.
2. Smith A, et al. A nephrolithometric nomogram to predict treatment success of percutaneous nephrolithotomy. J Urol. 2013;190(1):149–56.
3. Roberson D, Sperling C, Shah A, Ziemba J. Economic considerations in the management of nephrolithiasis. Curr Urol Rep. 2020;21:1–9.
4. Tasian GE, Kabarriti AE, Kalmus A, Furth SL. Kidney stone recurrence among children and adolescents. J Urol. 2017;197(1):246–52.
5. Rezaei M, et al. Role of artificial intelligence in the diagnosis and treatment of diseases. Kindle. 2023;3(1):1–160.
6. Goldfarb DS. The exposome for kidney stones. Urolithiasis. 2016;44(1):3–7.
7. Chen X, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79:102444.
8. Ngiam KY, Khor W. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20(5):e262–73.
9. Abraham A, et al. Machine learning prediction of kidney stone composition using electronic health record-derived features. J Endourol. 2022;36(2):243–50.
10. Chmiel JA, et al. Predictive modeling of urinary stone composition using machine learning and Clinical Data: implications for treatment strategies and pathophysiological insights. Journal of Endourology; 2024.
11. Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med. 2018;84:117–26.
12. Zhu B, et al. CT-based radiomics of machine-learning to screen high-risk individuals with kidney stones. Urolithiasis. 2024;52(1):91.
13. Bansal AD, Hui J, Goldfarb DS. Asymptomatic nephrolithiasis detected by ultrasound. Clin J Am Soc Nephrol. 2009;4(3):680–4.
14. Noh TI, et al. A comparison between asymptomatic and symptomatic ureteral stones. Sci Rep. 2023;13(1):2757.
15. Stritt K, et al. Risk factors for asymptomatic kidney stone passage in adults with recurrent kidney stones. Clin J Am Soc Nephrol. 2024;102215:p.
16. Li G, et al. Association between body fat distribution and kidney stones: evidence from a US population. Front Endocrinol. 2022;13:1032323.
17. Nakamura M, Kajiwara Y, Otsuka A, Kimura H. Lvq-smote–learning vector quantization based synthetic minority over–sampling technique for biomedical data. BioData Min. 2013;6:1–10.
18. Krishna K, Rayavarapu A, Vadlapudi V. Statistical and data mining aspects on kidney stones: a systematic review and meta-analysis. Open Access Sci Rep, 2012. 1(12).
19. Oladeji F et al. *Model for predicting the risk of kidney stone using data mining techniques.* 2019.
20. Levey AS, Coresh J. Chronic kidney disease. Lancet. 2012;379(9811):165–80.
21. Floege J, Johnson RJ, Feehally J. Comprehensive clinical nephrology E-book. Elsevier Health Sciences; 2010.
22. Ms P. Urologic disease in America project: urolithiasis. J Urol. 2005;173:848–57.
23. Shen X, et al. The association of urine creatinine with kidney stone prevalence in US adults: data from NHANES 2009–2018. Front Med. 2022;9:p819738.

24. Wang S, et al. Association between sleep quality and urolithiasis among general population in Western China: a cross-sectional study. BMC Public Health. 2022;22(1):1787.
25. Ticinesi A, et al. Salt and nephrolithiasis. Nephrol Dialysis Transplantation. 2016;31(1):39–45.
26. Andishgar A, et al. Machine learning-based models to predict the conversion of normal blood pressure to hypertension within 5-year follow-up. PLoS ONE. 2024;19(3):e0300201.
27. He S-K, et al. Sleep and circadian rhythm disturbance in kidney stone disease: a narrative review. Front Endocrinol. 2023;14:1293685.
28. Manzoor MA, Mujeeburahiman M, Rekha P. Association of serum biochemical panel with mineralogical composition of kidney stone in India. Acta Med Int. 2017;4(2):26–30.
29. Moreira DM, et al. Association of serum biochemical metabolic panel with stone composition. Int J Urol. 2015;22(2):195–9.
30. Simmons KE, et al. Risk factors for common kidney stones are correlated with kidney function independent of stone composition. Am J Nephrol. 2023;54(7–8):329–36.
31. Kagawa R, et al. Development of type 2 diabetes mellitus phenotyping framework using expert knowledge and machine learning approach. J Diabetes Sci Technol. 2017;11(4):791–9.

Publisher's note