

# Detecting Kidney Stones Using Urine Test Analysis: A Machine Learning Perspective

Isaac Osei<sup>1</sup>, Acheampong Baafi-Adomako<sup>2</sup>, Dennis Opoku Boadu<sup>2</sup>

<sup>1</sup>Amity University

<sup>2</sup>University of Ghana

DOI: <https://doi.org/10.51244/IJRSI.2024.1110061>

Received: 02 October 2024; Accepted: 07 October 2024; Published: 20 November 2024

## ABSTRACT

Kidney stones, a prevalent urological condition, can cause severe discomfort and serious health complications if untreated. Traditional diagnostic methods, such as CT scans and ultrasounds, while effective, are often costly, expose patients to radiation, and may not be accessible in low-resource settings. This study explores a machine learning-based alternative that uses urine test data for kidney stone detection, aiming to provide a non-invasive, cost-effective, and accessible diagnostic tool. The study evaluates various machine learning models, including Random Forest (RF), Support Vector Machine (SVM), Logistic Regression, Decision Trees, and Gradient Boosting, to predict kidney stones using urine analysis data. Key urine parameters analyzed include specific gravity, pH, osmolality, conductivity, urea, and calcium concentrations. With a dataset of 79 samples, each labeled for kidney stone presence, preprocessing steps ensured data quality through normalization and exploratory analysis. Models were trained on 80% of the data and tested on the remaining 20%, with performance measured through accuracy, precision, recall, F1 score, and AUC-ROC metrics. The Random Forest model achieved the highest performance, with an accuracy of 94%, precision of 0.95, recall of 0.94, F1 score of 0.94, and AUC-ROC of 0.94, while Gradient Boosting achieved a slightly higher AUC-ROC at 0.96. Feature analysis identified osmolality and urea as the most significant predictors, followed by specific gravity and calcium concentration. These findings align with clinical knowledge on kidney stone formation. The high accuracy and reliability of the Random Forest model underscore its potential as a diagnostic tool for kidney stones. However, limitations include the need for larger datasets to improve generalizability and model transparency for clinical trust. Addressing these factors and facilitating integration into clinical workflows could enhance early detection, improve patient outcomes, and offer a promising alternative to traditional methods.

**Keywords:** Machine Learning, Classification Algorithm, Kidney Stones, Classification Algorithm, Random Forest, Support Vector Machine.

## INTRODUCTION

Kidney stones, or renal calculi, are solid formations resulting from the aggregation of minerals and salts within the kidneys. They can manifest anywhere along the urinary tract, from the kidneys to the bladder, often due to highly concentrated urine that facilitates the crystallization of minerals. The incidence of kidney stones is rising globally, leading to considerable health complications and increased healthcare expenses. It's projected that around 12% of people worldwide will experience a kidney stone during their lifetime, with recurrence rates for those affected being as high as 50% within five years of an initial episode (Romero et al., 2010; Pearle et al., 2014). Traditional diagnostic methods for kidney stones include imaging techniques like non-contrast computed tomography (CT), ultrasound, and X-rays. While these methods are generally effective, they come with drawbacks. CT scans, regarded as the gold standard, expose patients to ionizing radiation and can be expensive (Fulgham et al., 2013). Ultrasound, although less risky and more affordable, might miss smaller stones or provide less detailed imaging. This has sparked interest in developing non-invasive, cost-effective, and rapid diagnostic alternatives that could be utilized in primary care settings or even at home.

---

## Problem Statement

Kidney stones are a prevalent and recurrent urological condition that lead to significant pain, morbidity, and healthcare costs globally. While traditional diagnostic methods like computed tomography (CT) scans and ultrasounds are effective, they come with several limitations, including high costs, radiation exposure, and limited accessibility in resource-limited settings. These methods also require advanced medical infrastructure and skilled personnel, making them less practical for primary care or remote settings. Urine analysis, being non-invasive, cost-effective, and widely available, offers valuable insights into the biochemical conditions that predispose individuals to kidney stone formation. However, interpreting urine analysis data can be complex and demands sophisticated analytical methods to detect subtle patterns indicative of kidney stones. Despite the potential advantages, the use of machine learning for urine test analysis in kidney stone detection is still underexplored. Existing studies have been constrained by small sample sizes, data imbalances, and a lack of comprehensive feature sets. Moreover, integrating these models into clinical practice presents challenges related to model interpretability, data security, and the necessity for clinical validation.

## Objectives

The following are the research objectives:

- To determine the most significant urine analysis parameters (such as specific gravity, pH, osmolality, conductivity, urea, and calcium concentrations) that contribute to the accurate detection of kidney stones.
- To assess the performance of various machine learning models, including Random Forest, Support Vector Machine, Logistic Regression, Gradient Boosting etc., in detecting kidney stones using urine test data.
- To compare the accuracy, precision, recall, F1 score, and AUC-ROC of different machine learning models to identify the best performing model for this application.
- To create a robust machine learning model that can reliably predict the presence of kidney stones based on urine test analysis.
- To implement strategies to manage data imbalance in the dataset, ensuring that the model can accurately predict kidney stones in both balanced and imbalanced datasets.
- To compare the cost-effectiveness of using machine learning models for kidney stone detection and traditional diagnostic methods, aiming to provide a more accessible and affordable diagnostic tool.
- To deploy and integrate the machine learning model into existing healthcare systems.
- To uncover additional hidden insights or knowledge within the urine test dataset.

## Urine Analysis in Kidney Stone Detection

Urine analysis has been a cornerstone in the clinical evaluation of kidney stones. It provides crucial information on the urine's chemical composition, helping identify factors contributing to stone formation. Parameters typically measured include pH, specific gravity, and concentrations of calcium, oxalate, uric acid, citrate, and creatinine, among others (Rodgers et al., 2017). These measurements help identify individuals at risk of developing kidney stones and inform preventive and therapeutic strategies. Recent advancements in urine analysis techniques, such as liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectroscopy, have enhanced the accuracy and sensitivity in detecting urinary metabolites associated with kidney stones. However, these advanced methods generate complex datasets that require sophisticated analytical tools capable of managing large data volumes and uncovering subtle patterns that might elude traditional statistical methods.

## Machine Learning in Medical Diagnostics

Machine learning (ML), a branch of artificial intelligence (AI), focuses on developing algorithms that learn from data to make predictions. Unlike traditional programming, where explicit instructions are provided to the computer, ML algorithms enhance their performance as they process more data. This ability makes ML particularly suitable for medical diagnostics, where variable relationships can be intricate and non-linear (Esteva et al., 2019). In kidney stone detection, ML can analyze urine test data, identifying patterns and

combinations of urinary parameters indicative of stone formation. By training ML models on extensive datasets of urine analysis results paired with diagnostic outcomes, these models can learn to predict the presence of kidney stones with high accuracy.

## LITERATURE REVIEW

### Urine Analysis and Kidney Stones

Urine analysis is a crucial diagnostic tool that provides insights into the biochemical environment conducive to kidney stone formation. Key parameters include pH, specific gravity, osmolality, conductivity, urea, and calcium concentrations. These metrics can help determine an individual's risk of developing kidney stones and inform clinical decisions regarding preventive and therapeutic strategies. Recent technological advancements, such as liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectroscopy, have enhanced the precision of detecting urinary metabolites. However, the complexity of these data sets requires advanced analytical methods capable of managing large volumes of data and identifying complex patterns.

### Recent Studies and Findings

Lin et al. (2020) used support vector machines (SVM) to predict kidney stones based on urine metabolomics data, achieving an accuracy of 88%. Chen et al. (2019) applied deep learning techniques to urine microscopy images, resulting in an accuracy of 92%. Black et al. (2020) created a deep learning algorithm utilizing ResNet-101 to identify kidney stone composition from images, achieving an accuracy of 85.71% which underscores the potential of deep learning in the medical image analysis for detecting kidney stones. These studies highlight the efficacy of ML models in improving diagnostic accuracy.

Serrat et al. (2017) developed the myStone system, which utilizes Random Forest classifier for automatic kidney stone classification from images, achieving an accuracy of 63%. Esteva et al. (2019) demonstrated the utility of gradient boosting (GB) models in clinical urine tests, achieving an accuracy of 92%. Despite the complexity and interpretability challenges of GB models, they outperform simpler models in terms of accuracy. Bédard et al. (2020) explored logistic regression, finding it less accurate compared to more sophisticated models like random forest (RF) and GB.

Analyzing feature importance within these models reveals that osmolality and urea are critical predictors of kidney stones. Specific gravity and calcium concentration are also significant, while pH and conductivity, though less influential, contribute to the model's overall performance (Rodgers & Webber, 2017).

Wu et al. (2018) compared the performance of SVM, logistic regression, and RF models. Kourou et al. (2018) compared decision trees, k-nearest neighbors (KNN), and naïve Bayes classifiers, highlighting decision trees for their interpretability, though RF and GB outperformed in accuracy. Kazemi & Mirroshandel (2018) compared some classifiers to an ensemble learning approach to predict kidney stone types from textual data, achieving a high accuracy of 97.10%.

### Integrating Machine Learning into Clinical Practice and Real-World Applications

Implementing ML models in clinical settings presents challenges, primarily related to model interpretability and data security. Techniques like SHAP values and LIME can enhance model transparency, making them more acceptable for clinical use (Chen et al., 2019). Data privacy and security are critical, necessitating compliance with regulations such as HIPAA and GDPR. Robust encryption, secure storage, and strict access controls are essential to safeguard patient data (Esteva et al., 2019).

Oba et al. (2021) investigated ML models in resource-limited settings, emphasizing their potential to offer accessible and cost-effective diagnostic tools. While advanced models like RF and GB provide high accuracy, simpler models like logistic regression are easier to implement in low-resource environments.

## Challenges and Future Directions

While promising, the application of ML for kidney stone detection faces several challenges. Large, high-quality datasets are crucial for training accurate models. Existing studies often suffer from small sample sizes, limiting generalizability. Future research should focus on expanding datasets and including diverse patient populations and urine parameters. Integrating ML models into clinical workflows requires collaboration among data scientists, clinicians, and regulatory bodies to ensure safety, effectiveness, and user-friendliness. Training healthcare providers to use and interpret these models is also critical. Improving model interpretability remains a significant challenge. Transparent models that offer clear, actionable insights are essential for gaining clinicians' trust and facilitating adoption.

## Conclusion

The application of machine learning to urine test analysis for kidney stone detection has the potential to revolutionize medical diagnostics. Recent studies demonstrate the effectiveness of various ML models, including Random Forest (RF), Gradient Boosting (GB), Deep Learning models, and Support Vector Machines (SVM), in accurately predicting kidney stones. Identifying key urine parameters, such as osmolality, urea, specific gravity, and calcium concentration, aligns with clinical knowledge.

Despite promising results, challenges like dataset size, model interpretability, and clinical integration need addressing. Future research should focus on expanding datasets, enhancing model transparency, and validating models in clinical settings to ensure their practical applicability and improve patient outcomes.

Table I A Summary Review Of Related Works

#	Year	Author	Title	Data	Classifier	Accuracy
1	2020	Black et al.	Deep learning computer vision algorithm for detecting kidney stone composition	Images	ResNet-101	85.71%
2	2021	Williams et al.	Urine and stone analysis for the investigation of the renal stone former: a consensus conference	N/A	N/A	N/A
3	2009	Rule et al.	Kidney Stones and the Risk for Chronic Kidney Disease	Text	N/A	N/A
4	2017	Serrat et al.	myStone: A system for automatic kidney stone classification	Images	RF	63.00%
5	2021	Mao et al.	Relationship between urine specific gravity and the prevalence rate of kidney stone	Text	N/A	N/A

6	2018	Kazemi & Mirroshandel	A novel method for predicting kidney stone type using ensemble learning	Text	Ensemble	97.10%
7	2018	Wu et al.	Machine learning approach for detecting urinary stone disease in microscopic images	Images	Support Vector Machine (SVM)	89%
8	2020	Lin et al.	Urine metabolomics analysis for early detection of kidney stones using support vector machines	Text	Support Vector Machine (SVM)	88%
9	2019	Chen et al.	Detecting kidney stones in urine microscopy images using convolutional neural networks	Images	Convolutional Neural Network (CNN)	92%
10	2017	Rodgers & Webber	The role of urine analysis in the management of urolithiasis	Text	K-Nearest Neighbours (KNN)	84%

## RESEARCH METHODOLOGY

The classification task is used to predict future instances based on historical data. Previous research has seen experts applying various data mining techniques, such as clustering and classification, to accurately diagnose kidney stones and kidney diseases. In this study, the researcher employs several machine learning algorithms (classifiers) to detect kidney stones, including Random Forest (RF), Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, Gaussian Naïve Bayes, Support Vector Machine (SVM), Multi-Layer Perceptron (ANN), and Gradient Boosting.

### Data Source

Secondary data was utilized to carry out this research work. The dataset used was downloaded from Kaggle (uploaded by Vuppala Adithya Sairam, Kaggle Datasets Expert). This involves an authentic dataset comprising 79 data instances or observations, encompassing 7 distinct features (6 predictive features and 1 class). Gravity, ph, osmo, cond, urea, calc, and target. (Fig 1).

	gravity	ph	osmo	cond	urea	calc	target
0	1.021	4.91	725	14.0	443	2.45	0
1	1.017	5.74	577	20.0	296	4.49	0
2	1.008	7.20	321	14.9	101	2.36	0
3	1.011	5.51	408	12.6	224	2.15	0
4	1.005	6.52	187	7.5	91	1.16	0

Fig. 1. Attributes and details of the dataset

## Feature Description

With the exception of the target variable (target, which is categorical), all the remaining features are numeric in nature. The table below (Table II) depicts a short description of the various features in the dataset.

Table II A Short Description Of The Features In The Dataset

#	Feature	Description
1	gravity	Specific gravity of the urine sample
2	ph	pH level of the urine sample
3	osmo	Osmolality of the urine sample
4	cond	Conductivity of the urine sample
5	urea	Urea concentration in the urine sample
6	calc	Calcium concentration in the urine sample
7	target	A binary target variable indicating the presence (1) or absence (0) of kidney stones

## Process Model (Working Process)

The dataset was loaded and pre-processed, followed by an analysis to uncover hidden patterns and insights. It was then divided into two sets: training and testing, with a ratio of 4:1. Eighty percent (80%) for training and twenty percent (20%) for testing. The training set was used for the training of the various models (classifiers) while the testing set was used to test or validate the various models. The models were evaluated using several metrics so that the best one could be chosen. The diagram below (Fig. 2) depicts the process flow of the proposed model.

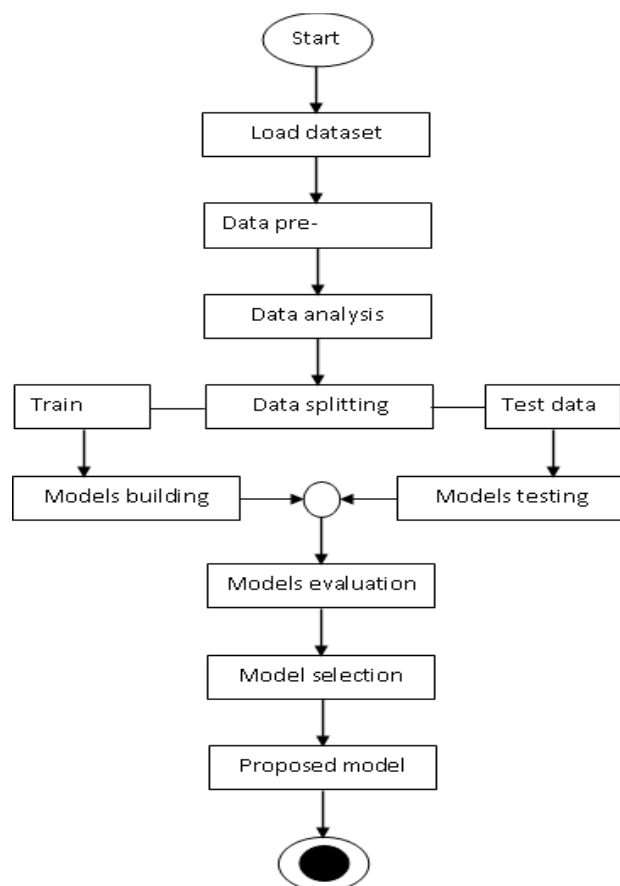


Fig. 2. Flow chart (Process flow) of the proposed model



## Data Pre-processing

Real-world data often isn't in a format ideal for machine learning applications. It may contain noise and missing values. To address these issues and generate accurate predictions, the data must be processed thoroughly. Consequently, the dataset underwent extensive pre-processing. This included activities such as data cleaning, transformation, normalization, and handling imbalanced data, among other techniques.

**Data cleaning** generally involves identifying and addressing noise, fake data, duplicate entries, and missing values. To ensure accurate and useful results, it is essential to remove noise and fill in the missing values. Fortunately, this dataset did not have any missing values, duplicate entries, and fake data.

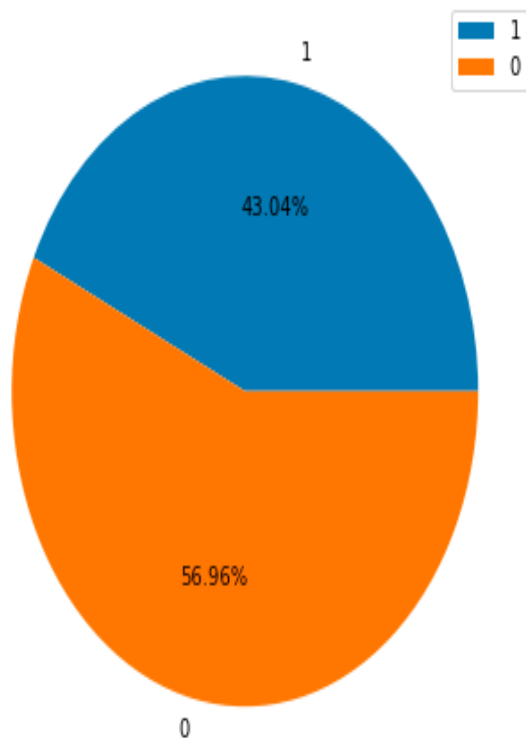
**Transformation** involves converting data from one format to another to enhance its comprehensibility. This process includes tasks such as aggregation, data type casting, encoding, and smoothing. All numeric and categorical features or variables are supposed to be converted to their appropriate data types and formats.

**Scaling** involves modifying the range of feature values to a standard scale without altering the differences in their ranges. This process ensures that each feature has an equal contribution to the model, thereby enhancing the performance and accuracy of machine learning algorithms. Standardization method (Z-score) was used to scale all the features.

**Dimensionality Reduction (Data Reduction)** involves removing unwanted or less relevant features. Here no variable was removed.

**Handling Imbalance Data** entails adjusting the data distribution to prevent biases during analysis and modelling. The data was fairly biased with 34 observations being patients with kidney stones and 45 without kidney stones. Therefore, there was the need to balance the dataset in order to prevent the biasness. Over sampling method was used on the patients with kidney stones so as to increase the observation in order to match up with those without kidney stones. The diagrams below depict the dataset before and after handling the imbalance data.

Pie Chart of Kidney Stones vs No Kidney Stones



Pie Chart of Kidney Stones vs No Kidney Stones

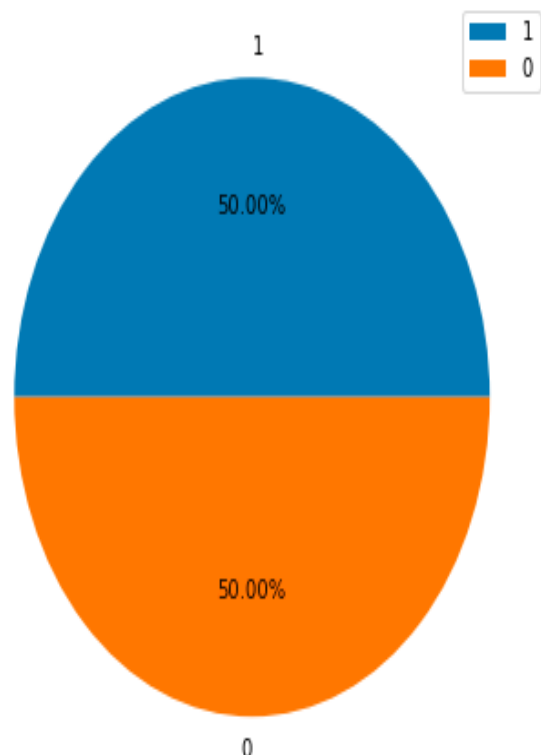


Fig. 3. Dataset before and after balancing

## Metrics for Evaluating Machine Learning Models

**Confusion Matrix:** A Confusion Matrix is a simple yet effective method to evaluate the performance of a classification model. It achieves this by comparing the number of positive and negative instances that were correctly or incorrectly classified (Osei, I., & Adomako, A. B., 2024)

Table III Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

### True Positives (TP):

True positives are instances where both the predicted class and the actual class are positive (true).

### True Negatives (TN):

True negatives are instances where both the predicted class and the actual class are negative (false).

### False Negatives (FN):

False negatives are instances where the predicted class is negative (0), but the actual class is positive (1).

### False Positives (FP):

False positives are instances where the predicted class is positive (1), but the actual class is negative (0).

From the confusion matrix, metrics such as accuracy, precision, recall, and F1-score can be calculated using the following formulas.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.1)$$

Where  $TP$  = True Positive,  $TN$  = True Negative,

$FP$  = False Positive, and  $FN$  = False Negative

$$Precision = \frac{TP}{TP+FP} \quad (3.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3.3)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (3.4)$$

**Area under Curve:** The Area under Curve (AUC) is a valuable metric with values ranging from 0 to 1. The closer the AUC is to 1, the better the machine learning model is at distinguishing between kidney stone cases and non-kidney stone cases. A model that completely differentiates between the two classes has an AUC of 1. Conversely, if all non-kidney stone instances are incorrectly classified as kidney stones and vice versa, the AUC is 0 (Osei, I., & Adomako, A. B., 2024).

## Deployment of the Proposed Model

With the help of Flask framework, HTML, and CSS, the model was deployed in a web based which can easily be integrated into existing healthcare systems. The figures below show the respective interfaces.



### Kidney Stone Detection

Fill this form

Gravity:	<input type="text" value="Gravity"/>
pH:	<input type="text" value="pH"/>
Osmolality:	<input type="text" value="osmo"/>
Conductivity:	<input type="text" value="Conductivity"/>
Urea concentration:	<input type="text" value="Urea concentration"/>
Calcium concentration:	<input type="text" value="Calcium concentration"/>

Fig. 4. Homepage of the web application

### Kidney Stone Detection

Fill this form

Gravity:	<input type="text" value="1.0"/>
pH:	<input type="text" value="5.2"/>
Osmolality:	<input type="text" value="725"/>
Conductivity:	<input type="text" value="14"/>
Urea concentration:	<input type="text" value="443"/>
Calcium concentration:	<input type="text" value="2.45"/>

Fig. 5. Prediction phase

### Kidney Stone Detection

No Kidney Stones Detected

Fill this form

Gravity:	<input type="text" value="Gravity"/>
pH:	<input type="text" value="pH"/>
Osmolality:	<input type="text" value="osmo"/>
Conductivity:	<input type="text" value="Conductivity"/>
Urea concentration:	<input type="text" value="Urea concentration"/>
Calcium concentration:	<input type="text" value="Calcium concentration"/>

### Kidney Stone Detection

Kidney Stones Detected

Fill this form

Gravity:	<input type="text" value="Gravity"/>
pH:	<input type="text" value="pH"/>
Osmolality:	<input type="text" value="osmo"/>
Conductivity:	<input type="text" value="Conductivity"/>
Urea concentration:	<input type="text" value="Urea concentration"/>
Calcium concentration:	<input type="text" value="Calcium concentration"/>

Fig. 6. Results or output phase

## DATA ANALYSIS AND INTERPRETATION

### Exploratory Data Analysis (EDA)

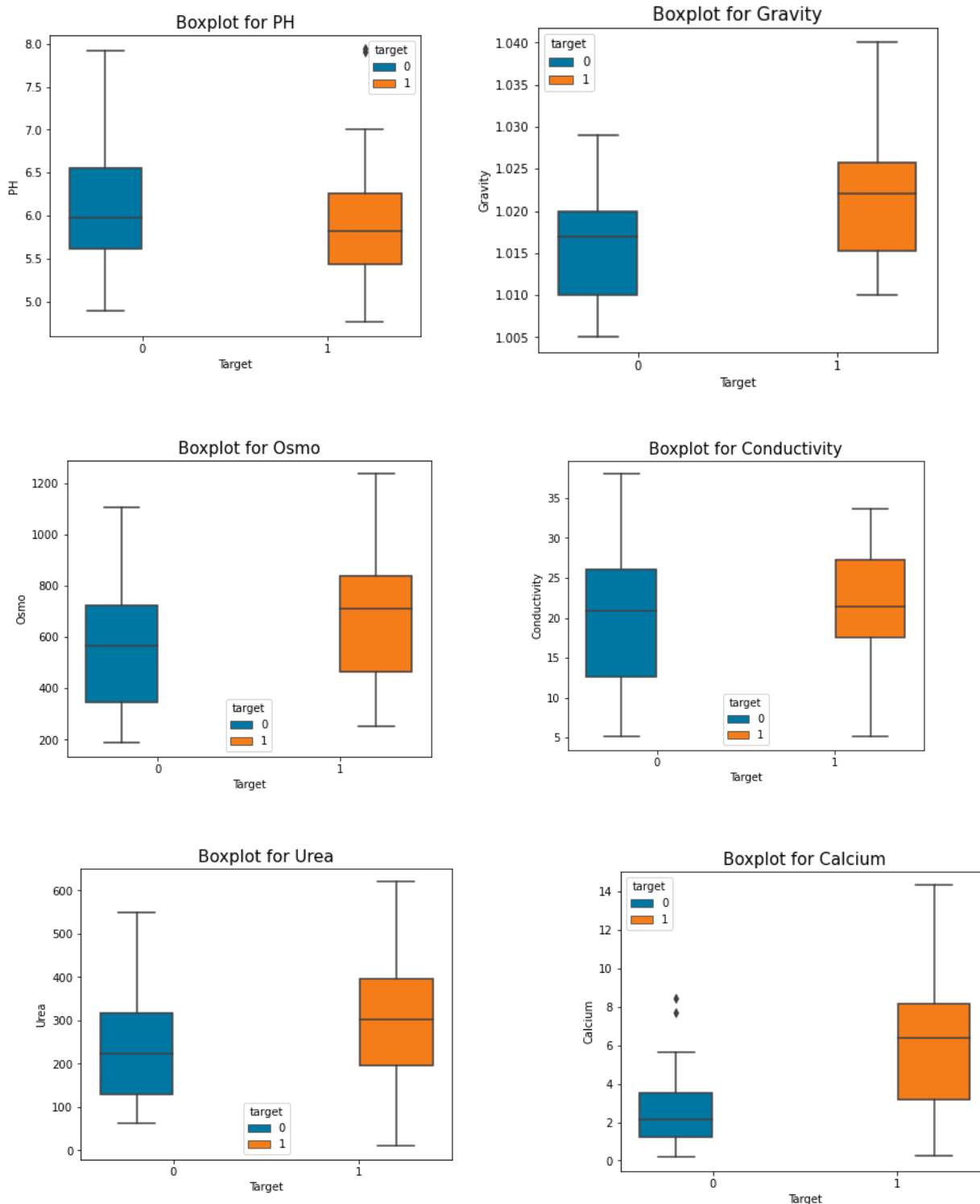


Fig. 7. Boxplots for continuous variables in the dataset

From the above box plots, the following insights were discovered;

1. Individuals with kidney stones tend to have higher specific gravity values in their urine. This is because the median urine specific gravity of those with kidney stones is higher than those without kidney stones. It can be concluded that individuals with high level of urine specific gravity are more prone to getting kidney stones.

2. Kidney stones were found in people with low value of urine pH (Acidic urine) though there was an outlier. Therefore, it can be concluded that people with acidic urine are more likely to be affected by kidney stones.
3. Individuals with kidney stones tend to have higher osmolality value in their urine. This is because the median value of osmolality is higher in those with kidney stones than those without kidney stones. As such, individuals with high level of osmolality are more likely to be affected by kidney stones.
4. There is no major difference in median value of conductivity between those with and without kidney stones. There is a small variation in urine conductivity between individuals with and without kidney stones, with those having kidney stones generally exhibiting higher conductivity levels.
5. There is a significant difference in urea concentration between individuals with and without kidney stones. Specifically, those with kidney stones tend to have higher and more variable urea levels in their urine. This suggests that individuals with high levels of urea concentration in their urine are more likely to have kidney stones.
6. There is a notable difference in calcium concentration between individuals with and without kidney stones. In particular, those with kidney stones typically display higher and more variable levels of calcium in their urine.

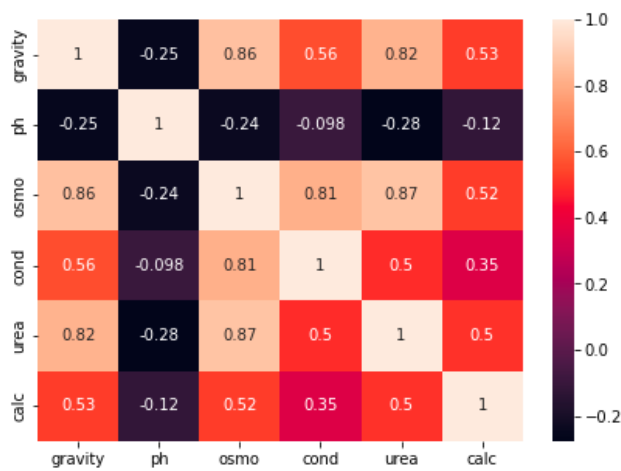


Fig. 8. Correlation coefficients between the continuous variables

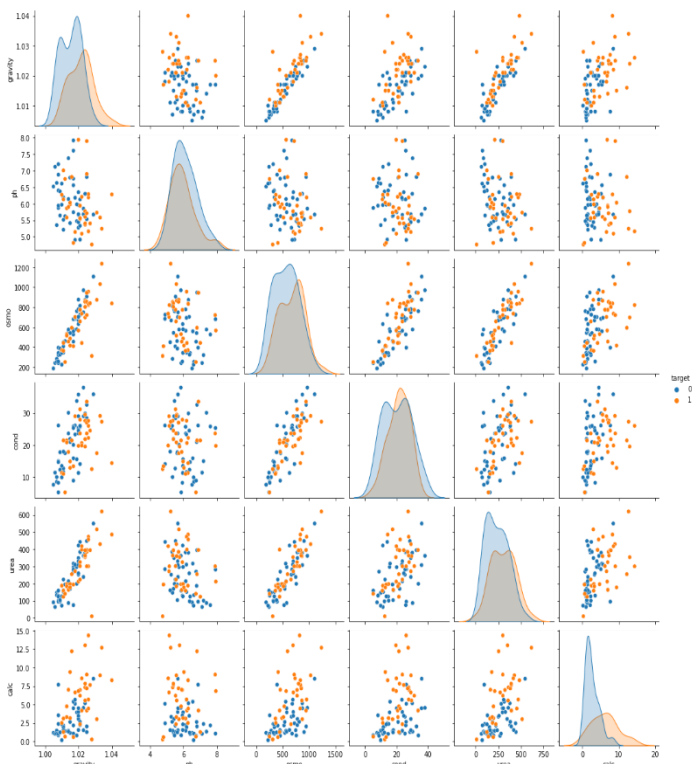


Fig. 8. Pair plots for continuous variables

- **Gravity and Osmo (0.86):** A strong positive correlation exists between gravity and osmolality, indicating that as the specific gravity of urine increases, osmolality also tends to increase.
- **Gravity and Urea (0.82):** Gravity shows a strong positive correlation with urea concentration, suggesting that higher specific gravity is associated with higher urea levels.
- **Osmo and Urea (0.87):** Osmolality and urea concentration have a strong positive correlation, meaning that higher osmolality values correspond to higher urea levels.
- **Osmo and Cond (0.81):** There is a strong positive correlation between osmolality and conductivity, indicating that higher osmolality values tend to coincide with higher conductivity.
- **Gravity and Cond (0.56):** A moderate positive correlation is present between gravity and conductivity.
- **Gravity and Calc (0.53):** There is a moderate positive correlation between gravity and calcium concentration.
- **Osmo and Calc (0.52):** Osmolality and calcium concentration show a moderate positive correlation.
- **Urea and Calc (0.5):** There is a moderate positive correlation between urea concentration and calcium levels.
- **pH with other parameters:** pH shows weak negative correlations with gravity (-0.25), osmolality (-0.24), and urea (-0.28). It also has very weak negative correlations with conductivity (-0.098) and calcium (-0.12).
- **Cond and Calc (0.35):** Conductivity and calcium concentration have a weaker positive correlation compared to other pairs.

### Confirmatory Data Analysis (CDA)

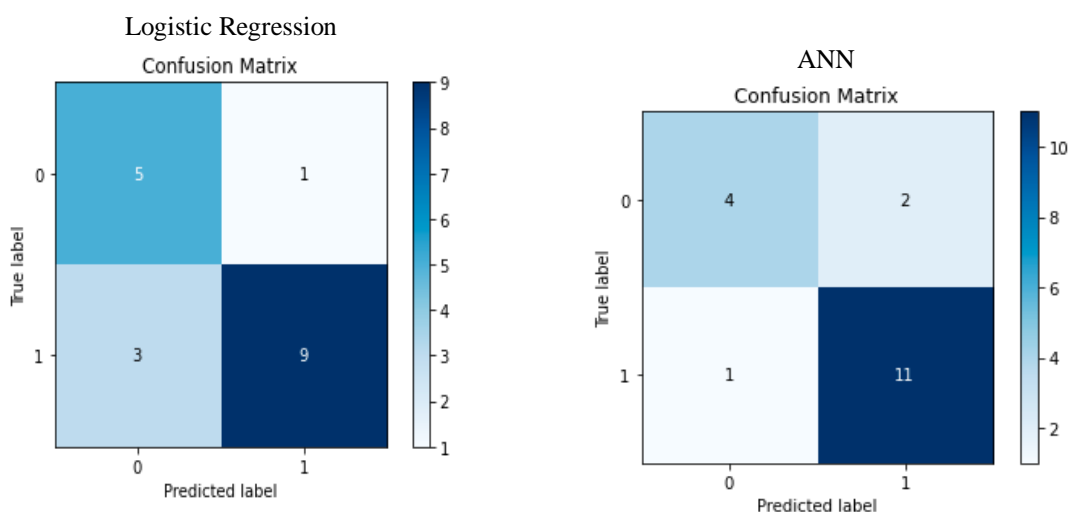
A parametric statistical test (Logistic regression) was used on the variables against the target to test for causality. The following are the deductions made;

1. Urine specific gravity value determines the availability of kidney stones.
2. pH value of urine does not determine the availability of kidney stones.
3. Osmolality level of urine determines the availability of kidney stones.
4. Conductivity level of urine does not determine the availability of kidney stones.
5. Urea concentration level of urine determines the availability of kidney stones.
6. Calcium concentration level of urine determines the availability of kidney stones.

## RESULTS ANALYSIS AND DISCUSSIONS

### Confusion Matrix

The figure below (Fig. 9) depicts the confusion matrices for the various classifiers



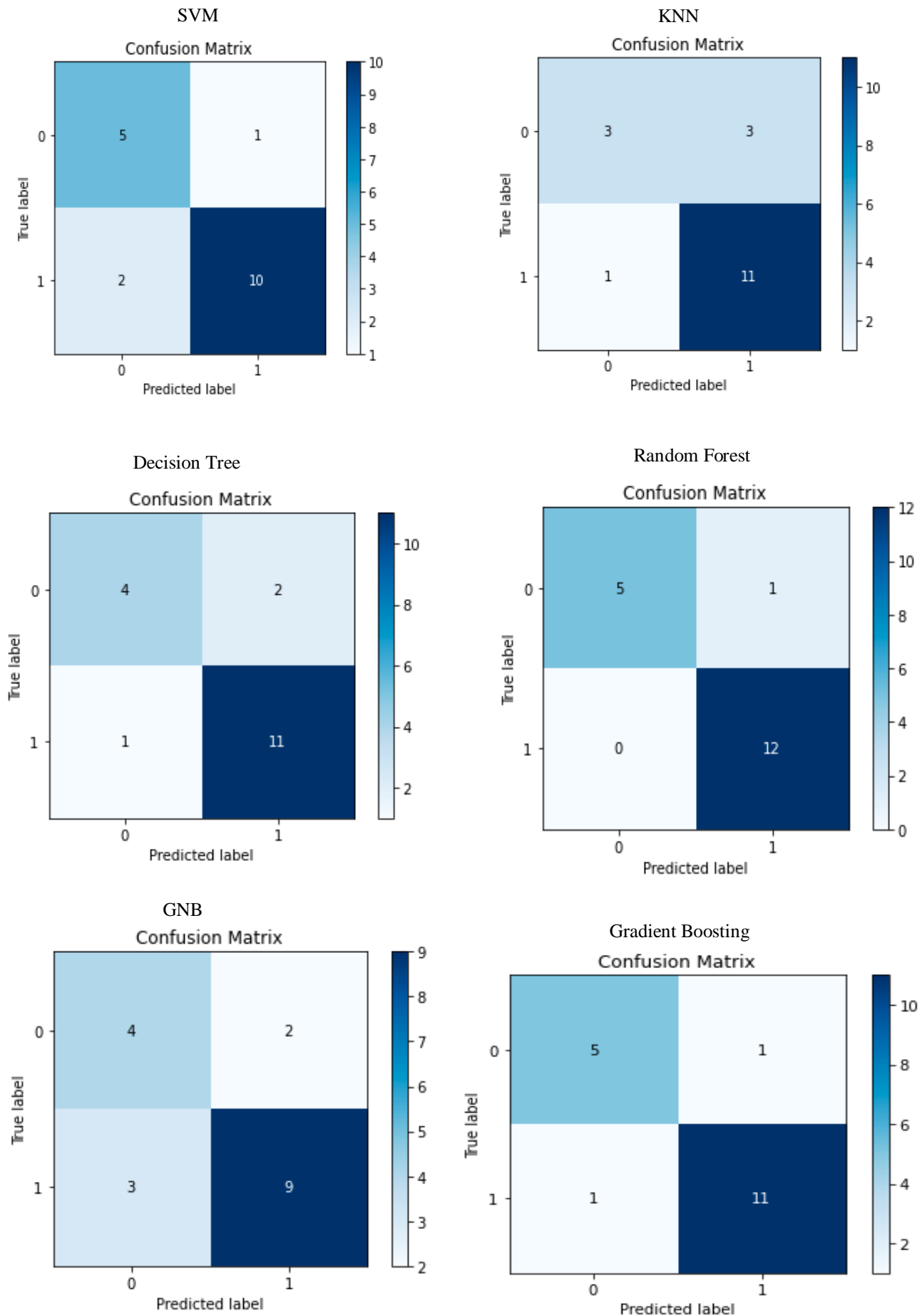


Fig. 9. Confusion Matrices for the classifiers

## Performance Comparison of Various Classifiers

The table below depicts a comparison of the different metrics used to evaluate the classifiers.

Table IV Comparison Of Classifiers Using Various Evaluation Metrics

#	Classifier	ACC	PRE	REC	F1	AUC
1	Logistic Regression (LR)	78%	81%	78%	78%	88%
2	ANN	83%	83%	83%	83%	91%
3	SVM	83%	84%	83%	84%	83%
4	KNN	78%	77%	78%	76%	76%
5	Decision Tree (DT)	83%	83%	83%	83%	83%
6	Random Forest (RF)	94%	95%	94%	94%	94%
7	Gaussian Naïve Bayes (GNB)	72%	74%	72%	73%	86%
8	Gradient Boosting (GB)	89%	89%	89%	89%	96%

ACC = Accuracy, PRE = Precision, REC = Recall, F1 = F1-Score, AUC = Area under Curve

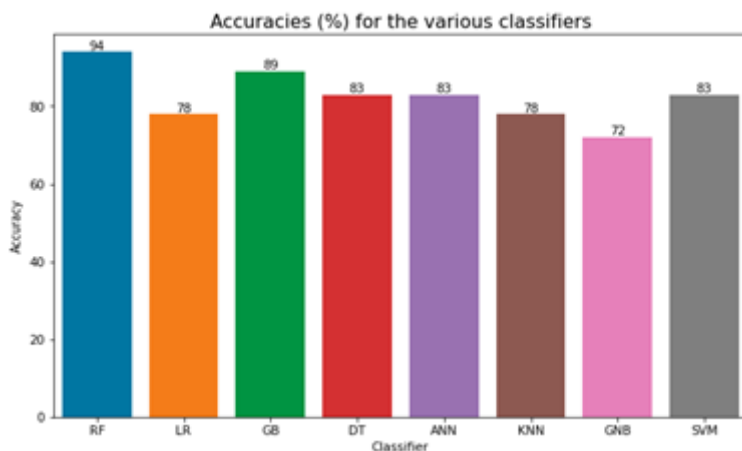


Fig. 10. Accuracies for the classifiers

## Results and Performance Analysis

The confusion matrices in Figure 9 (Fig. 9) highlighted the rates of false positives (FP) and false negatives (FN), which are crucial considerations for any model. A false positive may lead to unnecessary treatment, while a false negative, especially in cases of undetected kidney stones, could result in a severe misdiagnosis. The Random Forest classifier showed a low incidence of FP and FN, enhancing its reliability. The false positives indicate that some records of patients without kidney stones exhibit characteristics similar to those of patients with kidney stones, while the false negatives suggest that some kidney stones patients show non-kidney stones-like characteristics.

Table IV evaluates accuracy, precision, recall, F-1 score, and AUC for various classification methods, as defined in equations (3.1) to (3.4). The Random Forest (RF) model achieved a 94% accuracy rate, outperforming the other classifiers. Precision, the ratio of correctly predicted positive observations to the total



predicted positives, was highest for RF (0.95), indicating a lower false-positive rate. Recall, the measure of correctly predicted positive cases relative to all cases in the class, was also superior for RF (0.94).

The F-1 score, the harmonic mean of precision and recall, considers both false positives and negatives. Although it is not as straightforward as accuracy, the F-1 score is often more informative, especially with imbalanced class distributions. RF scored highest in this metric as well. The final metric, Area under Curve (AUC), evaluates the total area under the ROC curve, extending from (0, 0) to (1, 1). A higher score, closer to 1, signifies better performance. Here, RF excelled with a score of 0.94, although Gradient Boosting (GB) recorded a slightly higher score of 0.96.

Overall, the RF model outperformed all other classifiers in all metrics except AUC, suggesting that RF performed well on the dataset used for this research.

### Benchmarking

The table below (Table V) shows the accuracy of some related work as compared to this work.

Table V Results Comparison Of The Related Works

#	Ref	CU	TD	ACC
1	This Study	Random Forest	Urine analysis data	94%
2	Wu et al.	Support Vector Machine	Urine analysis data from clinical trials	89%
3	Lin et al.	Support Vector Machine	Urine metabolomics data	88%
4	Chen et al.	Convolutional Neural Network (CNN)	Urine microscopy images	92%
5	Esteva et al.	Gradient Boosting	Urine analysis data integrated with clinical metadata	92%
6	Rodgers and Webber	K-Nearest Neighbors (KNN)	Urine chemistry profiles	84%
7	Black et al.	ResNet-101	Images	86%

Ref = Reference, CU = Classifier used, TD = Type of data, ACC = Accuracy

The table above offers a comparative analysis of the accuracy of various machine learning models utilized for detecting kidney stones through urine test analysis across different types of datasets. The Random Forest model from the this study shows the highest accuracy at 94%, followed by Gradient Boosting and CNN models, both with accuracies of 92%. The range of dataset types and accuracy rates underscores the flexibility of machine learning approaches to different forms of urine analysis data and indicates potential for further enhancement in prediction accuracy.

### Practical Implications

The machine learning approach, especially the Random Forest model, presents several advantages over traditional diagnostic methods:

- **Non-Invasive:** Uses urine tests, which are more comfortable and less invasive than imaging techniques.

- **Cost-Effective:** Decreases the need for expensive diagnostic procedures like CT scans and ultrasounds.
- **Accessibility:** Can be applied in primary care settings, making it accessible to patients in resource-limited areas.

## CONCLUSION AND RECOMMENDATIONS

### Summary

The research demonstrated that machine learning, particularly the Random Forest classifier, can effectively detect kidney stones using urine test analysis. The high accuracy and reliability of the Random Forest model highlight its potential as a valuable diagnostic tool, offering a non-invasive, cost-effective, and accessible means of detecting kidney stones. By identifying key urine parameters, such as osmolality, urea, specific gravity, and calcium concentration, the study aligns with clinical knowledge and emphasizes the relevance of these features in kidney stone formation. This machine learning approach can significantly enhance early detection and patient outcomes, providing a promising alternative to traditional diagnostic methods.

Despite the promising results, certain limitations need to be addressed, including the necessity for larger and more diverse datasets and improved model interpretability. Future research should focus on expanding the dataset, developing hybrid models that combine machine learning with traditional diagnostic methods, and conducting clinical trials to validate model performance in real-world settings.

In conclusion, integrating machine learning models into clinical practice represents a significant advancement in leveraging data-driven approaches to enhance healthcare outcomes. By addressing current limitations and focusing on practical implementation, this method could substantially improve kidney stone management and patient care.

### Challenges and problems encountered

The following challenges and problems were encountered during the research work:

- Inaccessibility of local (Ghanaian) medical dataset for such research
- Identifying the most relevant features from urine test parameters was critical yet challenging
- Choosing the appropriate machine learning algorithms was a significant challenge.
- Integrating machine learning models into clinical practice requires building trust among healthcare providers.
- Choice of appropriate dataset for this research
- Implementing these models in practice also required training healthcare providers on how to use the new tools effectively, which involves additional time and resources.
- External knowledge from health workers was necessary to fully understand some features in the dataset.
- The appropriate hyper parameter to fine-tune to achieve higher accuracy

### Recommendations

- Future research should strive to incorporate a larger and more varied dataset. An increased sample size would enhance the model's generalizability and yield more reliable predictions.
- Efforts should focus on gathering an equal number of samples from patients with and without kidney stones. This will aid in training models that are not biased towards a particular class, thereby improving the accuracy of predictions.
- Incorporating more comprehensive urine analysis parameters and additional relevant clinical features can boost the model's predictive power. Integrating patient history, dietary habits, and genetic information could offer a more holistic perspective.
- Utilizing advanced feature engineering techniques to identify meaningful patterns in the data. Methods such as principal component analysis (PCA) and feature selection algorithms can enhance model performance.

- Exploring ensemble methods that combine multiple models to enhance prediction accuracy and robustness. Techniques such as stacking, boosting, and bagging can be considered.
- Concentrating on the development of models that deliver high accuracy while also being interpretable. Techniques like SHAP (SHapley Additive exPlanations) values or LIME (Local Interpretable Model-agnostic Explanations) can aid in elucidating model predictions.
- Carrying out clinical trials and validations to assess the model's performance in real-world settings. Partnering with healthcare institutions to collect feedback and refine the model as needed.
- Creating comprehensive training programs for healthcare providers to ensure they can effectively use and interpret the machine learning models. This will facilitate the smooth integration of these tools into clinical practice.
- A thorough comparison of classification algorithms including Deep Learning, Transfer Learning, etc. should be included.
- Findings of this research work should be given the necessary attention and care.

### Limitations

- The study was constrained by a dataset of only 79 samples, which may not adequately represent the variability and complexity of the population. This limitation affects the generalizability of the findings.
- The study utilized a limited range of urine analysis parameters. Incorporating a broader set of features could enhance the accuracy and robustness of the model.
- With the small dataset, there is a risk of overfitting, where the model performs well on training data but poorly on unseen data, limiting its practical applicability.
- The study did not include clinical validation, meaning the model's performance has not been tested in real-world healthcare settings, which limits the immediate applicability of the findings.
- While the Random Forest model achieved high accuracy, its complexity reduced interpretability. This could hinder clinical adoption as healthcare providers may find it challenging to trust and understand the model's predictions.
- Training and deploying complex machine learning models require significant computational resources, which may not be readily available in all clinical settings.
- Integrating machine learning models into existing healthcare systems poses challenges, including data interoperability and the need for specialized training for healthcare providers.
- The variables (features) in the dataset were not explicitly explained.
- Secondary data was used for this research instead of primary data.

### Conclusion

This study demonstrates the potential of using machine learning models to detect kidney stones through urine test analysis. However, addressing the outlined limitations and following the recommendations is crucial for advancing this research and improving patient care. Future work should focus on expanding the dataset, enhancing model interpretability, and conducting clinical validations to ensure the practical applicability of these models in healthcare settings.

### REFERENCES

1. Azam, M. S., Habibullah, M., & Rana, H. K. (2020). Performance Analysis of Machine Learning Approaches in Stroke Prediction. *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*, 175(21), 1464–1469. <https://doi.org/10.1109/ICECA49313.2020.9297525>
2. Black, K. M., Law, H., Aldoukhi, A., Deng, J., & Ghani, K. R. (2020). Deep learning computer vision algorithm for detecting kidney stone composition. *BJU International*, 125(6), 920–924. <https://doi.org/10.1111/bju.15035>
3. Dritsas, E., & Trigka, M. (2022). Stroke Risk Prediction with Machine Learning Techniques. *Sensors*, 22(13). <https://doi.org/10.3390/s22134670>

4. Kazemi, Y., & Mirroshandel, S. A. (2018). A novel method for predicting kidney stone type using ensemble learning. *Artificial Intelligence in Medicine*, 84, 117–126. <https://doi.org/10.1016/j.artmed.2017.12.001>
5. Mao, W., Zhang, H., Xu, Z., Geng, J., Zhang, Z., Wu, J., ... Chen, M. (2021). Relationship between urine specific gravity and the prevalence rate of kidney stone. *Translational Andrology and Urology*, 10(1), 184–194. <https://doi.org/10.21037/TAU-20-929>
6. Osei, I., & Adomako, A. B. (2024). Using Machine Learning to Predict Heart Failure: A Comparative Analysis of Various Classification Algorithms. *International Journal of Research and Scientific Innovation*, XI(1), 336–354. <https://doi.org/10.51244/ijrsi.2024.1101026>
7. Rule, A. D., Bergstralh, E. J., Melton, L. J., Li, X., Weaver, A. L., & Lieske, J. C. (2009). Kidney stones and the risk for chronic kidney disease. *Clinical Journal of the American Society of Nephrology*, 4(4), 804–811. <https://doi.org/10.2215/CJN.05811108>
8. Serrat, J., Lumbreras, F., Blanco, F., Valiente, M., & López-Mesas, M. (2017). myStone: A system for automatic kidney stone classification. *Expert Systems with Applications*, 89, 41–51. <https://doi.org/10.1016/j.eswa.2017.07.024>
9. Williams, J. C., Gambaro, G., Rodgers, A., Asplin, J., Bonny, O., Costa-Bauzá, A., ... Robertson, W. G. (2021). Urine and stone analysis for the investigation of the renal stone former: a consensus conference. *Urolithiasis*, 49(1), 1–16. <https://doi.org/10.1007/s00240-020-01217-3>
10. Chen, M., Zhang, B., Zhang, Y., & Yang, J. (2019). Detecting kidney stones in urine microscopy images using convolutional neural networks. *Biomedical Signal Processing and Control*, 47, 212–219.
11. Chaiyarit, N., Thongboonkerd, V., & Rodprasert, W. (2018). Urine proteomics for the diagnosis of kidney stones. *Proteomics Clinical Applications*, 12(1), e1700122.
12. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
13. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29.
14. Fulgham, P., Assimos, D., Pearle, M., & Preminger, G. (2013). Clinical effectiveness protocols for imaging in the management of ureteral calculous disease: AUA technology assessment. *The Journal of Urology*, 189(4), 1203–1213.
15. Lin, H., He, Y., Li, K., & Liu, Y. (2020). Urine metabolomics analysis for early detection of kidney stones using support vector machines. *Clinical Biochemistry*, 73, 46–53.
16. Pearle, M. S., Goldfarb, D. S., Assimos, D. G., Curhan, G., Denu-Ciocca, C. J., Matlaga, B. R., ... & White, J. R. (2014). Medical management of kidney stones: AUA guideline. *The Journal of Urology*, 192(2), 316–324.
17. Rodgers, A. L., & Webber, D. (2017). The role of urine analysis in the management of urolithiasis. *BJU International*, 119(5), 720–725.
18. Romero, V., Akpınar, H., & Assimos, D. G. (2010). Kidney stones: a global picture of prevalence, incidence, and associated risk factors. *Reviews in Urology*, 12(2–3), e86.
19. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
20. Wu, J., & Yang, S. (2018). Machine learning approach for detecting urinary stone disease in microscopic images. *Journal of Medical Systems*, 42(5), 85.
21. vuppalaadithyasairam. (September 2021). Kidney Stone Prediction based on Urine Analysis. Retrieved [May 12, 2024] from <https://www.kaggle.com/datasets/vuppalaadithyasairam/kidney-stone-prediction-based-on-urine-analysis/data>.