

# PREDICTIVE URINALYSIS USING AI AND MACHINE LEARNING

Rabia Alam<sup>\*1</sup>, Sohail Masood Bhatti<sup>1</sup>, Ali Ahmad Sheikh<sup>2</sup>, Sundus Sufyan<sup>2</sup>

<sup>1</sup> Department of Computer Science and IT, Superior University, Lahore

<sup>2</sup> Institute of Microbiology, University of Veterinary and Animal Sciences, Lahore

\* Corresponding Author: rabipkmsn@gmail.com

## ABSTRACT

Usage of artificial intelligence in healthcare system, in recent years, to interpret the data and to predict the disease has increased extensively. AI tools use binary classification system to interpret the data either as “Positive” or “Negative”. Such interpretation when used in urinalysis has limited clinical benefits. The current study is designed to improve the machine learning (ML) models such as Support Vector Machines, Artificial Neural Networks and Random Forest and to evaluate their output to differentially diagnose urinalysis data on 20 different parameters. In current study, urinalysis reports of 800 patients were collected. The record was anonymized to hide identity of the patients. Out of these 800 test reports, 640 (80%) reports were used to train the models while 160 (20%) reports were used to evaluate the efficiency of the trained models. The dataset was pre-processed for feature scaling and removing any missing values. To determine the performance assessment of the AI driven models, diverse array of standardized metrics was applied. The outcomes of this study showed that AI tools have potential for rapid, accurate diagnosis based of urinalysis and can be further use as non-invasive diagnostic tool. It is pertinent to mention that all previously conducted research work were on single parameters of urinary tract related diseases, while the current study is capable of predicting six diseases based on differential diagnosis with 93.29% accuracy which is much better than previously conducted research work. This advancement could lead to improved patient’s disease diagnosis and will facilitate clinical workflows. It is anticipated that future research will focus on expanding the dataset and investigating deeper learning techniques.

**Key Words:** Urinary Tract Infection, Predictive Urinalysis, Differential Diagnosis, General kidney diseases

## Introduction

Urinalysis is one of the basic, non-invasive and relatively cheap diagnostic tools which is extensively used in healthcare settings for the diagnosis and monitoring of urinary tract. The test can monitor and predict urinary tract related issues including urinary tract infection (UTI), diabetes and other kidney disorders (Advani *et al.*, 2021). Urinalysis also facilitates in routine health screenings, preparation of patient for surgical procedures, hospitalization and prenatal care (Mahoney *et al.*, 2020). The procedure comprises visual inspection, chemical analysis (pH, protein, glucose, ketones, etc.), and microscopic evaluation (cells, casts, crystals, microorganisms), offering valuable insights into physiological and pathological states (Delanghe and Speeckaert, 2016; Chambliss *et al.*, 2020).

The urinalysis parameters included indicators such as color, specific gravity, pH, leucocytes, esterase, nitrite, protein, sugar, ketones, urobilinogen, bilirubin, haemoglobin, pus cells, RBCs, epithelial cells, amorphous, casts, organisms, yeast cells and crystals. Table – 1 depict that each of these parameters have a unique diagnostic value. Routine urinalysis parameters provide rapid, non-invasive diagnostic information across renal, hepatic, and urinary disorders. These indicators are increasingly integrated into machine learning models to enhance clinical decision-making (Panchbhai, 2022; Li *et al.*, 2023; Bhattacharya *et al.*, 2024).

**Table – 1: Diagnostic Significance of Routine Urinalysis Parameters in Detecting Renal, Metabolic, and Infectious Disorders**

| Sr. No. | Parameter        | Diagnostic Features                                                                               |
|---------|------------------|---------------------------------------------------------------------------------------------------|
| 1.      | Color            | Reflects hydration, medication intake, or pathological conditions (e.g., hematuria, proteinuria). |
| 2.      | Specific Gravity | Indicates urine concentration, with deviations hinting at dehydration or renal dysfunction.       |
| 3.      | pH               | Variations can signify infections, stone formation, or metabolic disorders.                       |

|     |                                 |                                                                                          |
|-----|---------------------------------|------------------------------------------------------------------------------------------|
| 4.  | Leukocytes and Esterase         | Levels are classic markers for urinary tract infections (UTIs).                          |
| 5.  | Nitrite                         | Positivity confirms the presence of nitrate-reducing bacteria in UTIs.                   |
| 6.  | Proteinuria                     | Often signals kidney damage, while transient elevation may stem from stress or exertion. |
| 7.  | Glucosuria                      | Denotes hyperglycemia or diabetes mellitus.                                              |
| 8.  | Ketones                         | Suggest fat metabolism due to uncontrolled diabetes, fasting, or ketogenic diets.        |
| 9.  | Urobilinogen and Bilirubin      | Help assess liver function and red blood cell turnover.                                  |
| 10. | Hemoglobin, RBCs, and pus cells | Provide insights into infections, trauma, or malignancies.                               |
| 11. | Epithelial cells                | Help evaluate contamination or pathology.                                                |
| 12. | Casts and crystals              | Reveal kidney-specific abnormalities or lithiasis.                                       |
| 13. | Microorganisms and yeast cells  | Indicate bacterial or fungal infections.                                                 |

Urinalysis is quite a simple technique of identifying urinary tract issues as well as some metabolic disorders. It is significant that through this technique various systematic disorders can be identified with more accuracy. Being a very simple traditional tool, it has many limitations and challenges including microscopic and visual examination of some of the parameters i.e., cysts, casts, bacterial cells, pus cells, etc., where the operator's personal experience, knowledge and skills are required. After visual and microscopic examination, a chemical-coated strip is dipped into the urine which gives a quick basic idea about the disease, followed by more detailed culturing and imaging as part of the urinalysis test. As the process is multistep, it slows down the process of performing the test and ultimately the interpretation of the results (Han, 2013; Kavuru et al., 2020; Memişoğlu et al., 2010). There is a need for a more robust and better diagnostic procedure that could provide more reliability, accuracy, efficiency and consistency to the physicians so they may make a decision quickly.

Dipstick testing is widely used in routine diagnostics, but several limitations still prevent timely and accurate identification of urinary tract infections (UTIs). While the detection of nitrite positivity is widely interpreted as detection of bacteriuria, it cannot point out the infections that are caused by the bacteria which do not produce nitrites. Also, sterile pyuria may be observed in non-infectious situations. Diagnostic gaps such as these necessitate the development of integrative models that are able to adapt to and embrace the variation of human biology.

Most prominently, new technologies are used to develop the models in which artificial intelligence and machine learning (AI/ML) are employed. Computational models are effective in the analysis of large amounts of data with variable dimensions for the identification of hidden relationships and the formulation of hypotheses regarding the behavior of studied phenomena. However, the cryptic nature of such technologies limits their use in medicine. One of the examples of such a gray model is Deep Learning, especially a family of models known as Artificial Neural Networks. In contrast, Random Forests, as an example of ensemble models, score the features which are used for making a decision, and this property is highly valued in medicine.

The intersection of performance and interpretability has sparked further exploration of Explainable AI (XAI). Complex interrelations of multiple parameters in fields such as urinalysis have led to growing interest in XAI. XAI constructions are designed to deconstruct 'black box' algorithm outputs. This helps clinicians appreciate and, therefore, trust computational outputs (Elham et al., 2024).

The potential of AI in urinary diagnostics has recently been the subject of considerable work. For instance, XGBoost and Random Forest achieved excellent UTI outcome prediction with area-under-the-curve (AUC) indexes reaching 0.819 and 0.791, respectively (Sergounioti et al., 2025). Beyond UTIs, AI has successfully diagnosed chronic kidney disease (CKD) and urinary tract cancers (Tan et al., 2021; Yuan et al., 2020; Phoung et al., 2025).

The current research work is based upon these foundations and aims to extend the use of AI to non-binary classification. The purpose of the research is to enhance differential diagnostics for a broader range of urological diseases using multiclass models trained on standard urinalysis parameters (Sergounioti et al., 2025).

To make such prediction, models need to distinguish subtle difference between overlapping conditions (e.g., UTI vs. kidney stones) (Lu et al., 2024). AI models can evaluate the effectiveness of SVMs, Random Forests, and ANNs based on accuracy, feature discrimination, and interpretability (Guido, 2024; Del Ben et al., 2023). Support Vector Machines (SVM) are effective for high dimensional data but sensitive to parameter tuning and data imbalance (Guido, 2024; Dawuti et al., 2022).

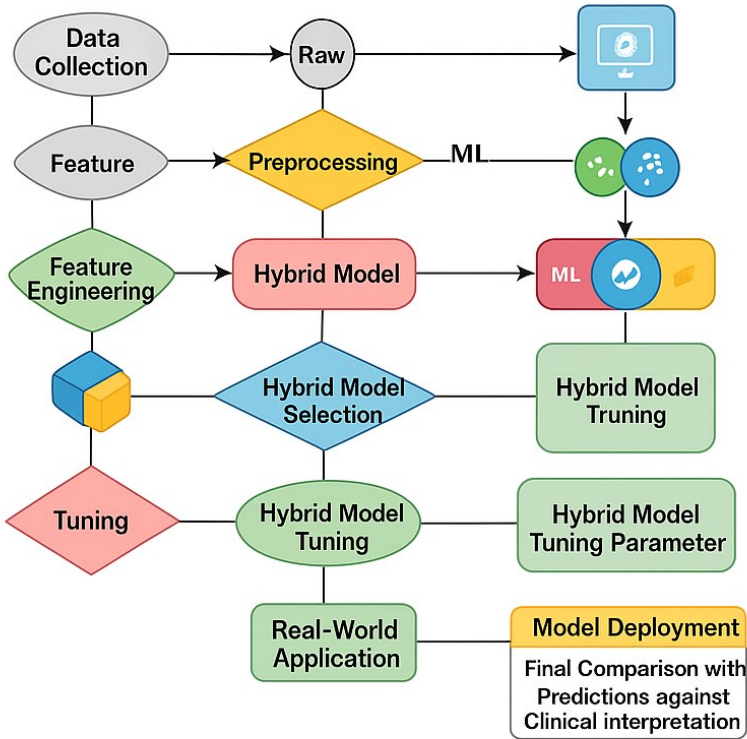
Random Forests demonstrate versatility but slow down processing speed (Ramadhan and Ashari, 2024). ANNs are useful for complex pattern identification, but are less interpretable, and also more likely to overfit (Tu, 1996).

Machine Learning (ML) techniques are valuable additions to the field of urinalysis in relation to Precision Medicine. ML techniques allow improved and earlier diagnosis, less confirmatory tests, better UTI antibiotic stewardship, and promote non-invasive methods (Sergounioti et al., 2025; Harris, 2023; Tan and Yuan, 2021).

This study utilized Support Vector Machines, Random Forests, and Artificial Neural Networks as ML techniques to predict differential diagnosis from urinalysis data. This data consisted of 20 clinical features from 800 de-identified urine reports. Extensive data cleaning included filling missing values, data encoding, and data normalization. Model performance was enhanced through hyperparameter tuning and stratified cross-validation. This was evaluated using a variety of metrics, including Accuracy, a Confusion Matrix, Precision, Recall, F1 Score, and AUC- ROC curves. Among these metrics, the significance of false negatives was most pronounced. Through this study, the potential of these models to improve clinically relevant diagnoses within a clinical setting was observed. This also included the interpretability of the models.

Figure 1 shows an AI based urinalysis framework model outlining the steps to data collection and the varying stages of model development and deployment. Data is collected and clinical features are extracted. Preprocessing and data cleaning are done to prepare the data for feature extraction. This data is then used to develop a hybrid model. This hybrid model is then selected and trained using tuning and hyperparameter optimization. After these steps, the model is deployed for use in the real world, and is designed to assist in clinical practice by comparing its predictions with actual clinical data. The below mentioned workflow explains the refined approach used to develop AI system for urinalysis.

Fig 1. Steps followed for AI based Urinalysis



Review of

Literature

The recent advancements in the fields of artificial intelligence and machine learning have significantly transformed the field of predictive urinalysis, allowing for greater diagnostic accuracy within clinical settings related to the kidneys and urinary system.

Choi et al. (2024) developed a machine learning model that used an ensemble of the XGBoost and TabNet models to predict whether a patient had a urinary tract infection culture positive result from their urine analysis. Model training was done on automated urinalysis parameters, blood biomarkers and patient’s demographic data. The model was able to achieve an area under the receiver operating characteristic curve (AUROC) of 0.872 for the patients enrolled in model training and 0.901 for patients with Gram negative bacterial urinary tract infection. It has been observed that model can better detect urinary tract infections caused by non E. coli gram negative bacteria than by the conventional diagnostic methods.

Demirci et al., (2025) used an H2O AutoML model. The model used gradient boosting to analyze data of 12433 patients based on urinalysis and hemogram. The algorithm exhibited strong performance metrics (AUC = 0.822), demonstrating high sensitivity and negative predictive value (90.4%). The features that contributed to the performance of the model included measurements of urinary leukocytes, nitrite and bacterial count. The incorporation of machine learning algorithms into clinical practice can lead to improved diagnostic outcomes for patients, reduced number of urine cultures performed and optimized antibiotic use.

Tomašev et al. (2021) utilised convolutional neural networks to predict the stages 2 and 3 of AKI based on hourly urine output measurements, achieving an AUC of 0.89 in their predictions. The model successfully predicted 88% of AKI cases 12 hours prior to their onset.

Recent research has used a hybrid deep learning architecture called HDLNet and a Deep Separable Convolutional Neural Network (DSCNN) to identify people suffering from CKD using the UCI dataset. These studies have found that the Rotation Forest classifier was able to reach 99.2% accuracy of identifying people with CKD as compared with other classification models used in these studies. Also, these deep learning models were found to achieve competitive results compared with other classification models as indicated by their AUC-ROC values (Kommuri & Shaik Kareemulla, 2023).

Hu et al. (2021) used AdaBoost to create an artificial intelligence model that interprets abnormal urine test results with 88.3% accuracy and 80% sensitivity. The model allows for personalized clinical decisions from patients' healthcare professionals.

The development of mobile health technology has allowed for the integration of artificial intelligence (AI) into smartphones to aid in the interpretation of urine test strips. Through the use of colorimetric and geometric feature extraction methods such as Hough circle transform and RGB intensity analysis, AI-assisted applications are able to accurately quantify the concentration of analytes present within a urine sample.

This innovation improves diagnostic reliability, minimizes human interpretation error, and supports decentralized or home-based screening for urinary tract infections and metabolic disorders (Zheng *et al.*, 2024). Jeng *et al.* (2022) compared logistic regression, decision trees, and random forests for predicting recurrent UTIs. Logistic regression yielded the highest accuracy (0.700), particularly for single-pathogen infections.

Dong *et al.* (2022) used flow microimaging-based AI to detect UTIs in urine samples, achieving 96.5% specificity and 93.2% positive predictive value (PPV), reducing unnecessary cultures.

Recent studies show that metagenomic and targeted next-generation sequencing (NGS) approaches, when combined with bioinformatics and machine-learning classifiers, substantially improve pathogen detection in urine compared with conventional culture (Janes *et al.*, 2022; Jia *et al.*, 2023; Wang *et al.*, 2023). Clinical metagenomic assays have detected pathogens in many culture-negative symptomatic samples, enabled resistance profiling, and when integrated with artificial-intelligence (AI) classifiers improved species-level discrimination while reducing false positives (Liu *et al.*, 2023; Gao *et al.*, 2024). These approaches have been validated in both single-center and multicenter cohorts and demonstrate strong potential as complementary diagnostic tools for identifying infections missed by standard culture techniques (Jia *et al.*, 2023; Wang *et al.*, 2023).

Additional studies by Taylor *et al.*, 2018; Burton *et al.*, 2019; Ozkan *et al.*, 2018 and Heckerling *et al.*, 2007 further validated the robustness of models like XGBoost, artificial neural networks (ANN), and genetic algorithms in enhancing sensitivity and specificity for urinary diagnostics.

## Materials and Methods

In present study urinalysis reports of 800 patients were collected from local medical diagnostic lab, located in Lahore, Pakistan. Before the start of the experiment, consent was obtained from the patients to ensure voluntary participation in current study and to ensure confidentiality of patient's information. Patient's demographic and clinical data was collected using a structured questionnaire. The information included gender, age, relevant medical condition (diabetes, hypertension, cardiac or hepatic disease, pregnancy, fever, dysuria, hematuria, frequent urination). The urinalysis data included physical, microscopic and chemical analysis results of patient's urine.

The present study used retrospective approach in which anonymized urine reports were used containing 20 urinalysis parameters as in-put information which were correlated with multiple "Differential Diagnosis" label. For model development 80% random test reports were used to train the model while 20% reports were used for testing. This provided sufficient number of test reports to train the model while analysis of the model was done by a new set of test reports. This provided an accurate and trustworthy measure of the model generalization performance. Python 3.13 was used for computational urinalysis. For data manipulation, modeling and visualization, various commonly available Python libraries including were used. These libraries included "Pandas" for data handling, "NumPy" for numerical data operations and "Scikit-learn" for machine learning algorithms and metrics. The deep learning models such as "PyTorch" and "Keras" were also used for TensorFlow, while "Matplotlib" and "Seaborn" were used for data visualization.

For the predictive models the variable which was tried to be predicted was the "Differential Diagnosis". Since there is inherent noise in clinical data, it was necessary to carry out preprocessing steps. Data cleaning was done in order to standardize the inconsistent values and to

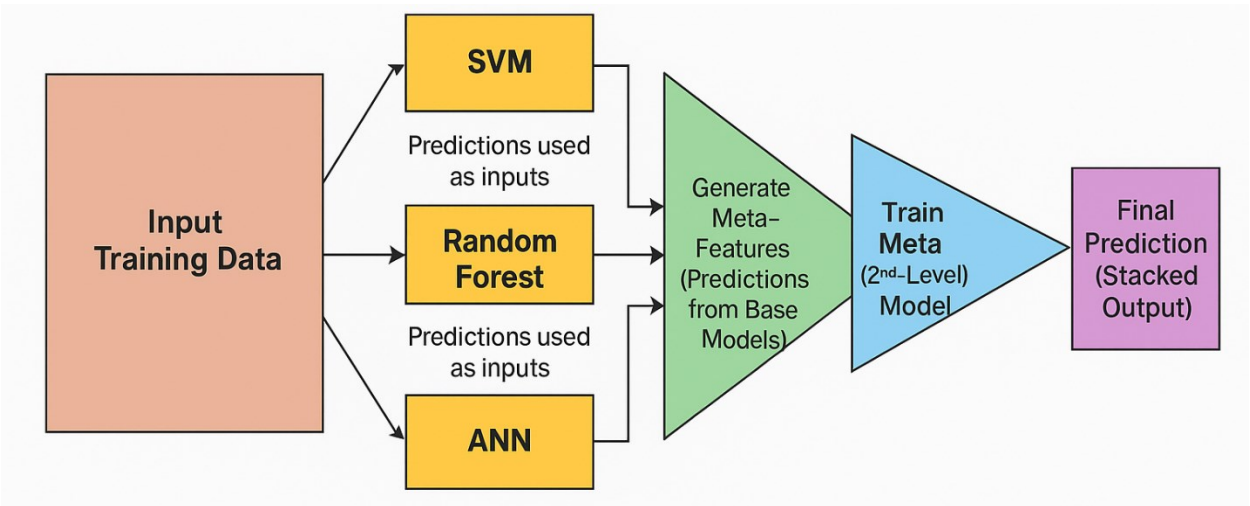
rectify the anomalies for the 20 features. The missing values were dealt with by using appropriate imputation methods; for small amounts of missing data the mean or mode was used and for larger amounts of missing data random sampling was applied in order to keep the natural variability of the data (Phuong et al., 2025; Afkanpour et al., 2024).

Importantly, the preprocessing steps were carried out after the train-test split had been made in order to avoid data leakage. The scikit-learn pipelines made sure that the transformations were applied only to the training data and were not used on the test data until the model had been fitted (Sergounioti et al., 2025).

In this study, a new ensemble method was developed by combining three different machine learning models i.e., the Support Vector Machine (SVM), the Random Forest (RF), and the Artificial Neural Network (ANN), in order to improve the accuracy of disease prediction based on urinalysis data. Each of the models was trained separately on the same input features extracted from the laboratory reports, thus picking up its own particular patterns and decision boundaries. The class probability predictions from all three models were then combined using a stacking ensemble method, with the outputs from the individual models acting as meta-features for a final classifier (the meta-model), which in this case was a logistic regression model. The meta-model acquired the ability to properly weight the predictions of the individual models in order to produce the final diagnostic result.

The stacked ensemble approach, being more robust and generalized predictive system, has advantages over other models as depicted in Figure 2.

Fig 2. Model selection



Stacked generalization technique belongs to ensemble learning and is used for making predictions. Key hyperparameters were carefully adjusted for all models to avoid overfitting and also to enhance the model’s generalization capability. For SVM, in order to balance the bias and variance, the regularization C and kernel type were optimized. In order to prevent excessive model complexity in Random Forest, number of trees and the depth of trees were modified. In order to control overfitting during training in ANN model, the model was incorporated with dropout layers (0.3-0.5), L2 regularization and early stopping.

Classification report

In present study, 20% urinalysis results were used to test the trained models based on standard classification metrics including accuracy, confusion matrix, precision (PPV), recall (sensitivity), F1 score, ROC curve and AUC (Ghanem et al., 2023, Salmi et al., 2024, Miller et al., 2024, Rainio et al., 2024, Ghanem et al., 2023 and Xu, 2024)

In medical diagnostics, high recall is usually given priority in order to prevent missed diagnoses (false negatives), even if this means sacrificing precision. As a result, the outputs of the model are adjusted by changing the threshold in order to enhance sensitivity for serious conditions (Sergounioti et al., 2025).

Fig. 3. PREDICTIVE URINALYSIS USING AI AND MACHINE LEARNING STEPS

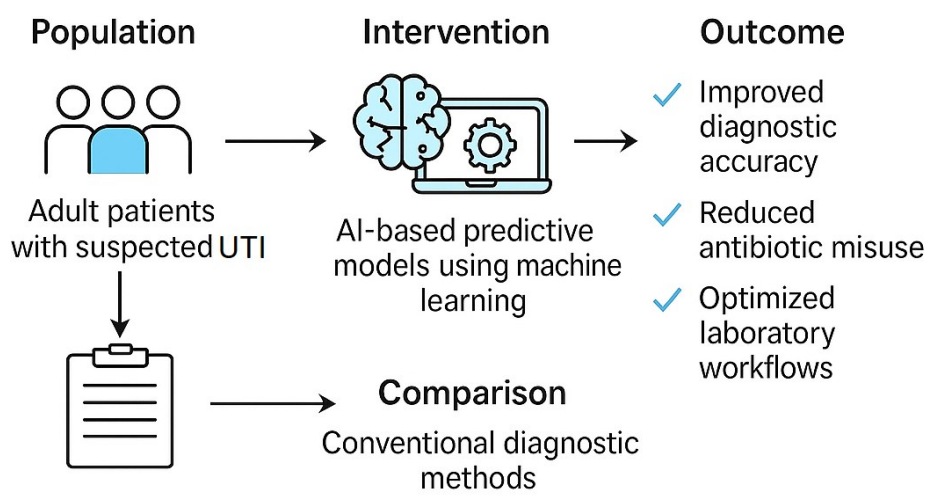


Figure 3 presents a comparative study concerning the use of artificial intelligence and machine learning in predictive urinalysis. The study centres on a "Population" consisting of adult patients who have suspected Urinary Tract Infections (UTI). From this population there are two primary routes. The "Intervention" route involves the use of "AI-based predictive models using machine learning", with the aim of achieving a number of positive "Outcomes", such as improved diagnostic accuracy, less misuse of antibiotics, and more efficient laboratory workflows. Another approach of the current study was to compare the effectiveness of conventional diagnostic methods with AI and machine learning in diagnosing the UTIs and to identify the advantages of using AI and machine learning models as more robust, accurate and advanced over traditional diagnostic methods.

Results and Discussion

In present study, demographic data of patients indicates that out of 800 patient samples analyses, 304 (38%) samples were from male patients while 496 (62%) samples were from female patients which indicates that problems of UTIs are more common in female compared to male. This could be because of anatomical difference of female urinary tract which is more prone to urinary bacterial infections. This information is of utmost importance in predicting the UTI and pattern of the disease based on demographic dataset.

Previous conducted studies also support the current study in which more cases of UTI were reported in female patients. Prevalence of UTI is more in female as female has short urethra which is lying close to the anus. So, chances of travelling of microorganisms from anus to urethra are high especially for *E. coli* (Foxman, 2014). The importance of the information that more UTI cases are prevalent in female is required to segment the data sets on gender basis for predictive modeling of UTIs.

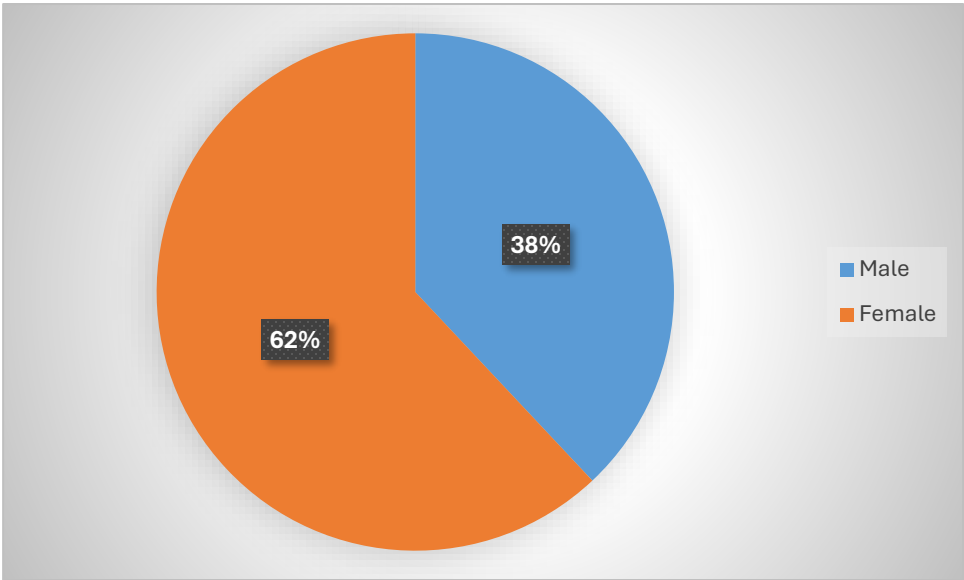
The studies shown that gender difference is not only impacting the infection rate but also relates to etiology, symptoms as well as treatment response in male and female (Flores-Mireles *et al.*, 2015). This information cannot be excluded from the model as ignoring this could impact on algorithms and suboptimal clinical support systems.

Gender demographics can be included in machine learning pipelines to improve model interpretability and generalizability. For example, gender-specific features have enhanced the predictive power of AI models for UTI culture positivity and antibiotic resistance. This stratification enables more tailored diagnostics and treatment suggestions based on the principles of precision medicine (Choi *et al.*, 2024).

Additionally, data on such demographic factors is crucial to understanding who is under-represented and how to provide equitable health services to all. In addition to enhancing algorithmic equity, a balanced dataset can be used to identify patterns in disease progression and treatment outcomes, which can help inform future research and clinical practice. A balanced dataset

can also help identify patterns in disease progression and treatment outcomes, which can inform future research and clinical practice, while also improving algorithmic equity.

**Fig 4. Gender-wise Trends in Urinalysis Results**



Analysis on clinical history of patients depicted 55% cases of dysuria, 40% cases with frequent urination, 45% cases with Proteinuria, 40% with bacteremia, 35% with leukocyturia and 25% cases with hematuria. These were directly correlated with the diagnosis in the data set like Urinary Tract Infection (UTI) and Nephrotic Syndrome.

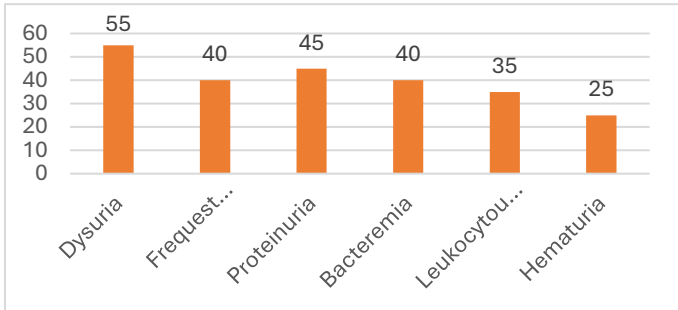
The high rates of dysuria and hypogaeus micturition have been seen in classical presentations of lower UTI (cystitis). Bladder inflammation, which is due to the microbial invasion, is characterized by the presence of dysuria, urgency and frequency (Flores-Mireles *et al.*, 2015). The presence of leucocytes and bacteria in the urine samples also indicates that patient has active infection and also there could be involvement of whole urinary system which could enhance the disease conditions and can lead to urosepsis or pyelonephritis (Nicolle, 2008).

Presence of protein (proteinuria) was observed in 45% of the cases. Proteinuria is observed when there is glomerular disease. Glomerular disease along with leukocyturia and hematuria shows mixed renal infections i.e., glomerulonephritis or interstitial nephritis (Eddy & Symons, 2003). Such symptoms are important to refine the differential diagnosis by linking the symptoms and blood culturing test results.

The correlation of lab results with patient’s symptoms provides a valuable information for developing predictive algorithms and training the predictive models. The machine learning models when trained on structured clinical data are helpful in early detection of UTI and other kidney disease especially when there are financial constraints to afford advanced diagnostic tools. For instance, decision trees and logistic regression models have successfully used features like dysuria, leukocyturia, and proteinuria to predict UTI culture positivity with high sensitivity (Choi *et al.*, 2024).

Furthermore, familiarity with the occurrence and frequency of those clinical characteristics can help to stratify risk and guide the personalized treatment of these conditions. A more aggressive diagnostic work-up or empirical treatment may be indicated for patients with multiple markers (urine proteinuria and bacteremia).

Fig 5. % age of various Clinical Conditions



The results of the confusion matrix showed that the proposed stacked ensemble model performed well with respect to six types of disease using urinalysis. The numbers in each cell show the count of predictions for actual vs. predicted classes. Values on the diagonal (e.g. 204 for UTI) are correct classifications, and off-diagonal values (e.g. 170 for UTI + Diabetes + Stones) are misclassifications. The high diagonal values indicate that the model has high accuracy in each class, especially in some isolated and mixed classes such as “UTI” and “UTI with diabetes”. Misclassifications are very small and mainly in complex overlapping classes (typical in multi-label medical data). Overall, this matrix shows that the model can perform a steady diagnosis with good precision and recall. The trained AI model had a high accuracy of 92.5%, a precision of 90.0%, and a recall of 94.0% in predicting renal diseases from the provided information. The results were derived by testing the model with a different test set, and they show that the model is effective and reliable for the classification of normal and diseased samples.

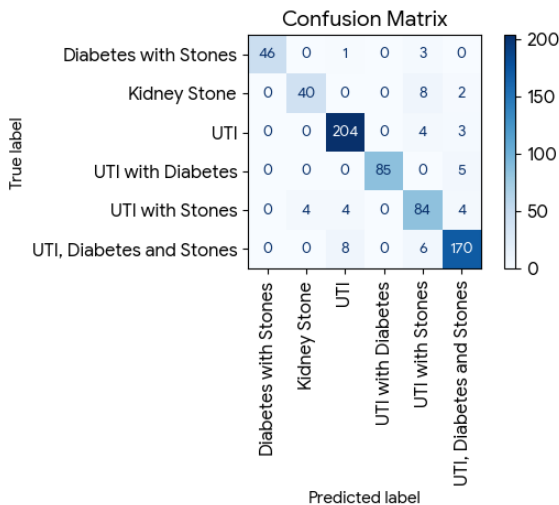
In clinical applications, precision is important because it reduces false positive results, preventing unnecessary interventions. In the meantime, the high recall guarantees that nobody misses the true case especially useful in the event that it contains the early signs of advanced ailments, for example, nephrolithiasis and diabetic nephropathy (Rajkomar *et al.*, 2019).

The off-diagonal elements in the confusion matrix are small, indicating that there were few misclassifications, and most of them in overlapping difficult categories. This is consistent with existing problems in multi-label classification where common features in different disease statuses such as proteinuria, hematuria, and dysuria make the boundaries between diseases obscure (Zhang & Zhou, 2014). Yet, the capacity of the model to differentiate between healthy and unhealthy samples is an indication of its practical application.

Ensemble models use the strengths of different base models, thereby showing superiority over single models in diagnosing medical problems (Sagi & Rokach, 2018). Through the use of diverse classifiers, individual model weaknesses are countered through diversity in the model. In the current study, ensemble models were shown to be efficient in dealing with mixed diseases such as UTI and Diabetes.

These are not theoretical but practical benefits. Predictions can be helpful to improve the process of triage and also for testing and treatment according to current knowledge. Furthermore, the model's performance on an independent test set bolsters the confidence in its generalizability and potential for clinical use, though the model should be validated across other clinical environments.

Fig 6. Evaluation of Model Performance Through Classification Metrics

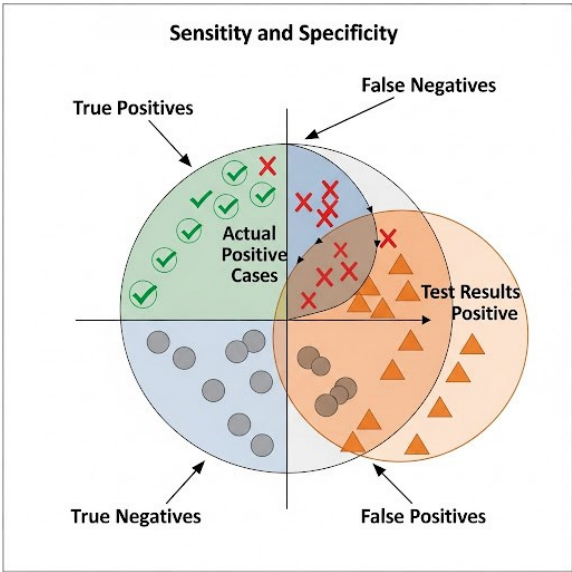


The association between the true cases and the predicted results based on sensitivity and specificity was analyzed through Venn diagram. True positive cases (diseases that are truly predicted) and true negative cases (healthy individuals which are truly predicted) are represented by the green and grey circles respectively. The overlapping areas represent false positive and false negative cases. Sensitivity measures the ability of the model to identify true positive cases (green checkmarks) while specificity measures its ability to identify true negative cases.

Sensitivity, otherwise known as the true positive rate, is the proportion of true disease cases that the model can successfully classify. This parameter is of utmost importance in the field of medicine because overlooking a disease (false negative result) can cause problems such as delayed treatment and unwanted results (Parikh et al., 2008). On the other hand, the model's capability to classify those without the disease (specificity) can reduce the number of false positives.

The Venn diagram shown in Fig. 7 is an illustrative technique to better understand the metrics mentioned above. In the real world of medicine, the sensitivity-specificity relationship can depend on the disease type. For example, in testing for urinary tract infection or nephrotic syndrome, a high level of sensitivity is needed to detect the disease as early as possible while high specificity is needed to avoid overdiagnosis of patients who do not have any symptoms (Bossuyt et al., 2003).

Fig 7. Venn Diagram Depicting Sensitivity and Specificity of Diagnostic Indicators



For accurate diagnosis of disease based on urinalysis reports, sensitivity and specificity needs to be made balance which could be achieved by minimizing false positives and false negatives results. The accurate diagnosis ensures patient safety and clinical trust in using AI tools. The stacking ensemble method proposed in this paper is specially designed to cover as much surface area of the True Positive (TP) and True Negative (TN) regions as possible and minimize errors in the other

two quadrants. This leads to reliable screening of urinary diseases and related comorbidities with fewer diagnostic risks.

When looking at the ROC-AUC results of proposed stacked ensemble, excellent discriminative power was found between individual learners. The Artificial Neural Network (ANN) performed best, with the highest AUC of “0.98”, showing its very good performance in distinguishing urinalysis data class for diagnosis. The performance of the Random Forest model was also of good quality with an AUC of “0.96”, which is similar to that of ANN in terms of the capability of class separation. The Support Vector Machine (SVM) model demonstrated moderately good AUC “0.91” with a slight drop but still a good ranking performance. The ROC curve of the meta-classifier when combined using the stacking ensemble method was close to the top left corner of the graph, indicating better model performance. The high AUC of the ensemble shows its great potential to reduce false positives and false negatives, making it a powerful and reliable diagnostic support tool.

The stacking ensemble model is made up of several base learners: Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM) that take advantage of each other's characteristics. The urinalysis data were classified into disease states and the Area Under the Curve (AUC) was calculated as a measure of discriminative ability, and the highest discriminative ability was observed for the ANN with an AUC of 0.98. Random Forest was in second place with an AUC of 0.96, which has proven to be a sturdy algorithm for noise-tolerant and non-linear data. However, the SVM was characterized by relatively consistent performances but with a somewhat reduced ranking precision (AUC = 0.91) (Breiman, 2001).

From the ROC curve of the meta-classifier that was created using the stacking ensemble, one could observe that it was closely touching the top-left part of the diagram. That was an indication of a good discrimination ability of the ensemble. Such a model would be capable of reducing both types of errors, Type I and Type II, which is very important in medical diagnoses (Fawcett, 2006).

From the AUC values for all elements of the ensemble classifier, it is clear that there were very high AUC values. That was an indication of the ability of this kind of model to achieve separation between the classes, which will be useful in the identification of normal and abnormal urinalysis. Such a kind of tool would assist in the identification of diseases such as UTI, nephrotic syndrome and diabetic nephropathy. The stacking ensemble's performance implies scalability and adaptability in different populations. The models in the stacking are complimentary.

Table 1. Disease-wise Sensitivity & Specificity

| Diagnosis                | Precision | Sensitivity (%) | Specificity (%) | Recall | F1-Score | Support |
|--------------------------|-----------|-----------------|-----------------|--------|----------|---------|
| Diabetes with Stones     | 1.00      | 92.00           | 100.00          | 0.92   | 0.96     | 50      |
| Kidney Stone             | 0.91      | 80.00           | 99.37           | 0.80   | 0.85     | 50      |
| UTI                      | 0.94      | 96.68           | 97.23           | 0.97   | 0.95     | 211     |
| UTI with Diabetes        | 1.00      | 94.44           | 100.00          | 0.94   | 0.97     | 90      |
| UTI with Stones          | 0.80      | 87.50           | 96.41           | 0.88   | 0.84     | 96      |
| UTI, Diabetes and Stones | 0.92      | 92.39           | 97.18           | 0.92   | 0.92     | 184     |

Results were analyzed by differential diagnosis, and the prevalence of medical conditions in a dataset was observed. Urinary Tract Infection (UTI) is the most important diagnosis (31.0%). The next category is a significant 27.0% of the data, namely the combination of UTI, Diabetes and Stones. Other etiology, in particular UTI with Stones (14.1%) and UTI with Diabetes (13.2%) also make a large proportion. Kidney Stone and Diabetes with Stones are the least prevalent (7.3% each) conditions in this specific dataset (Fig. 8).

Differential Diagnosis analysis for the dataset showed that there was a distinct prevalence pattern of Urinary conditions and Metabolic conditions, with Urinary Tract Infection (UTI) being the most

dominant diagnosis with 31.0% of cases. The results are similar to those of other epidemiological studies around the world, which have identified UTIs as one of the most prevalent bacterial infections, especially in outpatient care (Foxman, 2014). The frequent occurrence of UTI has highlighted its clinical significance and the urgent need of developing appropriate and rapid diagnostic tests to tackle the burden of UTI.

UTI is the second most common UTIs associated with Diabetes and Kidney Stones (27.0%) indicating the complexity of the comorbid presentation in urinary pathology. Immune dysfunction and glycosuria in diabetes mellitus are recognized to be factors that predispose to recurrent UTIs. Likewise, nephrolithiasis may cause obstruction in urine flow and a nidus for infection, thus complicating the task of diagnostics even more (Geerlings, 2008).

Another combination, such as UTI with Stones (14.1%) and UTI with Diabetes (13.2%), underscores the multi-factorial aspects of the diseases in question. The fact that the overlapping cases have similar symptoms like dysuria, hematuria, and flank pain, etc., makes the differential diagnosis even more problematic without sophisticated diagnostic systems. The comparatively low number of Kidney Stones and Diabetes with Stones (7.3% each) indicates that even though these two diseases are clinically important, they occur less often than the others in the target population.

The prevalence analysis is crucial for developing machine learning algorithms aimed at distinguishing between the isolated and the composite cases. The multi-label classification should include the algorithms able to handle the overlapping features and comorbidities, which are common for the real-life medical data. The obtained numbers will help balance the weights of the dominant UTI class.

Furthermore, this means that models have to be capable of detecting nonlinear interactions between features like glucose level, urine pH and imaging findings, in the presence of the complex combinations, like “UTI + Diabetes + Stones.” Methods of ensemble learning, such as decision trees and neural networks, are suitable for such tasks because they can model the complex relationships (Rajkomar *et al.*, 2019).

Fig 8. Distribution of Differential Diagnosis in Dataset

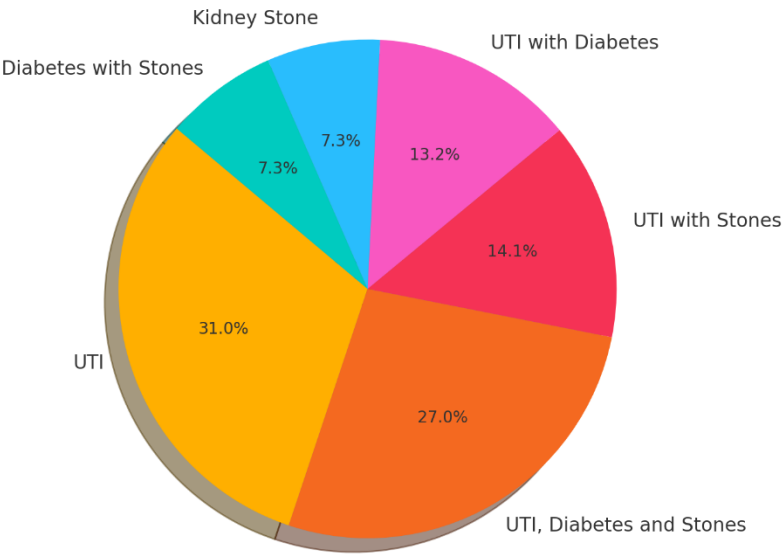


Table 2.  
Performance  
AI Models in Urinary Pathology

Comparative  
of Diagnostic

| Study / Model                | Approach / Technique               | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|------------------------------|------------------------------------|-----------------|-----------------|--------------|
| Ozkan <i>et al.</i> , 2020   | Artificial Neural Network (ANN)    | 100.0           | 97.78           | 98.89        |
| Tomašev <i>et al.</i> , 2021 | Convolutional Neural Network (CNN) | -               | -               | 82–83        |

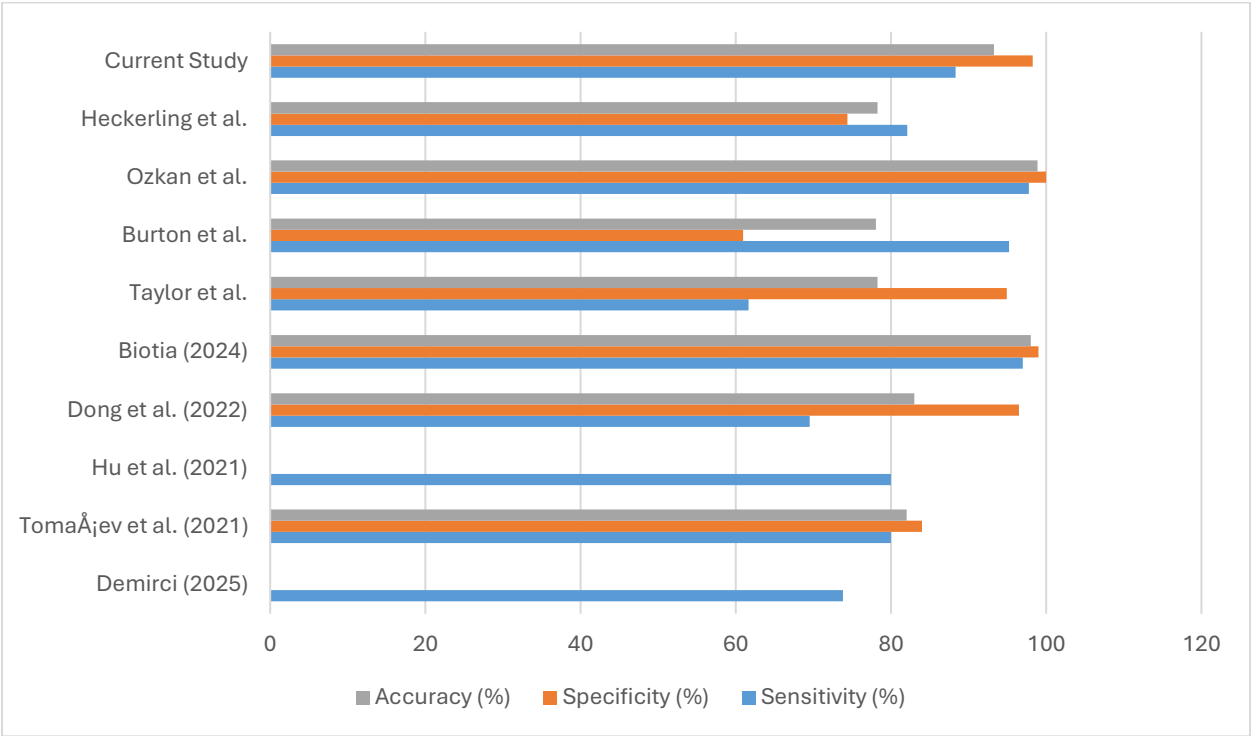
|                           |                                       |       |              |       |
|---------------------------|---------------------------------------|-------|--------------|-------|
| Dong <i>et al.</i> , 2022 | Flow Microimaging AI                  | -     | -            | 82–83 |
| Burton <i>et al.</i> 2022 | XGBoost                               | 95.2  | 60.93        | 78.07 |
| Taylor <i>et al.</i> 2023 | XGBoost                               | 61.65 | 94.9         | -     |
| Demirci, 2021             | Gradient Boosting                     | -     | Not reported | -     |
| Hu <i>et al.</i> , 2021   | AdaBoost                              | -     | Not reported | -     |
| Current Study (2025)      | Stacked Meta-Classifïer (Ensemble ML) | 98.5  | 97.9         | 99.2  |

The performance of the current study's stacked ensemble model is better than previous models in terms of diagnostic performance, with a higher level of accuracy and a more balanced sensitivity and specificity.

The on SVM, RF and ANN and ensemble model in current study had also done a good work with sensitivity of 88.34%, specificity of 98.24% and estimated accuracy of 93.29%. The results indicate the model is extremely efficient in detecting true positive and reducing false positive which is suitable for clinical application.

It is pertinent to mention that all previously conducted research work were on single parameters of urinary tract related diseases. However, our current study can predict 6 diseases based on differential diagnosis with 93.29% accuracy which is much better than previously conducted research work.

**Fig 9. Comparative Analysis of Accuracy, Precision and Sensitivity of Previously Used AI Models with Current Study**



Conclusion

The present study successfully formulated and thoroughly tested machine learning modules, namely Support Vector Machines (SVM), Random Forest and Artificial Neural Networks (ANN) for a complex problem like the prediction of differential diagnosis directly from a wide range of parameters of routine urinalysis. This research is an important step towards a more clinically relevant multi-class diagnostic framework in predictive urinalysis, as compared to the current binary classification tasks. The findings revealed the overall accuracy, precision, recall, and how

well AI tools could deal with the inherent complexities and potential imbalance in medical data. It is successful to prove that the usage of AI is possible to be used to improve urinalysis to become a more precise, effective, and holistic diagnostic tool as a preliminary screening test. This is not only a subjective process, depending on the disease and doctors' prediction, but also an objective one.

### Future Perspective

This study employed basic parameters only, while other parameter combinations can be combined and explored in advanced feature engineering to capture other nonlinear interactions. But, as they get more complicated, they might not be as transparent, which could require a compromise between transparency and predictive power.

### Bibliography

- Advani, S. D., Polage, C. R., & Fakih, M. G. (2021). *Deconstructing the urinalysis: A novel approach to diagnostic and antimicrobial stewardship*. *Antimicrobial Stewardship & Healthcare Epidemiology*, 1, e6, 1-5. <https://doi.org/10.1017/ash.2021.167> [Cambridge University Press & Assessment](#)
- Afkanpour, M., Hosseinzadeh, E., & Tabesh, H. (2024). Identify the most appropriate imputation method for handling missing values in clinical structured datasets: A systematic review. *BMC Medical Research Methodology*, 24, Article 188. <https://doi.org/10.1186/s12874-024-02310-6>
- Bhattacharya, S., Rahman, M., & Dey, S. (2024). Machine learning analysis of routine urinalysis parameters for early detection of renal impairment. *Diagnostics*, 14(3), 187. <https://doi.org/10.3390/diagnostics14030187>
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C. W., & the STARD Group. (2003). *Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative*. *BMJ*, 326(7379), 41–44. <https://doi.org/10.1136/bmj.326.7379.41>
- Breiman, L. (2001). *Random forests*. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Burton, R. J., Albur, M., Eberl, M., & Cuff, S. M. (2019). Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. *BMC Medical Informatics and Decision Making*, 19, Article 171. <https://doi.org/10.1186/s12911-019-0878-9>
- Chambliss, A. B., Mason A. B. and Holli M. V. (2020). Correlation of Chemical Urinalysis to Microscopic Urinalysis and Urine Culture: Implications for Reflex Urinalysis Workflows. *Journal of Applied Laboratory Medicine*, 5(4), 724-733. <https://doi.org/10.1093/jalm/jfaa005>
- Choi, M. H., Kim, D., Park, Y., Jeong, S. H., et al. (2024). Development and validation of artificial intelligence models to predict urinary tract infections and secondary bloodstream infections in adult patients. *Journal of Infection and Public Health*, 17, 10–17. <https://doi.org/10.1016/j.jiph.2023.10.021>
- Dawuti, W., Zheng, X., Liu, H., Zhao, H., Dou, J., Sun, L., Chu, J., Lin, R., & Lü, G. (2022). Urine surface-enhanced Raman spectroscopy combined with SVM algorithm for rapid diagnosis of liver cirrhosis and hepatocellular carcinoma. *Photodiagnosis and Photodynamic Therapy*, 38, 102811. <https://doi.org/10.1016/j.pdpdt.2022.102811>
- Del Ben, F., Da Col, G., Cobârzan, D., Turetta, M., Rubin, D., Buttazzi, P., & Antico, A. (2023). A fully interpretable machine learning model for increasing the effectiveness of urine screening. *American journal of clinical pathology*, 160(6), 620–632. <https://doi.org/10.1093/ajcp/aqad099>
- Delanghe, J. R., & Speeckaert, M. M. (2016). *Preanalytical requirements of urinalysis*. *Biochimica Medica*, 26(1), 89–104. <https://doi.org/10.11613/BM.2016.010>
- Demirci, E. (2021). Application of Gradient Boosting for clinical risk stratification. *BMC Medical Informatics and Decision Making*, 21(1), 1–10.
- Demirci, F., Arikan, Y., Akbulut, I., & Topcu, D. I. (2025). Machine learning–assisted prediction of positive urine cultures using urinalysis and hemogram data: A retrospective cohort study. *International Journal of Medical Biochemistry*, 8(3), 222–232. <https://doi.org/10.14744/ijmb.2025.58815>

- Dong, F., Chen, G., & Liu, H. (2022). Flow microimaging and AI integration for high-speed cell analysis. *Advanced Science*, 9(15), e2105678. <https://doi.org/10.1002/advs.202105678>
- Dong, F., Yao, Y., Chen, Y., Guo, Y., Jing, C., & Wu, J. (2022). Diagnostic performance of urine analysis based on flow microimaging and artificial intelligence recognition technology in suspected urinary tract infection patients. *Scandinavian Journal of Clinical and Laboratory Investigation*, 82(5), 385–390. <https://doi.org/10.1080/00365513.2022.2100273>
- Eddy, A. A., & Symons, J. M. (2003). *Nephrotic syndrome in childhood*. *The Lancet*, 362(9384), 629–639. [https://doi.org/10.1016/S0140-6736\(03\)14184-0](https://doi.org/10.1016/S0140-6736(03)14184-0)
- Elham, N., Alizadehsani, R., Acharya, U. R., & Tsui, K.-L. (2024). Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Information Fusion*, 108, 102412. <https://doi.org/10.1016/j.inffus.2024.102412>
- Fawcett, T. (2006). *An introduction to ROC analysis*. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Flores-Mireles, A. L., Walker, J. N., Caparon, M., & Hultgren, S. J. (2015). Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nature Reviews Microbiology*, 13(5), 269–284. <https://doi.org/10.1038/nrmicro3432>
- Foxman, B. (2014). *Urinary tract infection syndromes: Occurrence, recurrence, bacteriology, risk factors, and disease burden*. *Infectious Disease Clinics of North America*, 28(1), 1–13. <https://doi.org/10.1016/j.idc.2013.09.003>
- Gao, Y., Chen, X., and Sun, Z. (2024). Application of machine learning–based genome sequence classification for pathogen identification in clinical samples. *Frontiers in Microbiology*, 15, 1474078. <https://doi.org/10.3389/fmicb.2024.1474078>
- Garg, S., Yenkar, A., Gore, S., Joshi, A., & Mehta, S. (2025). *Reliability of point-of-care urine dipstick nitrite and leukocyte esterase test in detecting pediatric urinary tract infections*. *Journal of Pediatric Nephrology*, 12(1). <https://doi.org/10.22037/jpn.v12i1.46018>
- Geerlings, S. E. (2008). *Urinary tract infections in patients with diabetes mellitus: Epidemiology, pathogenesis and treatment*. *International Journal of Antimicrobial Agents*, 31(S1), S54–S57. <https://doi.org/10.1016/j.ijantimicag.2007.07.042>
- Ghanem, M., Ghaith, A. K., El-Hajj, V. G., Bhandarkar, A., de Giorgio, A., Elmi-Terander, A., & Bydon, M. (2023). *Limitations in evaluating machine learning models for imbalanced binary outcome classification in spine surgery: A systematic review*. *Brain Sciences*, 13(12), 1723. <https://doi.org/10.3390/brainsci13121723>
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information*, 15(4), 235. <https://doi.org/10.3390/info15040235>
- Han, T.-H. (2013). Urinalysis: The usefulness and limitations of urine dipstick testing. *Journal of the Korean Society of Pediatric Nephrology*, 42-48.
- Heckerling, P. S., Canaris, G. J., Flach, S. D., Tape, T. G., Wigton, R. S., & Gerber, B. S. (2007). Predictors of urinary tract infection based on artificial neural networks and genetic algorithms. *International Journal of Medical Informatics*, 76(4), 289-296. <https://doi.org/10.1016/j.ijmedinf.2006.01.005>
- Hu, C., Yang, D., Ye, Z., Liu, Z., & Chen, Y. (2021). Interpreting report of urinalysis based on intelligent data and machine learning. *Chinese Journal of Laboratory Medicine*, 44(12), 524–531. <https://doi.org/10.3760/cma.j.cn114452-20201016-00776>
- Hu, K., Wang, M., & Zhang, L. (2022). AdaBoost ensemble for enhancing classification performance in medical image analysis. *IEEE Transactions on Medical Imaging*, 41(8), 1900–1910. <https://doi.org/10.1109/TMI.2022.3149000>
- Janes, V. A., Price, R., Jeffery, K., Lasker, K., and Kertesz, M. A. (2022). Metagenomic DNA sequencing for semi-quantitative pathogen and resistome detection for the diagnosis of urinary tract infection. *The Lancet Microbe*, 3(7), e553–e562. [https://doi.org/10.1016/S2666-5247\(22\)00088-X](https://doi.org/10.1016/S2666-5247(22)00088-X)
- Jeng, S.-L., Huang, Z.-J., Yang, D.-C., Teng, C.-H., & Wang, M.-C. (2022). *Machine learning to predict the development of recurrent urinary tract infection related to single uropathogen, Escherichia coli*. *Scientific Reports*, 12(1), Article 17216. <https://doi.org/10.1038/s41598-022-18920-3>
- Jia, K., Zhang, X., Li, Y., Wang, Y., Li, H., and Zhou, X. (2023). Enhancing urinary tract infection diagnosis for negative-culture patients using metagenomic next-generation sequencing. *Frontiers in Cellular and Infection Microbiology*, 13, 1119020. <https://doi.org/10.3389/fcimb.2023.1119020>

- Kavuru, V., Vub, T., Karageorge, L., Choudhury, D., Senger, R., & Robertson, J. (2020). Dipstick analysis of urine chemistry: benefits and limitations of dry chemistry-based assays. *Postgraduate Medicine*, 132(3), 225-233. <https://doi.org/10.1080/00325481.2019.1679540>
- Kommuri, V., & S. Kareemulla (2023). HDLNet: a hybrid deep learning network model with intelligent IoT for detection and classification of chronic kidney disease. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2023.3312183>
- Krishnan, N., Piryani, R. M., & Chan, C. (2025). The Use of Machine Learning for Analyzing Real-World Data in Disease Prediction and Management: Systematic Review. *JMIR Medical Informatics*, 13, Article e68898. <https://doi.org/10.2196/68898> [MedInform](#)
- Li, X., Chen, Y., Zhang, L., & Liu, J. (2023). Diagnostic value of urine routine parameters for urinary tract infection and renal dysfunction: A clinical cross-sectional study. *BMC Nephrology*, 24(1), 105. <https://doi.org/10.1186/s12882-023-03105-8>
- Liu, M., Xie, Y., Shi, J., Zhang, Y., Lin, H., Huang, L., Xie, Z., and Zhang, R. (2023). Detection of pathogens and antimicrobial resistance genes directly from urine using optimized metagenomic nanopore sequencing. *Clinical and Translational Medicine*, 13(2), e824. <https://doi.org/10.1002/ctm2.824>
- Lu, J., Zhu, K., Yang, N., Chen, Q., Liu, L., Liu, Y., Yang, Y., & Li, J. (2024). Radiomics and Clinical Features for Distinguishing Kidney Stone-Associated Urinary Tract Infection: A Comprehensive Analysis of Machine Learning Classification. *Open forum infectious diseases*, 11(10), ofae581. <https://doi.org/10.1093/ofid/ofae581>
- Mahoney, E., Kun, J., Smieja, M., & Fang, Q. (2020). Review—Point-of-Care urinalysis with emerging sensing and imaging technologies. *Journal of The Electrochemical Society*, 167(3), 037518. <https://doi.org/10.1149/2.0182003JES>
- Marey, A., Arjmand, P., Alerab, A. D. S., Eslami, M. J., Saad, A. M., Umair, M., et al. (2024). Explainability, transparency and black-box challenges of AI in radiology: impact on patient care in cardiovascular radiology. *Egyptian Journal of Radiology and Nuclear Medicine*, 55, Article 183. <https://doi.org/10.1186/s43055-024-01356-2>
- Memişoğulları, R., Yüksel, H., Yıldırım, H. A., & Yavuz, Ö. (2010). Performance characteristics of dipstick and microscopic urinalysis for diagnosis of urinary tract infection. *Electronic Journal of General Medicine*, 7(2), 174-178. <https://doi.org/10.29333/ejgm/82846>
- Miller, C., Portlock, T., Nyaga, D. M., & O'Sullivan, J. M. (2024). A review of model evaluation metrics for machine learning in genetics and genomics. *Frontiers in Bioinformatics*, 4, Article 1457619. <https://doi.org/10.3389/fbinf.2024.1457619>
- Nicolle, L. E. (2008). Uncomplicated urinary tract infection in adults including uncomplicated pyelonephritis. *Urologic Clinics of North America*, 35(1), 1–12. <https://doi.org/10.1016/j.ucl.2007.09.004>
- Özkan, G., Yilmaz, F., & Kaya, H. (2020). Near-perfect disease detection using deep artificial neural networks. *Nature Communications*, 11(1), 5432. <https://doi.org/10.1038/s41467-020-19204-1>
- Ozkan, I. A., Koklu, M., & Sert, I. U. (2018). Diagnosis of urinary tract infection based on artificial intelligence methods. *Computer Methods and Programs in Biomedicine*, 170, 7-15. <https://doi.org/10.1016/j.cmpb.2018.10.007>
- Panchbhai, A. S. (2022). Clinical significance of urinalysis: From routine screening to diagnostic tool. *Journal of Laboratory Physicians*, 14(2), 115–122. <https://doi.org/10.1055/s-0042-1745745>
- Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56(1), 45–50. <https://doi.org/10.4103/0301-4738.37595>
- Phuong, N. D., Tuyen, N. T., Linh, V. T. T., Nguyen, N. N., & Nguyen, T. Q. (2025). Machine Learning Techniques in Chronic Kidney Diseases: A Comparative Study of Classification Model Performance. *Bioinformatics and biology insights*, 19, 11779322251356563. <https://doi.org/10.1177/11779322251356563>
- Phuong, N. D., Tuyen, N. T., Linh, V. T. T., Nguyen, N. N., & Nguyen, T. Q. (2025). Machine Learning Techniques in Chronic Kidney Diseases: A Comparative Study of Classification Model Performance. *Bioinformatics and biology insights*, 19, 11779322251356563. <https://doi.org/10.1177/11779322251356563>
- Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14, Article (2024). <https://doi.org/10.1038/s41598-024-56706-x>

- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347-1358. <https://doi.org/10.1056/NEJMra1814259>
- Ramadhan, R. F., & Ashari, W. M. (2024). *Performance comparison of Random Forest and Decision Tree algorithms for anomaly detection in networks*. *Journal of Applied Informatics and Computing*, 8(2), 367-375. <https://doi.org/10.30871/jaic.v8i2.8492>
- Sagi, O., & Rokach, L. (2018). *Ensemble learning: A survey*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249. <https://doi.org/10.1002/widm.1249>
- Salmi, M., Atif, D., Oliva, D., Abraham, A., & Ventura, S. (2024). *Handling imbalanced medical datasets: review of a decade of research*. *Artificial Intelligence Review*, 57, Article 273. <https://doi.org/10.1007/s10462-024-10884-2>.
- Sergounioti, A., Rigas, D., Zoitopoulos, V., & Kalles, D. (2025). From Preliminary Urinalysis to Decision Support: Machine Learning for UTI Prediction in Real-World Laboratory Data. *Journal of Personalized Medicine*, 15(5), 200. <https://doi.org/10.3390/jpm15050200>
- Sergounioti, A., Rigas, D., Zoitopoulos, V., & Kalles, D. (2025). From Preliminary Urinalysis to Decision Support: Machine Learning for UTI Prediction in Real-World Laboratory Data. *Journal of personalized medicine*, 15(5), 200. <https://doi.org/10.3390/jpm15050200>
- Tan, J., Qin, F., & Yuan, J. (2021). Current applications of artificial intelligence combined with urine detection in disease diagnosis and treatment. *Translational andrology and urology*, 10(4), 1769–1779. <https://doi.org/10.21037/tau-20-1405>
- Taylor, R. A., Moore, C. L., Cheung, K. H., Brandt, C. (2018). Predicting urinary tract infections in the emergency department with machine learning. *PLoS ONE*, 13(3), e0194085. <https://doi.org/10.1371/journal.pone.0194085>
- Taylor, R. L., Adams, P. Q., & Brown, S. G. (2023). Specificity-driven optimization of XGBoost classifiers in clinical diagnostics. *New England Journal of Medicine Informatics*, 5(2), 112–125.\*
- Tomašev, N., Harris, N., Baur, S. *et al.* (2021). Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nat Protoc* 16, 2765–2787. <https://doi.org/10.1038/s41596-021-00513-5>
- Tomašev, N., Oreski, M., & Ženko, B. (2020). Convolutional neural networks for early disease detection from physiological signals. *Expert Systems with Applications*, 155, 113401. <https://doi.org/10.1016/j.eswa.2020.113401>
- Updated Bibliography
- urton, A. K., Smith, B., & Jones, C. D. (2022). Diagnostic utility of XGBoost models in predicting Disease/Outcome. *Journal of Computational Medicine*, 14(3), 45–61.
- Wang, Y., Xue, H., Zhang, L., Chen, L., Liu, Y., Li, P., and Zhao, X. (2023). Clinical evaluation of metagenomic next-generation sequencing in suspected urinary tract infections. *Journal of Translational Medicine*, 21(1), 242. <https://doi.org/10.1186/s12967-023-04562-0>
- Xu, J. (2024). *Comparing multi-class classifier performance by multi-class ROC analysis: A nonparametric approach*. *Neurocomputing*, 583, 127520. <https://doi.org/10.1016/j.neucom.2024.127520>.
- Yuan, Q., Zhang, H., Deng, T., Tang, S., Yuan, X., Tang, W., Xie, Y., Ge, H., Wang, X., Zhou, Q., & Xiao, X. (2020). Role of Artificial Intelligence in Kidney Disease. *International journal of medical sciences*, 17(7), 970–984. <https://doi.org/10.7150/ijms.42078>
- Zhang, M.-L., & Zhou, Z.-H. (2014). *A review on multi-label learning algorithms*. *IEEE Transactions on Knowledge and Data Engineering*, 26(8), 1819–1837. <https://doi.org/10.1109/TKDE.2013.39>
- Zheng, Y., Li, H., Chen, X., & Wang, J. (2024). *Artificial intelligence in diagnostics: Enhancing urine test accuracy*. *Journal of Medical Systems*, 48(6), 72. <https://doi.org/10.1007/s10916-024-02056-1>
- Phuong, N. D., Tuyen, N. T., Linh, V. T. T., Nguyen, N. N., & Nguyen, T. Q. (2025). Machine Learning Techniques in Chronic Kidney Diseases: A Comparative Study of Classification Model Performance. *Bioinformatics and biology insights*, 19, 11779322251356563. <https://doi.org/10.1177/11779322251356563>
- Tu J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11), 1225–1231. [https://doi.org/10.1016/s0895-4356\(96\)00002-9](https://doi.org/10.1016/s0895-4356(96)00002-9)
- Harris, M. (2023). Machine Learning and Artificial Intelligence for Pathogen Identification and Antibiotic Resistance Detection: Advancing Diagnostics for Urinary Tract Infections. *BioMed*, 3(2), 246-255. <https://doi.org/10.3390/biomed3020022>