IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155283

SPECTROSCOPIC BIOMARKER DETECTION FOR URINE DISEASES USING MACHINE LEARNING

Mr. Surendhiran S[1] Mr. Saran K[2], Ms. Savina S[3], Mrs. Nirupashri G[4]

Final Year Student, Biomedical Engineering, Dhanalakshmi Srinivasan College of Engineering and Technology (Autonomous), Mamallapuram, Chennai, India[1]

Final Year Student, Biomedical Engineering, Dhanalakshmi Srinivasan College of Engineering and Technology (Autonomous), Mamallapuram, Chennai, India[2]

Final Year Student, Biomedical Engineering, Dhanalakshmi Srinivasan College of Engineering and Technology (Autonomous), Mamallapuram,Chennai, India[3]

Assistant Professor, Biomedical Engineering, Dhanalakshmi Srinivasan College of Engineering and Technology (Autonomous), Mamallapuram, Chennai,India[4]

Abstract: Urinary diseases such as kidney disorders, urinary tract infections, and diabetic nephropathy are becoming major health concerns worldwide. Early detection of these diseases is important to prevent severe complications and improve treatment outcomes. Traditional diagnostic methods are usually time-consuming, expensive, and dependent on laboratory testing. This paper proposes a machine learning–based framework for urine disease classification using spectroscopic biomarker analysis. The system analyses important urine parameters such as glucose, protein, pH, RBC, WBC, ketone, and bacteria to predict disease severity. Various machine learning algorithms including Support Vector Classifier, Logistic Regression, and Bernoulli Naive Bayes are used for prediction. The trained model is deployed using the Django framework for real-time disease prediction through a web application. The proposed system provides a rapid, accurate, non-invasive, and cost-effective solution for urine disease diagnosis and monitoring.

Keywords: Machine Learning, Urine Disease Classification, Spectroscopy, Biomarker Detection, Support Vector Machine, Non-Invasive Diagnosis.

I. INTRODUCTION

Urinary diseases such as kidney stones, urinary tract infections, and chronic kidney disease can cause serious health problems if not diagnosed early. Traditional diagnostic methods are often time-consuming, expensive, and require laboratory analysis. Urine contains important biomarkers such as glucose, protein, pH, RBC, and WBC that help identify urinary diseases. Recent advancements in spectroscopy and machine learning have improved automated disease prediction systems. The proposed project develops a machine learning–based urine disease classification system using spectroscopic biomarker analysis for fast, accurate, and real-time disease prediction through a web application.

II. LITERATURE SURVEY

Related Work

Kavuru et al. (2023) analysed urine Raman spectra from patients with diabetes mellitus and renal diseases. Principal Component Analysis and classification algorithms were used to identify disease-specific spectral variations in urine samples.

Farooq et al. (2023) proposed a diabetes monitoring system using ATR-FTIR spectroscopy and machine learning techniques. Their work demonstrated that urine spectroscopy combined with machine learning can provide rapid and reliable disease prediction.

Wang et al. (2025) reviewed spectroscopic analytical approaches for kidney disease biomarker detection. Their study explained the effectiveness of spectroscopic methods in identifying biochemical changes associated with renal disorders.

DOI: 10.17148/IJARCCE.2026.155283

Singh et al. (2025) evaluated different deep learning methods for disease classification using biomedical datasets. Their results showed that machine learning and deep learning algorithms improve classification performance and prediction accuracy in healthcare applications.

III. EXISTING SYSTEM

The existing urine disease detection systems mainly depend on traditional laboratory testing and manual medical analysis. Doctors analyse urine samples using microscopic examination, chemical testing, and imaging techniques to identify urinary diseases such as kidney stones, urinary tract infections, and chronic kidney disease. Although these methods provide accurate results, they are time-consuming, expensive, and require skilled medical professionals.

Several traditional machine learning systems have also been developed for urine disease prediction. These systems use basic algorithms and manual feature extraction techniques for disease classification. However, conventional methods often fail to analyse complex biochemical patterns present in urine data, resulting in lower prediction accuracy and reduced efficiency.

Existing systems also face limitations such as delayed diagnosis, lack of real-time prediction, high computational cost, and limited automation. Most traditional diagnostic methods are not integrated with intelligent web-based healthcare platforms and require continuous human involvement. Therefore, an efficient machine learning–based automated urine disease classification system is required for accurate and real-time disease prediction.

Limitations of Existing Systems

The existing urine disease diagnosis systems depend mainly on traditional laboratory methods and manual examination of urine samples. These methods require expensive equipment, trained healthcare professionals, and longer processing time. In many cases, delayed diagnosis may increase disease severity and reduce treatment effectiveness. Traditional systems also lack intelligent automation and real-time prediction features, making disease classification less efficient and time-consuming.

IV. PROPOSED SYSTEM

The proposed system is a machine learning–based urine disease classification framework developed using spectroscopic biomarker analysis. The system is designed to detect and classify urinary diseases using important urine parameters such as glucose, protein, pH, RBC, WBC, ketone, bilirubin, and bacteria. These parameters are analysed using machine learning algorithms to predict disease severity accurately and efficiently.

The proposed framework includes data collection, preprocessing, feature analysis, model training, and prediction stages. During preprocessing, missing values, duplicate records, and inconsistent data are removed to improve dataset quality. Exploratory Data Analysis and visualization techniques are used to identify feature relationships and disease patterns.

Machine learning algorithms such as Support Vector Classifier (SVC), Logistic Regression, and Bernoulli Naive Bayes are trained using the processed dataset. The best-performing model is selected and deployed using the Django framework for real-time prediction through a web application.

The system provides a user-friendly interface where users can register, login, enter urine-related clinical parameters, and obtain disease prediction results instantly. The proposed system offers fast, accurate, non-invasive, and cost-effective diagnosis while supporting early disease detection and clinical decision-making.

V. SYSTEM ARCHTECTURE

The system architecture of the proposed urine disease classification framework is designed to provide accurate, efficient, and real-time disease prediction using machine learning techniques. The architecture integrates urine biomarker analysis, machine learning algorithms, database management, and web application deployment to support automated disease diagnosis. The system consists of multiple modules including data collection, preprocessing, feature analysis, machine learning model training, prediction, result visualization, and database storage.

Initially, urine-related datasets are collected from reliable medical and public sources such as Kaggle. The dataset contains important urine parameters including glucose, protein, pH, RBC, WBC, ketone, bilirubin, and bacteria, which are useful for identifying urinary diseases. The collected dataset is stored and managed for further processing and analysis.

DOI: 10.17148/IJARCCE.2026.155283

After data collection, the preprocessing module performs operations such as missing value handling, duplicate removal, normalization, and data cleaning to improve the quality of the dataset. These preprocessing techniques help remove noise and inconsistencies from the data, resulting in better model performance and prediction accuracy. Feature scaling and data transformation techniques are also applied to maintain uniformity among input parameters.

The processed dataset is then transferred to the Exploratory Data Analysis and visualization module. In this stage, graphs, heatmaps, histograms, and correlation plots are generated using Python libraries such as Matplotlib and Seaborn. These visualization techniques help identify important disease-related features, feature relationships, and hidden patterns within the dataset.

After feature analysis, the machine learning module performs model training and classification. Machine learning algorithms such as Support Vector Classifier (SVC), Logistic Regression, and Bernoulli Naive Bayes are trained using the processed dataset. The trained models analyse urine biomarker patterns and classify disease severity levels accurately. Performance evaluation metrics such as accuracy, precision, recall, and F1-score are used to compare model performance and select the best-performing algorithm.

The selected machine learning model is then deployed using the Django framework and integrated into a web-based application. The frontend interface is developed using HTML, CSS, and JavaScript to provide a simple and user-friendly environment. Users can register, login, enter urine-related clinical parameters, and obtain real-time disease prediction results instantly through the web application.

The prediction output, user details, and disease history are stored in the SQLite database management system for future reference and monitoring. The database module supports secure data storage, retrieval, and tracking of prediction history. Overall, the modular system architecture improves scalability, automation, prediction efficiency, and system reliability while supporting intelligent urine disease diagnosis and clinical decision-making.

VI. WORKING SCENARIO

The proposed urine disease classification system works as an automated machine learning–based healthcare framework designed to provide accurate and real-time disease prediction. The workflow begins when the user enters urine-related clinical parameters such as glucose, protein, pH, RBC, WBC, ketone, bilirubin, and bacteria through the web application interface.

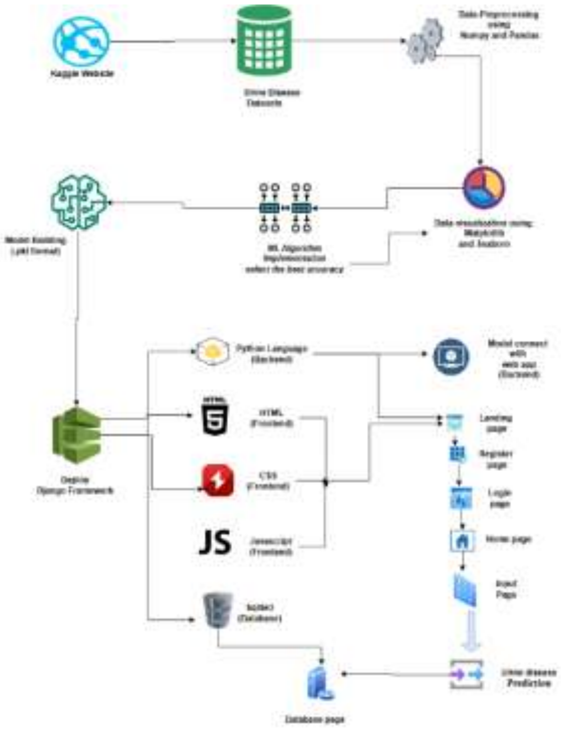

Fig.1 Working Scenario of Urine Disease Prediction System

DOI: 10.17148/IJARCCE.2026.155283

This Fig.3 shows the workflow of the proposed urine disease prediction system, including data input, preprocessing, machine learning analysis, and disease prediction result generation.

Initially, the entered data undergoes preprocessing operations such as missing value handling, normalization, and data cleaning to improve data quality and maintain consistency. These preprocessing techniques help remove noise and improve prediction accuracy.

After preprocessing, the refined data is transferred to the machine learning module for feature analysis and classification. Machine learning algorithms such as Support Vector Classifier (SVC), Logistic Regression, and Bernoulli Naive Bayes analyse the urine biomarker patterns and identify disease-related abnormalities. The trained model then predicts the urine disease type and severity level based on the input parameters.

Once the prediction process is completed, the generated output is displayed through the web application interface. The result page provides disease classification, severity level, and additional medical recommendations for better clinical understanding. The prediction history and user details are stored securely in the SQLite database for future monitoring and analysis.

The proposed automated framework reduces manual workload, improves prediction efficiency, and supports fast, accurate, and non-invasive urine disease diagnosis through a user-friendly web-based platform.

VII. TECHNICAL METHODOLOGY

The proposed urine disease classification system begins with the collection of urine-related datasets containing parameters such as glucose, protein, pH, RBC, WBC, ketone, and bacteria. The collected data is preprocessed using Python libraries like NumPy and Pandas to remove missing values, duplicate records, and inconsistent data.

After preprocessing, Exploratory Data Analysis and visualization techniques are performed using Matplotlib and Seaborn to identify important disease-related patterns and feature relationships. The processed dataset is then divided into training and testing datasets for machine learning analysis.

Machine learning algorithms such as Support Vector Classifier (SVC), Logistic Regression, and Bernoulli Naive Bayes are trained using the processed data. The trained models analyse urine biomarker patterns and classify disease severity accurately. Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate model performance.

The best-performing model is selected and deployed using the Django framework. A web-based application is developed using HTML, CSS, and JavaScript, allowing users to enter urine parameters and obtain real-time disease prediction results. The prediction output and user details are stored in the SQLite database for future reference and monitoring.

Fig.2 Home page

This Fig.2 shows the homepage of the machine learning–based urine disease prediction system with navigation modules and user-friendly healthcare interface.

Fig.3 Model Dataset

This Fig.3 shows the urine disease dataset containing various urine parameters used for training and testing the machine learning prediction model.

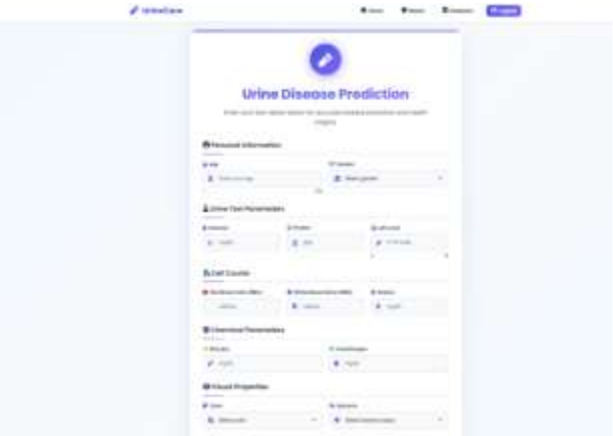

Fig.4 Model page

This Fig.4 shows the machine learning–based urine disease prediction page where users enter urine parameters and personal details for disease analysis and prediction.

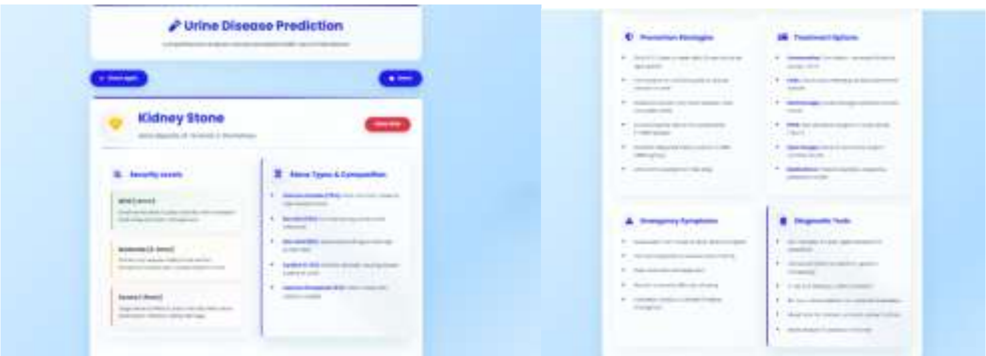

Fig.5 Urine Prediction Result Page

This Fig.5 shows the urine disease prediction result page generated by the machine learning–based web application. The page displays the detected disease type, severity levels, symptoms, treatment options, and diagnostic information based on the entered urine parameters.

VIII. CONCLUSION

This study presented a machine learning–based urine disease classification system using spectroscopic biomarker analysis. The proposed system uses important urine parameters such as glucose, protein, pH, RBC, WBC, and ketone for disease prediction and severity classification. Machine learning algorithms such as Support Vector Classifier, Logistic Regression, and Bernoulli Naive Bayes were used to improve prediction accuracy.

DOI: 10.17148/IJARCCE.2026.155283

The trained model was deployed using the Django framework to provide real-time disease prediction through a user-friendly web application. The proposed system offers a fast, accurate, cost-effective, and non-invasive solution for urine disease diagnosis. Overall, the framework supports early disease detection, clinical decision-making, and efficient healthcare monitoring.

REFERENCES

[1] Delrue, C., Speeckaert, M. M., & De Bruyne, S. (2026). Deep learning algorithms and Raman spectroscopy in the clinical laboratory setting. *Critical Reviews in Clinical Laboratory Sciences*, 63(1), 80–108.

[2] Kavuru, V., Senger, R. S., Robertson, J. L., & Choudhury, D. (2023). Analysis of urine Raman spectra differences from patients with diabetes mellitus and renal pathologies. *PeerJ*, 11, e14864.

[3] Farooq, S., & Zezell, D. M. (2023). Diabetes monitoring through urine analysis using ATR-FTIR spectroscopy and machine learning. *Chemosensors*, 11(11), 589.

[4] Wang, W., Mao, J., Wang, J., Yuan, X., & Huang, K. (2025). Spectroscopic analytical approaches for kidney disease biomarker detection. *Analytical Methods*, 17(4), 210–225.

[5] Singh, A. P., Nigam, A., & Garg, G. (2025). Evaluation of deep learning methods for disease classification. *Current Medical Imaging*, 21(1), 45–58.

[6] Sharma, R., Patel, D., & Verma, S. (2024). Machine learning approaches for urine disease prediction using biomedical data. *International Journal of Medical Informatics*, 175, 105012.

[7] Rahman, M. A., Islam, M. T., & Kabir, M. H. (2025). Ensemble learning framework for spectroscopic disease severity prediction. *Artificial Intelligence in Medicine*, 140, 102580.

[8] Jiao, Z., Gao, C., Li, Y., et al. (2019). A survey on deep learning for disease diagnosis in medical images. *Pattern Recognition*, 102, 107123.

[9] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.

[10] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 25, 1097–1105.

[13] Kutlu, H., & Avci, E. (2020). A novel method for classifying medical datasets using deep learning techniques. *Multimedia Tools and Applications*, 79(25), 18551–18578.

[14] Islam, M. T., Rahman, M. M., & Kabir, M. H. (2021). Deep learning-based medical image classification: A review. *Artificial Intelligence in Medicine*, 113, 102112.

[15] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks (DenseNet). *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708.