An integrated data-driven approach for Chronic Kidney Disease of Unknown Etiology (CKDu) risk profiling and prediction in Sri Lanka

Nethmini Rajapaksha*[a], Himasha Rajawasan[a], Ruvini Ubeysinghe[a], Shanika Perera[a], N. H. P. Ravi Supunya Swarnakantha[a], Manori Gamage[a], Nishntha Nanayakkara[b], Janaka Wijayakulasooriya[c], Dulanjali Herath[d], Madhusha Lakmali[e]

[a]Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka, [b]Nephrology and Kidney Transplant Unit, National Hospital Kandy, Sri Lanka, [c]Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka, [d]Centre for Research, National Hospital Kandy, Sri Lanka, [e]Department of Geology, Faculty of Science, University of Peradeniya, Sri Lanka

*Corresponding author: nethminiirajapaksha@gmail.com

ABSTRACT

CHRONIC KIDNEY DISEASE of UNKNOWN ETIOLOGY is a significant public health issue in Sri Lanka, especially in rural farming communities. The exact causes remain unclear, with potential links to environmental and socio-economic factors. This research employs Biological Data and Geographic Information Systems to analyze risk factors such as water quality, agricultural practices, climatic conditions, Demographic Factors, Socio-economic Factors. This study uses data from government health records, the Centre for Research - National Hospital Kandy, and field surveys. By identifying patterns and correlations, the study aims to inform public health interventions and reduce the impact of CKDu, ultimately improving health outcomes for affected populations. This will greatly contribute to preventing the disease, reducing the risk, and identifying patients at an early stage.

Keywords: CKDu, Machine Learning, GIS, environmental risk factors, Risk Factor Profiling, Epidemiological Analysis, public health, eGFR

1. INTRODUCTION

Chronic Kidney Disease of Unknown Etiology (CKDu) is a significant and growing public health concern, particularly in Sri Lanka. This disease predominantly affects rural agricultural communities, posing challenges to public health systems and impacting the socio-economic stability of affected regions. CKDu is characterized by its enigmatic etiology, with no clear links to conventional causes such as diabetes or hypertension, which necessitates a more nuanced approach to understanding its prevalence and progression [1], [2].

CKDu research demands a multidisciplinary perspective, integrating epidemiology, environmental science, and data analytics to identify its root causes and contributing factors. These factors encompass a wide spectrum, including water quality, occupational hazards, and socio-economic conditions, each playing a role in disease prevalence [3], [4]. This complexity makes it crucial to develop innovative methodologies to profile risk factors, predict disease trends, and design targeted interventions. The presented research aims to address these challenges through an integrated approach that combines advanced data- driven methodologies and epidemiological insights. The focus areas include profiling and simulating CKDu risk factors, particularly emphasizing water quality in Sri Lanka, and conducting spatial distribution analysis with baseline prediction models. The study further delves into understanding the impact of environmental factors, such as agrochemical exposure and climatic conditions, on CKDu prevalence.

The subsequent sections of this paper detail the comprehensive methodology adopted to achieve the research objectives. The Related Work section highlights existing literature on CKDu research and identifies gaps that necessitate this study. The Methodology section elaborates on the multi-stage approach, including data preprocessing, statistical and spatial analyses, backward analysis and forward prediction, and simulation modelling. Techniques such as correlation analysis, Principal Component Analysis (PCA), and machine learning models are employed to enhance the accuracy of profiling and prediction. The integration of Agent-Based Modelling (ABM) and System Dynamics (SD) further enriches the simulation framework, providing actionable insights into disease progression under various scenarios [1], [4].

Seventh International Conference on Information Technology and Computer Communications
(ITCC 2025), edited by Chaofeng Zhang, Proc. of SPIE Vol. 13977, 1397709

The Results and Discussion section presents the findings derived from the analytical and Machine learning models, emphasizing their effectiveness in identifying high-risk groups and assessing the impact of proposed interventions. Visual tools, including heat maps and trend graphs, facilitate a clearer understanding of CKDu prevalence patterns and intervention outcomes [2], [3].

Finally, the Conclusion section provides a comprehensive summary of the research, discussing its implications for public health strategies and potential avenues for future exploration. This study aims to bridge the gap between data- driven methodologies and epidemiological research, offering a robust framework for addressing CKDu and its associated challenges in Sri Lanka.

2. RELATED WORK

Chronic Kidney Disease of Unknown Etiology (CKDu) is a growing public health concern in Sri Lanka, and numerous studies have explored its potential causes and risk factors. Several research efforts have focused on the role of water quality, environmental contamination, and advanced computational techniques in understanding CKDu prevalence and progression.

Numerous studies have explored the relationship between water quality and CKDu prevalence in Sri Lanka. Jayasumana et al. [8] identified high arsenic and fluoride concentrations in groundwater, indicating potential links to CKDu. Hettithanthri et al. [10] analyzed water security in Sri Lanka's dry zone, emphasizing heavy metal contamination and poor sanitation practices as key risk factors. Similarly, Balasooriya et al. [11] examined groundwater geochemistry in the Ginnoruwa region, identifying possible connections between trace elements such as cadmium, lead, and fluoride with CKDu incidence. These studies highlight environmental contamination as a major contributor to CKDu, reinforcing the need for region-specific investigations.

Advancements in computational techniques have enabled more robust CKDu risk analysis. Silva et al. [12] utilized machine learning models to predict CKDu risks using environmental and patient data, while Vasan et al. [13] developed predictive models employing gradient boosting algorithms to enhance early-stage CKDu detection. Sharma et al. [14] further extended this approach by applying deep learning techniques (CNNs, DNNs) to analyze large-scale CKDu datasets, improving risk classification and prediction accuracy. However, most studies focus on national-level datasets, leaving a gap in localized CKDu risk analysis. Addressing this, our study integrates CKDu risk datasets, Multi-parameter Test (MPT) data, and water quality data to develop a comprehensive risk assessment framework, leveraging machine learning and simulation models for targeted prevention and intervention strategies.

Several studies have explored predictive modelling of estimated Glomerular Filtration Rate (eGFR) to assess kidney function decline and CKDu risk. Rutter et al. [15] and Caplin et al. [16] analyzed international patterns of eGFR prevalence through population-based surveys, highlighting regional disparities and socioeconomic influences on kidney disease progression. Their research contributed to the Disadvantaged Populations eGFR Epidemiology Study (DEGREE), which provided a framework for studying eGFR in low- and middle-income populations to better identify at-risk individuals. While these studies emphasize early disease detection, there is still a need for more localized predictive models that incorporate region-specific risk parameters for CKDu.

Spatial analysis techniques have played a crucial role in identifying CKDu hotspots and understanding environmental risk factors. Marasinghe et al. [17] applied geographic information systems (GIS) to analyze spatial patterns of CKDu prevalence, revealing high-risk areas correlated with groundwater fluoride contamination. Their study demonstrated how machine learning-assisted GIS mapping can enhance risk prediction and intervention planning.

Machine learning techniques have been increasingly applied to analyze environmental interactions and predict CKDu risk. Dharmawardhane [3] explored multi-ion interactions in water using machine learning models, demonstrating their role in CKDu development. Athuraliya et al. [4] applied predictive modelling by integrating historical patient records with environmental data to identify at-risk populations. Perera and Samarasekara [8] emphasized the importance of data cleaning in health research. These studies highlight the critical role of AI and data-driven methodologies in improving CKDu risk assessments.

Despite advancements in computational techniques, existing approaches primarily focus on large-scale regional data, highlighting the need for integrated spatial and predictive models tailored to localized CKDu-affected regions. Addressing this gap, our study aims to incorporate localized data sources and predictive modelling techniques to develop a more comprehensive CKDu risk assessment framework.

Biological, environmental, and demographic factors also play a significant role in CKDu research. Dharma-wardana [1] identified tubule interstitial damage as a key pathological feature of CKDu, reinforcing the need for integrating biological markers with geospatial and environmental data. While existing studies have contributed to CKDu risk profiling, most focus on national-level data, leaving a gap in localized risk factor analysis. Additionally, current research lacks a unified framework incorporating genetic, epigenetic, environmental, and behavioral aspects. To address this, the proposed study aims to develop a comprehensive CKDu risk assessment model, integrating simulation techniques and predictive analytics to support targeted intervention strategies.

3. METHODOLOGY

This study employs an integrated methodology to profile CKDu risk factors, analyze spatial distribution, and predict disease progression using machine learning and statistical modelling. The research is structured around four key components: (A) Profiling CKDu Risk Factors, (B) Cohort Study and Spatial Distribution Analysis, (C) Comparative Analysis of CKDu Risk Factors, and (D) Identifying the Impact of Environmental Factors on CKDu. Each component focuses on a different aspect of CKDu assessment, collectively providing a comprehensive understanding of its prevalence and progression in Sri Lanka.

3.1. Profiling CKDu risk factors focusing on water quality as a primary risk factor

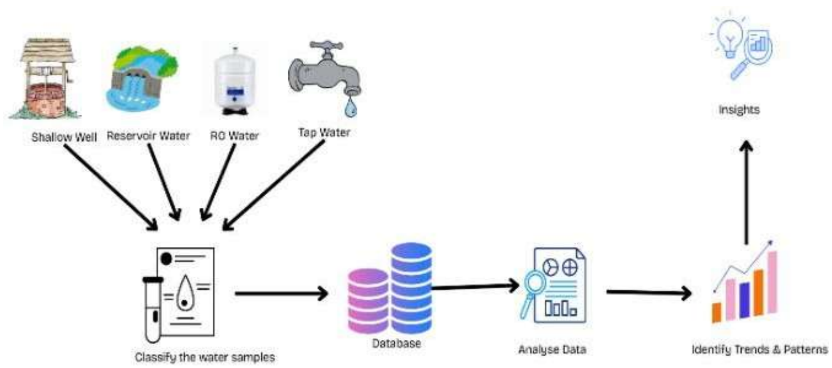

Figure 1. Profiling CKDu Risk Factors in Sri Lanka component diagram

This study focuses on identifying the relationship between water quality and CKDu prevalence.Data is collected from high-risk and low-risk regions in Wilgamuwa, a recognized CKDu-affected area. Wilgamuwa area is a risk area for CKDu. Water samples were obtained from various sources, including dug wells, tube wells, spring water, and filtered water, with key water quality parameters analyzed to assess their role in disease occurrence.

The dataset underwent rigorous preprocessing, including normality testing using the Shapiro-Wilk test and statistical significance evaluation through T-tests and Mann-Whitney U tests. The Logistic Regression Model, Random Forest Model is developed to classify and predict CKDu risk, while feature importance analysis using both statistical and model-based techniques, identifying critical water quality parameters influencing CKDu progression.

Additionally, risk factor distribution is analyzed both location-wise and water source-wise to understand regional variations and the impact of different water sources on CKDu prevalence. This component mainly analyzes water quality data in CKDu risk areas in Sri Lanka.

3.2. Cohort study and spatial distribution analysis of CKDu using machine learning

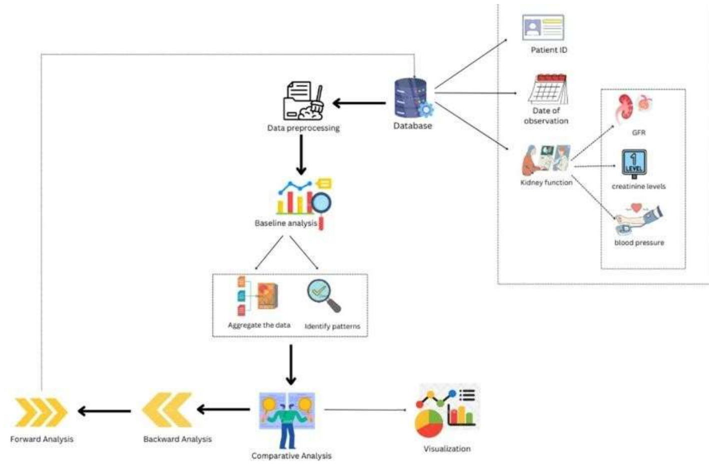

Figure 2. Cohort Study and Spatial Distribution Analysis of CKDu component diagram

A cohort study design was adopted, incorporating demographic, biological, and geographic data from multiple sources. The dataset included information on patient demographics such as age and sex, clinical markers like serum creatinine and eGFR, and environmental factors such as land use patterns and water source composition. The data underwent preprocessing, including outlier detection using the IQR method and missing data imputation using interpolation and KNN techniques. Feature engineering was applied to enhance the dataset by deriving new variables related to environmental exposure and disease progression.

A two-pronged analytical approach was used to study CKDu progression:

1. Backward Analysis: Historical patient data was examined to identify factors contributing to eGFR decline. Feature importance analysis was performed using Random Forest and Gradient Boosting models, while retrospective studies established critical eGFR thresholds for CKDu onset.

2. Forward Prediction: Machine learning models were developed to predict future CKDu progression trends. Long Short-Term Memory (LSTM) networks were applied to analyze time- series data, while XGBoost Models provided robust forecasting. The performance of these models was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R- squared metrics.

3.3. Comparative analysis and profiling of CKDu risk factors

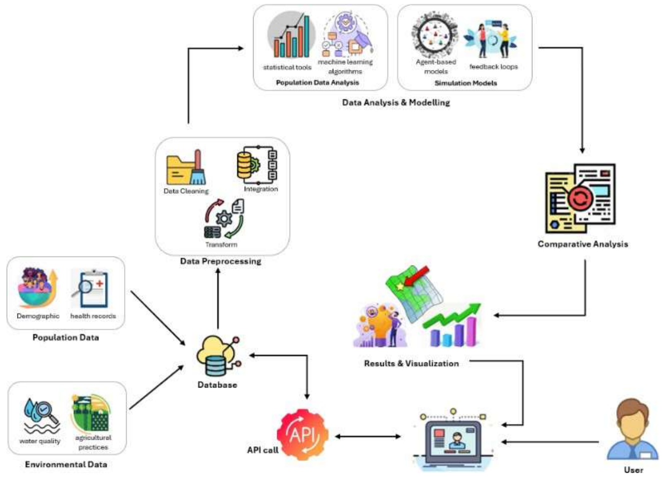

Figure 3. Comparative Analysis and Profiling of Risk Factors component diagram

This component emphasizes the identification and simulation of epidemiological risk factors associated with CKDu prevalence in Sri Lanka. The methodology includes key phases: Data collection, Preprocessing, Epidemiological Profiling, and Modelling.

Data on CKDu cases, demographic details, and environmental parameters were sourced from government health records, Centre for Research - National Hospital Kandy, and field surveys. The dataset underwent preprocessing to ensure consistency and reliability:

- Missing Data Handling: Missing values were addressed using K-Nearest Neighbors (KNN) Imputation for continuous variables and Mode Imputation for categorical variables.
- Outlier Detection: The Interquartile Range (IQR) Method was used to detect and remove outliers from variables such as Age, Serum Creatinine, and eGFR.
- Data Normalization: Variables such as Blood Pressure, Potassium, and Hemoglobin Levels were Z-Scored Normalized to ensure comparability.

  Key epidemiological factors were analyzed using statistical and machine learning techniques to determine their influence on CKDu progression.
- Feature Selection: Random Forest Feature Importance Analysis was used to rank variables such as Fertilizer Usage, Herbicide Usage, Well Water Consumption, and Occupation.
- Statistical Analysis:
  1. Kruskal-Wallis Test was applied to identify significant differences in CKD risk across categorical variables such as Education Level and Water Source.
  2. Multivariate Logistic Regression was used to evaluate the relationship between CKDu Status (binary outcome) and predictor variables.
- Clustering Analysis: K-Means Clustering was applied to categorize high-risk and low-risk patient groups based on biomarker distributions.
- Data Visualization: Heat Maps and Scatter Plots were generated to display the spatial distribution of CKDu prevalence and the correlation between risk factors.

Simulation Modelling: The simulation models predict the potential progression of CKDu under various scenarios, such as changes in environmental factors (e.g., water quality, agrochemical use) or demographic shifts.

3.4. Identifying impact of environmental factors on CKDu

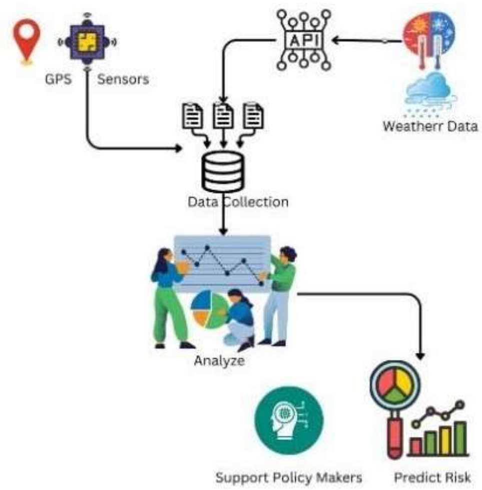

Figure 4. Identifying Impact of Environmental Factors on CKDu component diagram

Environmental data is collected from center for Clinical Audit and Research National Hospital – Kandy, weather data APIs, and field surveys, covering various water sources such as dug well, tube well, natural springs and key indicators such as Mg^{2+}, Ca^{2+}, Na^{+}. Missing values are imputed based on the data distribution using mean, median, or KNN imputation. The statistical analysis begins with exploratory data visualization using boxplots to assess the distributions of physicochemical parameters (e.g., Mg^{2+}, Ca^{2+}, Na^{+}) across CKDu risk groups. Normality of the data is rigorously evaluated via Shapiro-Wilk tests, which confirm non-normal distributions for all variables ($p < 0.05$). Consequently, non-parametric Mann-Whitney U tests are employed to compare median differences between risk (CKDu-positive) and non-risk groups, complemented by Cliff's Delta to quantify the magnitude and direction of observed effects. For categorical data, a chi-square test of independence is applied to evaluate the association between water source types (e.g., dug wells, tube wells) and CKDu risk status, using a contingency table to analyze frequency distributions. This multi-faceted approach ensures robust identification of risk factors while accounting for non-normality and categorical dependencies in the dataset.

This component mainly analyzes water quality data between CKDu risk areas and non-risk areas in Sri Lanka.

4. RESULTS AND DISCUSSION

This section presents the key findings of the study, integrating Statistical Analysis, Machine Learning Models, and environmental assessments to evaluate the factors contributing to CKDu risk. The results are systematically structured to provide a comprehensive understanding of the relationships between water quality, biomarkers, environmental factors and CKDu progression. The discussion contextualizes these findings by comparing them with previous research and highlighting their implications for public health interventions and policymaking.

The statistical analysis of water quality parameters across CKDu-affected (Naminigama) and unaffected (Hasalaka) regions revealed notable differences. Statistical tests, including the Shapiro-Wilk normality test, T-test, and Mann- Whitney U test, identified Na^{+}, Mg^{2+}, Ca^{2+}, and F^{-} as significantly elevated in CKDu-affected areas. But EC and TDS are not having significant relationship with CKDu.

Fisher's exact test and descriptive statistical analysis further identified Na^{+}, Ca^{2+}, Mg^{2+}, Si^{4+}, HCO_3^{-}, Cl^{-}, and SO_4^{2-} as significant contributors to water quality variations in affected regions.

The application of machine learning models for CKDu risk classification provided high accuracy. Logistic Regression achieved 93.75% accuracy, with Na^{+}, Mg^{2+}, Ca^{2+}, and F^{-} emerging as the most influential predictors. Random Forest modelling identified key environmental variables impacting CKDu risk, with an overall accuracy of 98%.

Feature importance analysis revealed Mg^{2+}, Ca^{2+}, and Na^{+} as the top contributors to CKDu prediction. Scatter plots demonstrated significant differences in Na^{+}, Mg^{2+}, and Ca^{2+} concentrations between risk and non-risk datasets, emphasizing their predictive relevance. According to Accuracy VS Features graphs Top contributors for CKDu is Na+, Mg2+, K+, Br-. Considering all those outputs and results, Na+ and Mg+2 have a strong relationship with CKDu.

Table 1. Summary of Significant Variables

Test	Significant Variable
1.T Test / U Test	Remove TDS ,EC
2.Fishers Exact Test Descriptive statistical Analysis	Na+ ,Ca2+ ,Mg2+ ,Si4+, HCO3-,Cl-,SO42-
3.Machine Learning Models	Na+ ,mg2+ ,ca2+ ,F-
4.Scatter Plots	Na+ ,mg2+ ,ca2+
5. Accuracy VS Features graphs	Na+,Mg2+,K+,Br-

A trend analysis of eGFR values across multiple patient visits indicated a consistent decline in kidney function. The computed slope of -3.11 from the linear regression model suggested a steady reduction in eGFR over time. Classification based on individual eGFR decline rates showed that most patients (169 individuals) remained stable, whereas only one patient experienced rapid decline, and another showed improvement.

Further categorization of patients into rapid decline, stable, and improvement groups reinforced these findings. Predictive modelling using Random Forest Regression yielded strong predictive performance (MAE: 0.23, RMSE: 0.31), with serum creatinine at the latest visit (scrmv6) identified as the most influential factor.

Backward analysis confirmed historical eGFR values and serum creatinine levels as significant determinants of current kidney function ($R^2 = 0.95$), emphasizing the progressive nature of CKD and the importance of routine monitoring.

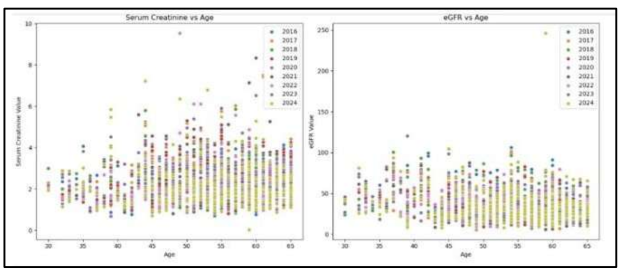

Figure 5. Age- wise eGFR and Serum Creatinine Value Change

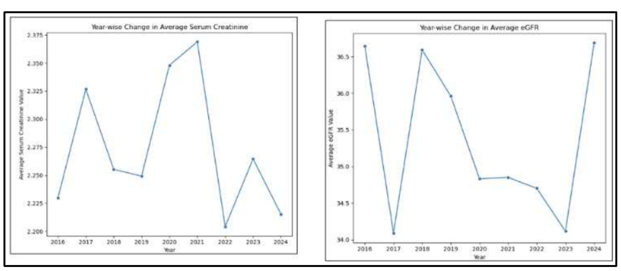

Figure 6. Year-wise changes for Serum Creatinine and eGFR values (Avg)

The findings of Comparative Analysis and Profiling of Risk Factors largely align with previous research on CKDu while introducing new insights into biomarker interactions and predictive modelling techniques. Similar to past studies, statistical tests such as ANOVA and Kruskal-Wallis confirmed the significance of age, well-water history, herbicide usage, hemoglobin (Hb), uric acid, and potassium in CKDu progression (p-values < 0.05). These results reinforce existing knowledge on environmental and biochemical risk factors associated with CKDu. However, this study identified nickel (Ni) as a previously underexplored predictor, with Random Forest feature importance analysis ranking Ni concentrations among the top contributing factors.

Additionally, machine learning models demonstrated superior predictive accuracy compared to conventional epidemiological methods. Logistic Regression and Random Forest achieved 94.4% and 94.3% accuracy, respectively, outperforming earlier studies that relied primarily on statistical correlation methods. Simulation experiments further validated these models, illustrating how modifying key risk factors, such as water quality and agrochemical exposure, directly influences CKDu progression. Unlike previous research, which primarily focused on static analysis, this study integrates predictive modelling and scenario-based simulations to assess long-term trends in disease progression.

Table 2. Proportional Impact of Interventions on CKD Progression

Intervention	Reduction in CKD Progression (%)
Herbicide Exposure Reduction	12.5%
Well Water Filtration Improvements	18.0%
Early Anaemia Treatment (Hb Levels)	15.2%

The environmental analysis reveals critical insights into water quality risk factors and their association with source types in the Kandy region. Mann-Whitney U tests identify significant differences ($p < 0.05$) between risk and non-risk groups for Na^+, Ca^{2+}, Mg^{2+}, F^-, Total Alkalinity, NO_3^-, Depth of Water Column, SO_4^{2-}, and PO_4^{3-} (Table 1). Large effect sizes (Cliff's $\Delta \geq 0.43$) are observed for Ca^{2+} ($\Delta = 0.73$), Mg^{2+} ($\Delta = 0.86$), F^- ($\Delta = 0.49$), and Depth of Water Column ($\Delta = -0.48$), indicating substantial practical significance. These findings suggest geogenic contamination, likely from groundwater interactions with mineral-rich bedrock, particularly in shallow sources. Elevated Mg^{2+} and Ca^{2+} align with limestone dissolution patterns in tropical regions, while high F^- may reflect fluoride-rich geological strata—a known issue in Sri Lanka's dry zones. The negative effect of Depth of Water Column underscores the vulnerability of shallow water sources to surface contaminants, such as agricultural runoff or fecal matter, consistent with studies linking lower depths to microbial and chemical infiltration.

NO_3^- exhibits a medium negative effect ($\Delta = -0.37$), implying lower nitrate levels in risk-associated samples, which contrasts with typical agricultural contamination patterns. This anomaly may reflect localized dilution effects or rapid denitrification processes in specific aquifers. Conversely, Cl^- and Sample Depth show negligible effects, suggesting these variables are less critical to risk stratification in this context.

The chi-square test confirms a strong association between water source type and risk status ($p < 0.001$). Drilled wells (57.4% risk) and filtered sources (87.5% risk) are disproportionately associated with contamination. The prevalence of risk in dug wells—a common rural water source—highlights infrastructural vulnerabilities, such as inadequate sealing or proximity to septic systems. The high risk in filtered systems may reflect improper maintenance or recontamination post-treatment. In contrast, tap lines, natural springs, and tube wells are predominantly low-risk, likely due to regulated infrastructure or natural protection from surface pollutants. However, small sample sizes for categories like *stock tanks* ($n = 1$) and *surface water* ($n = 4$) limit conclusive inferences for these sources.

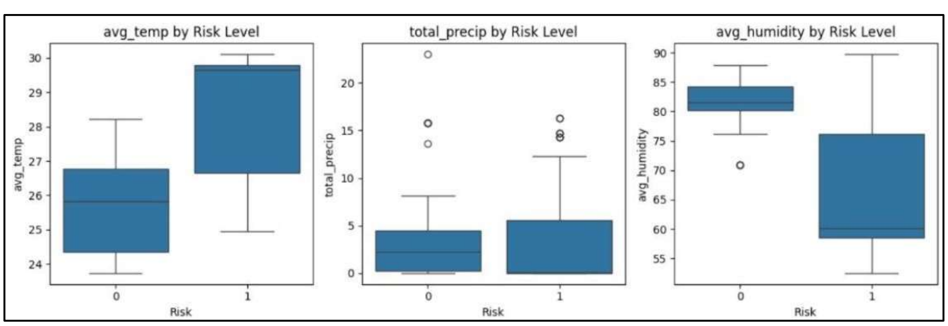

Figure 7. Weather & Risk

5. CONCLUSION

This research presents a comprehensive approach to analyzing and profiling chronic kidney disease of Unknown Etiology (CKDu) risk factors, with a focus on developing robust models to better understand its progression and prevalence. By integrating epidemiological data, simulation models, and spatial distribution analyses, this study highlights the significant influence of environmental, demographic, and lifestyle factors on CKDu. The methodologies employed ranging from water quality assessments to advanced statistical and machine learning techniques have enabled the identification and profiling of key risk factors, which are essential for preventive measures and effective policymaking.

The study of comparative analysis of CKDu risk factors, identifying key biomarkers and socio-environmental influences that contribute to disease progression. The findings highlight the significance of hemoglobin levels, uric acid, potassium, and water quality parameters as critical predictors of CKDu severity. Machine learning models, particularly Random Forest and Logistic Regression, demonstrated high accuracy in classifying risk groups, while statistical analyses confirmed significant variations across CKD stages. These insights enhance the understanding of CKDu etiology, supporting targeted prevention strategies and early intervention efforts to mitigate disease progression in affected populations.

This environmental study demonstrates that elevated concentrations of geogenic minerals specifically Mg^{2+}, Ca^{2+}, and Na^{+} are strongly associated with CKDu risk, driven by interactions between groundwater and mineral-rich bedrock. Shallow water sources, particularly dug wells, emerge as high-risk due to their susceptibility to surface contamination and geological leaching, while regulated infrastructure (e.g., tap lines) shows protective effects. Non-parametric statistical methods reinforce these findings, highlighting the practical significance of geogenic contaminants and the vulnerability of shallow water systems. However, limitations such as seasonal data gaps and underrepresented source types underscore the need for longitudinal and expanded sampling. These results call for urgent interventions, including community education on water source management, mitigation strategies for geogenic contaminants, and policy reforms to protect vulnerable populations. Future research should prioritize spatial-temporal analyses and interdisciplinary approaches to address CKDu’ s multifactorial etiology in endemic regions.

REFERENCES

[1] C.Dharma-wardana, “Chronic Kidney Disease of Unknown Aetiology (CKDu)—The search for causes and the impact of its politicization,” arXiv preprint arXiv:2303.02479, 2023.

[2] P. Mayadunna, G. Salibi, and N. Tzenios, “Chronic Kidney Disease of Unknown Etiology in Sri Lanka,” Special Journal of the Medical Academy and Other Life Sciences, vol. 2, no. 4, 2024.

[3] M. W. C. Dharma-wardana, “Chronic Kidney Disease of Unknown Aetiology (CKDu) and multiple-ion interactions in drinking water,” arXiv preprint arXiv:1704.07906, 2017.

[4] A. Fernando and N. Sivakumaran, “A comprehensive review of Chronic Kidney Disease of Unknown Etiology,” arXiv preprint arXiv:1803.11253, 2018.

[5] C. Jayasumana, P. Paranagama, S. Agampodi, C. Wijewardane, S. Gunatilake, S. Siribaddana, "Drinking well water and occupational exposure to herbicides is associated with chronic kidney disease, in Padavi- Sripura, Sri Lanka," Environmental Health, vol. 14, no. 6, pp. 1-9, 2015.

[6] M. Jayasumana, P. Paranagama, I. Amarasinghe, C. Wijewardane, S. Dahanayake, D. Gunatilake, "Possible link of chronic arsenic toxicity with chronic kidney disease of unknown etiology in Sri Lanka," Journal of Natural Sciences Research, pp. 69-79, 2015.

[7] O. Hettithanthri, S. Sandanayake, D. Magana- Arachchi, R. Wanigatunge, A. U. Rajapaksha, X. Zeng, Q. Shi, H. Guo, M. Vithanag, "Risk factors for endemic chronic kidney disease of unknown etiology in Sri Lanka: Retrospect of water security in the dry zone," Science of the Total Environment, vol. 657, no. March, pp. 116-127, 2019.

[8] S. Balasooriya, H. Munasinghe, A. T. Herath, S. Diyabalanage, O. A. Ileperuma, H. Manthrithilake, C. Daniel, K. Amann, C. Zwiener, J. A. C. Barth, and R. Chandrajith, "Possible links between groundwater geochemistry and chronic kidney disease of unknown etiology (CKDu): an investigation from the Ginnoruwa region in Sri Lanka," Exposure and Health, p. 12, 823–

[9] R. Silva and J. Fernando, "Agent-based modeling of CKDu interventions in Sri Lanka," Journal of Environmental Simulation, vol. 12, no. 4, p. 45–56, 2021.

[10] K. Vasan, H. Thanushkodi, and M. Senthilkumar, "Predictive Modeling of CKDu Using Machine Learning Algorithms: A Case Study from Sri Lanka, pp. 1005-1015, 2019.

[11] R. Sharma, S. Pathak, and M. Singh, "AI-Based Modeling for Predicting CKDu Risk in Sri Lankan Population," Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Learning Applications (AIMLA), pp. 237-243, 2021.

[12] C. E. Rutter, M. Njoroge, P. J. Cooper, and N. Pearce, "International patterns of prevalence of low estimated glomerular filtration rate (eGFR) in population-based surveys of working-age adults," vol. 35, no. June, 2024.

[13] B. Caplin, K. Jakobsson, J. Glaser, D. Nitsch, V. Jha, A. Singh, R. Correa-Rotter, and N. Pearce, "International collaboration for the epidemiology of eGFR in low and middle-income populations - rationale and core protocol for the Disadvantaged Populations eGFR Epidemiology Study (DEGREE)".

[14] B. S. Marasinghe, W. A. U. Vitharana, R. Chandrajith, N. Nanayakkara, and H. P. U. Fonseka, "Spatial analysis of chronic kidney disease of unknown aetiology (CKDu) prevalence in Badulupura Village, Sri Lanka," vol. 35, 2024.

[15] G. Perera and A. Samarasekara, "Data cleaning and handling in health research," Sri Lankan Journal of Medicine, vol. 29, no. 1, pp. 10-19, 2022.

[16] A. M. P. B. Alahakoon, M. M. Nibraz, P. M. S. S. B. Gunarathna, S. Thenuja, K. A. D. C. P. Kahandawaarchchi, "Water Quality Index Based Prediction of Ground Water Properties for Safe Consumption," in 2020 2nd International Conference on Advancements in Computing (ICAC), 2020.

[17] R. Silva, A. Pathirana, N. Vithanage, K. Ratnayake, "Machine Learning Techniques in Analyzing Environmental Data for Chronic Kidney Disease of Unknown Etiology (CKDu) in Sri Lanka," Proceedings of the International Conference on Environmental and Climate Change, pp. 55-61, 2020.

[18] S. J. Wimalawansa, "The role of ionicity in drinking water in the pathogenesis of chronic kidney disease of multifactorial origin," Ceylon Medical Journal, vol. 60, p. 10–18, 2015.

[19] S. Rajapakse, M. C. Shivanthan, and M. Selvarajah, "Chronic kidney disease of unknown etiology in Sri Lanka," International Journal of Occupational and Environmental Health, vol. 22, no. July, pp. 259-264, 2016.