

A Cohort Study and Spatial Analysis for Predicting Future eGFR Using Machine Learning in Chronic Kidney Disease of Unknown Etiology (CKDu)

W.H.H.S. Rajawasan^{1*}, R.G.N.D. Rajapaksha¹, W.A.R.R. Ubeyasinghe¹, G.S.H. Perera¹,
H.M.D.K. Herath², M. Lakmali³, J.V. Wijayakulasooriya⁴, N. Nanayakkara⁵,
N.H.P.R.S Swarnakantha¹, M.P. Gamage¹

¹*Department of Information Technology, Sri Lanka Institute of Information Technology,
Malabe, Sri Lanka,*

²*Centre for Clinical Audit and Research, National Hospital Kandy, Sri Lanka,*

³*Department of Geology, Faculty of Science, University of Peradeniya, Sri Lanka,*

⁴*Department of Electrical and Electronic Engineering, University of Peradeniya,
Sri Lanka*

⁵*Nephrology and Kidney Transplant Unit, National Hospital Kandy, Sri Lanka,
himashasandanika@gmail.com

Chronic Kidney Disease of Unknown Etiology (CKDu) is a progressive, non-traditional form of chronic kidney disease predominantly affecting rural agricultural communities in Sri Lanka. Early prediction of kidney function decline is crucial for effective intervention. This study analyses longitudinal clinical data to forecast future trends in estimated glomerular filtration rate (eGFR) using machine learning. A dataset of 300 CKDu patients from Wilgamuwa, Matale district (2016–2024), approved by the Kandy Kidney Unit, was used. Each record includes eGFR, serum creatinine, blood pressure, and demographic data. After data cleaning and normalization, machine learning models were applied for both forward prediction (forecasting future eGFR values) and backward analysis (identifying early risk indicators). The models tested include Long Short-Term Memory (LSTM) networks for non-linear sequence prediction, ARIMA for statistical forecasting under stationarity assumptions, and Random Forest and Linear Regression for non-temporal modelling. The performance of each model was evaluated using metrics such as RMSE, MAE, and R². Results revealed a rapid decline in eGFR with a slope of 3.11, indicating severe disease progression. Boxplot analysis exposed significant outliers, while gender-based analysis showed that female patients had higher median eGFR than males. Among the models, ARIMA achieved the best performance, with 78.02% accuracy, MAE of 8.98, and RMSE of 8.99. The findings confirm ARIMA as a robust and interpretable model for CKDu-related eGFR forecasting. Accurate prediction supports early clinical decision-making, targeted care, and improved patient outcomes. Additionally, observed gender differences underline the need for personalised healthcare strategies in CKDu management.

Keywords: CKDu, eGFR prediction, cohort study, spatial analysis, chronic kidney disease