Predicting the Disease Progression of CKDu Patients Based on Clinical, Environmental and Socio-Demographic Factors

H.M.D.K. Herath[1]*, H.T.K. Abeysundara[2], N. Nanayakkara[3], P. Hewavitharane[1]

[1]Centre for Research, National Hospital, Kandy, Sri Lanka
[2]Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka
[3]Nephrology and Kidney Transplant Unit, National Hospital, Kandy, Sri Lanka
**dulanjalikh@gmail.com*

Chronic Kidney Disease of Uncertain Etiology (CKDu) has emerged as a significant global health concern in recent years and in Sri Lanka, it was first identified in 1994. The disease predominantly afflicts certain regions within the country's dry zone, encompassing key agricultural provinces like North Central, Uva, Eastern, and North Western. The causative factors of this disease are still unknown and it is not caused by the traditional risk factors. The aim of this study is to develop a predictive model for estimating the annual eGFR decline of CKDu patients using a comprehensive dataset encompassing clinical, environmental, and socio-demographic factors. The dataset included 304 patients and 45 variables which were cleaned and was split into 80:20 training and testing sets. Initially, Ridge Regression was employed, but its performance with a relatively low R-squared value (25.92%) and higher Root Mean Square Error (RMSE) (2.92) did not meet expectations, indicating the need for an alternative. Least Absolute Shrinkage and Selection Operator (LASSO) regression was then utilized, demonstrating higher accuracy and predictive performance, making it the preferred final model. The results revealed 14 significant predictor variables in the LASSO model, including water sources (public water, protected dug well water, and spring water), alcohol consumption, systolic blood pressure, baseline eGFR, weight, serum sodium and calcium levels, as well as health symptoms such as joint pain, dehydration, nausea, and fever. With an R-squared value of 65.77% and a reduced RMSE of 1.82, the LASSO model performed more predictably than Ridge Regression. The actual vs. predicted graph further highlighted the model's precision in estimating the annual eGFR decline for CKDu patients. The predictive model can be used to estimating the annual eGFR decline of CKDu patients based on identified factors and it has practical implications for healthcare professionals in understanding disease progression and making informed decisions regarding personalized interventions.

Keywords: Chronic Kidney Disease, Annual eGFR Decline. Ridge Regression, LASSO Model

Acknowledgement: Financial assistance from University Research Grant (URG2021/38/S) is acknowledged.