International Journal of Advances and Applications in Convergent Studies & Transdisciplinary (IJ-AACST)

Bridging Disciplines, Bridging Knowledge

Aligned with the Vision of NEP 2020 and UGC Guidelines

**Website**: https://aacstij.org || **Email:** journaleditor@aacstij.org

# Tree and Regression based Variants for Early Kidney Disease Prediction using Machine Learning: Performance and Interpretability

**DOI:** https://doi.org/10.5281/zenodo.22339351

**Nita Dakhare[1]*, Dr. Nilay Karade[2]**

*Corresponding Author: [1]Research Scholar, School of Allied Sciences, Datta Meghe Institute of Higher Education and Research (Deemed to be University), Sawangi (Meghe), Wardha, Maharashtra, India.
*nitamalekar29@gmail.com*

ORCID: https://orcid.org/0009-0008-0084-8008

[2]Professor, School of AI and Future Technology, Universal AI University, Karjat, Mumbai, Maharashtra, India.
*nilay.karade@universalai.in*

## ABSTRACT

It is very important to detect chronic kidney disease (CKD) early in order to prevent the progression to end-stage renal failure. However, the conventional markers used for diagnosis typically identify the loss of function only after a large amount of damage has been done to the kidneys. In this work, a comparative review is given of the recent Machine Learning (ML) approaches for early CKD prediction, with a particular focus on the tree-based and regression-based variants. Six representative studies published between 2020 and 2023 were analysed according to dataset characteristics, preprocessing methods, model architectures, evaluation metrics and interpretability strategies. The results show that Random Forest is the method that often yield the highest predictive performance on the UCI CKD dataset, with accuracy ranging from 88% to 89.5%. Nevertheless, some issues such as the small dataset size, class imbalance, lack of model explainability, and limited external validation still pose major challenges. The review brings together the strengths and limitations of the current methods and emphasizes the need for developing future CKD prediction models that combine powerful feature selection, balanced learning, and explainability techniques to be used in clinical decision-making.

**KEYWORDS**: CKD, ML, Prediction Models, Decision Trees, Random Forest, Logistic Regression.