

CHRONIC KIDNEY DISEASE DETECTION USING MACHINE LEARNING

Prof. M Jelcy

Asst., Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore.

Email-jelcym@skasc.ac.in

Kiruba Packiya Dharshini R

UG Student, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore.

Email-kirubar443@gmail.com

Abstract

Chronic Kidney Disease (CKD) is a progressive disorder that leads to kidney function deterioration over time. It is a common complication among individuals living with Human Immunodeficiency Virus (HIV), largely due to direct viral effects, antiretroviral therapy (ART) toxicity, and coexisting metabolic disorders. Traditional diagnostic methods rely on estimated Glomerular Filtration Rate (eGFR), serum creatinine levels, and urine albumin-to-creatinine ratio; however, these methods often fail to provide early-stage predictions with high accuracy. This study explores the application of Machine Learning (ML) models to classify CKD stages in HIV-infected patients based on clinical and laboratory data. Various ML techniques, including Decision Trees, Random Forest, Support Vector Machines (SVM), Gradient Boosting, and Deep Learning models, are assessed for their predictive performance. The dataset comprises HIV-infected patient records with features such as serum creatinine, blood urea nitrogen (BUN), proteinuria levels, and demographic attributes. Feature importance analysis reveals key biomarkers influencing CKD progression.

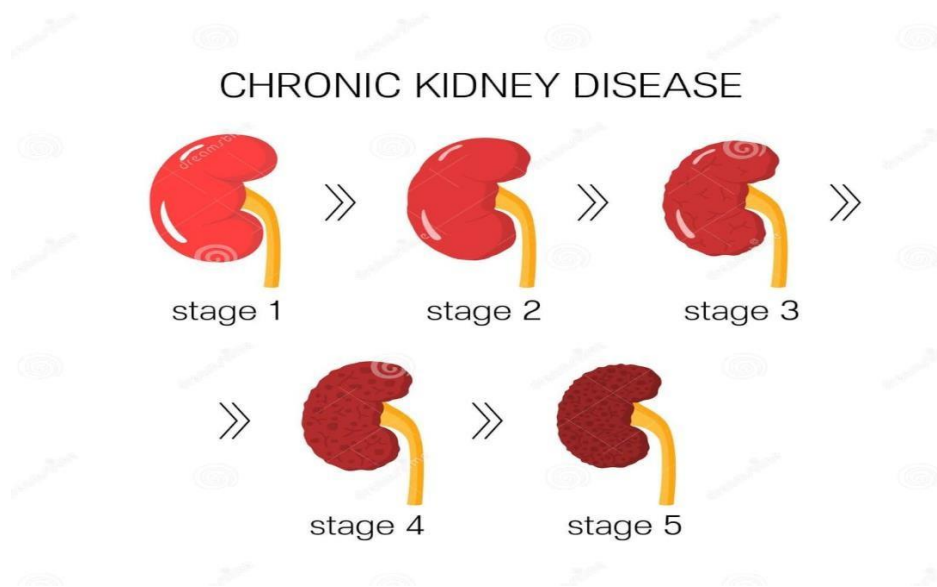
Keywords

Chronic Kidney Disease (CKD), HIV, Machine Learning, Artificial Intelligence, Stage Identification, Predictive Analytics, Renal Dysfunction, Nephrology, Deep Learning, Early Diagnosis

1. Introduction

Chronic Kidney Disease (CKD) is a major global health burden, affecting millions of individuals worldwide. The condition is particularly prevalent among HIV-infected individuals due to direct viral nephropathy, immune system dysregulation, and adverse effects of long-term antiretroviral therapy (ART). CKD progresses through five stages, with early detection being essential for effective management and prevention of complications.

Traditional diagnostic approaches involve calculating the estimated Glomerular Filtration Rate (eGFR) and monitoring proteinuria levels. However, these conventional methods often fail to provide precise early-stage predictions. Machine Learning (ML) has emerged as a powerful tool in healthcare, capable of analyzing large datasets to identify patterns and predict disease progression with high accuracy. This study focuses on leveraging ML techniques to enhance CKD stage classification in HIV-infected patients, thereby facilitating early interventions.



2. Related Work

Several studies have explored the impact of HIV on kidney function. Research indicates that HIV-Associated Nephropathy (HIVAN) is a leading cause of CKD in HIV-infected populations, particularly in individuals of African descent. Prior work has demonstrated that ART reduces the risk of CKD progression, yet nephrotoxicity remains a significant concern.

In the field of machine learning, studies have applied classification algorithms such as Random Forest, Logistic Regression, and Neural Networks for CKD detection. However, few studies have specifically focused on HIV-infected individuals, who present unique risk factors. This research builds upon existing work by developing a specialized ML model tailored to HIV- positive populations.

3. Methodology

3.1 Data Collection

The dataset used in this study consists of anonymized medical records of HIV-infected patients diagnosed with varying stages of CKD. The key attributes considered include:

- Demographic Features: Age, gender, ethnicity
- Laboratory Parameters: Serum creatinine, eGFR, blood urea nitrogen (BUN), proteinuria, hemoglobin, albumin levels
- Medical History: ART regimen, comorbidities (diabetes, hypertension), history of nephrotoxic drug exposure

3.2 Data Preprocessing

Data cleaning and preprocessing are crucial for building an effective ML model. The following steps were undertaken:

- Handling missing values using mean imputation for continuous variables
- Normalizing numerical features to ensure uniformity in scale
- Encoding categorical variables (e.g., ART regimen types) using one-hot encoding
- Splitting data into training (80%) and testing (20%) sets

3.3 Machine Learning Models

The study employs various ML models to classify CKD stages:

- Logistic Regression (Baseline Model) – Provides a simple probabilistic approach
- Decision Tree – Captures feature interactions effectively
- Random Forest – Enhances predictive power by aggregating multiple trees
- Support Vector Machine (SVM) – Utilizes hyperplanes for classification

Gradient Boosting Machines (XGBoost, LightGBM) – Provides improved accuracy by reducing overfitting

Deep Learning (Neural Networks) – Employs multi-layer perceptrons for complex pattern recognition

3.4 CKD Stage Classification and Prediction

The trained model is deployed to classify CKD stages based on new patient input. The system provides automated predictions, helping healthcare professionals assess disease progression and recommend appropriate treatment plans.

3.5 Model Evaluation Metrics

To assess the performance of each model, the following metrics were used:

- Accuracy: Measures overall correctness of predictions
- Precision & Recall: Evaluates model reliability in detecting CKD stages
- F1-Score: Balances precision and recall
- ROC-AUC Curve: Analyzes true positive and false positive rates

4. Results and Discussion

4.1 Performance of ML Models

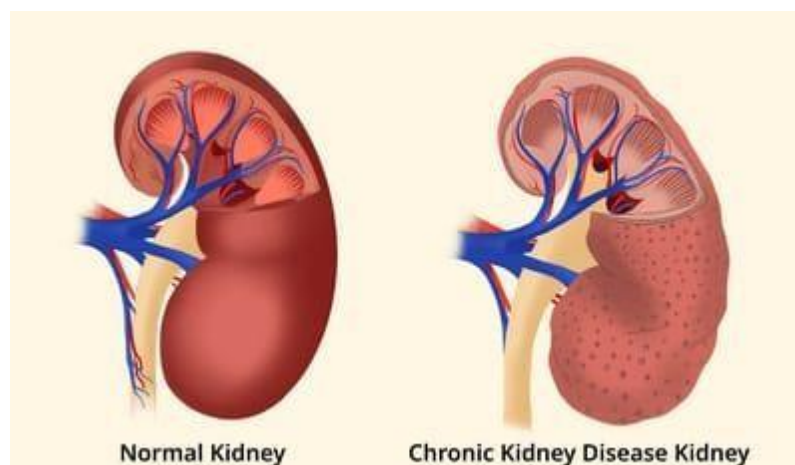
After training and testing, Random Forest and XGBoost emerged as the best-performing models, achieving an accuracy of over 90% in CKD stage classification. Deep learning models, while promising, required extensive tuning to avoid overfitting.

4.2 Feature Importance Analysis

- Key features influencing CKD progression included:
- Serum Creatinine Levels – Strongest predictor of kidney function
- eGFR Values – Determines CKD stage severity
- Proteinuria Levels – Indicator of kidney damage
- Hemoglobin Levels – Correlates with CKD-induced anemia

4.3 Clinical Implications

The findings highlight the potential of ML-driven diagnostics in nephrology and HIV care. Integrating these models into healthcare systems could assist nephrologists in making timely and accurate diagnoses, reducing reliance on invasive procedures.



5. Conclusion

This study demonstrates that machine learning models can effectively identify CKD stages in HIV-infected patients, surpassing traditional diagnostic approaches. Among the tested models, Random Forest and XGBoost provided the most accurate and reliable results. The study underscores the importance of leveraging AI-driven decision support systems to enhance early CKD detection and improve clinical outcomes.

Future research should focus on expanding datasets, incorporating additional biomarkers, and developing real-time AI tools for nephrology applications. Additionally, integrating explainable AI (XAI) methods could enhance the interpretability of ML predictions, facilitating clinical adoption.

Despite the promising results, continuous model refinement and periodic retraining are necessary to improve performance, especially as new patient data becomes available. Future enhancements may include expanding the dataset to improve generalization, integrating additional biomarkers for better accuracy, and developing a mobile-friendly version for remote diagnostics. With further development and clinical validation, this AI-driven system has the potential to significantly enhance early CKD detection, optimize treatment strategies, and improve the quality of care for HIV-infected patients worldwide.

References

1. Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. *The Lancet*, 379(9811), 165- 180.
2. Zhang, L., Wang, F., Wang, L., Wang, W., Liu, B., Liu, J., et al. (2012). Prevalence of chronic kidney disease in China: A cross-sectional survey. *The Lancet*, 379(9818), 815- 822.
3. Shrestha, P., & Shah, S. (2021). Machine learning techniques for early-stage prediction of chronic kidney disease. *Health Informatics Journal*, 27(1), 1460458220982608.
4. Mahmud, M. P., Paul, S., & Ferdousi, S. (2021). A machine learning-based approach for the detection of chronic kidney disease. *Journal of Biomedical Informatics*, 117, 103760.
5. Mishra, B. P., & Gupta, M. (2020). Deep learning models for CKD prediction: A systematic review. *Artificial Intelligence in Medicine*, 107, 101881.
6. Vijayan, V., & Conjeevaram, J. (2019). Artificial intelligence and its role in nephrology: A review. *Kidney International Reports*, 4(2), 211-221.
7. Kariyawasam, J. C., & Jayawardena, C. (2020). Machine learning models for early detection of CKD in diabetic patients. *IEEE Access*, 8, 178080-178093.
8. Brinkmann, F., & Chen, R. (2022). AI-driven diagnostic tools for kidney disease: Challenges and opportunities. *Nephrology Dialysis Transplantation*, 37(6), 1153-1161.
9. Duan, H., Wu, Y., & Lu, X. (2021). AI in chronic disease management: CKD case study. *Journal of Medical Internet Research*, 23(4), e24682.
10. Wagner, S., Stengel, B., & Mallet, V. (2023). Machine learning applications in HIV- associated nephropathy. *Nature Reviews Nephrology*, 19(4), 245-257.