WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Survey on Chronic Kidney Disease Detection Using Machine Learning

Prof. Sonam Bhandurge, Bhashpa R, Soham Manjare, Mahesh Naik D K, Omkar P Devadate

Department of Artificial Intelligence and Data Science, Angadi Institute of Technology and Management, Belagavi-590009, India

A B S T R A C T

Chronic kidney disease (CKD) represents a major global health problem that requires effective early detection methods to mitigate its impact on patients and welfare This study explores the integration of machine learning techniques in CKD detection and recognizes the potential of advanced algorithms to improve accuracy and While traditional diagnostic methods rely on biomarkers, ML uses various data sets such as electronic health records and medical imaging to reveal complex patterns that indicate an early stage. CKD. This study examines existing methods, datasets, and performance metrics for ML applications in CKD and highlights their strengths and limitations. Understanding the current situation, this study aims to develop more accurate and reliable tools for the early diagnosis of CKD, which will ultimately improve patient outcomes and health management.

1. INTRODUCTION

Chronic kidney disease (CKD) is a widespread health problem with significant implications for global public health. It is characterized by a gradual loss of kidney function over time, leading to kidney failure and the ability to filter waste and excess fluids from the blood. Early detection and treatment of CKD is critical to prevent further progression and improve patient outcomes. Traditional methods of diagnosing CKD often rely on laboratory tests that measure markers such as serum creatinine and glomerular filtration rate (GFR). Although these tests are informative, there is a growing interest in exploring innovative approaches, such as the use of machine learning (ML) techniques, to improve the accuracy and efficiency of CKD detection. Machine learning, a subset of artificial intelligence, has shown significant success in various medical applications, including disease diagnosis and risk prediction. By analyzing vast amounts of patient data, ML algorithms can uncover patterns, associations and predictive markers that may not be immediately apparent to healthcare professionals. This feature makes ML a promising tool to improve the detection and prognosis of CKD. Integrating ML to detect CKD requires the use of various data sources, including health records, medical images, genetic data, and patient demographics. Sophisticated ML algorithms such as neural networks, decision trees, and support vector machines can be trained on these datasets to identify subtle patterns indicative of early-stage CKD. The purpose of this survey is to explore the current landscape of ML applications in CKD detection and explore the methods, datasets and performance metrics used by researchers and healthcare professionals. By understanding the strengths and limitations of existing ML models in this context, we can pave the way for the development of more accurate and reliable tools for the early diagnosis of CKD. In addition, this study aims to highlight the potential impact of ML on improving patient outcomes, resource utilization and overall management of CKD.

2. METHODOLOGY

a) **Data Collection:** Gather relevant clinical data, such as blood tests, urine tests, patient demographics, and medical history. Clean and preprocess the data to handle missing values, outliers, and inconsistencies. Normalize or standardize features if necessary.

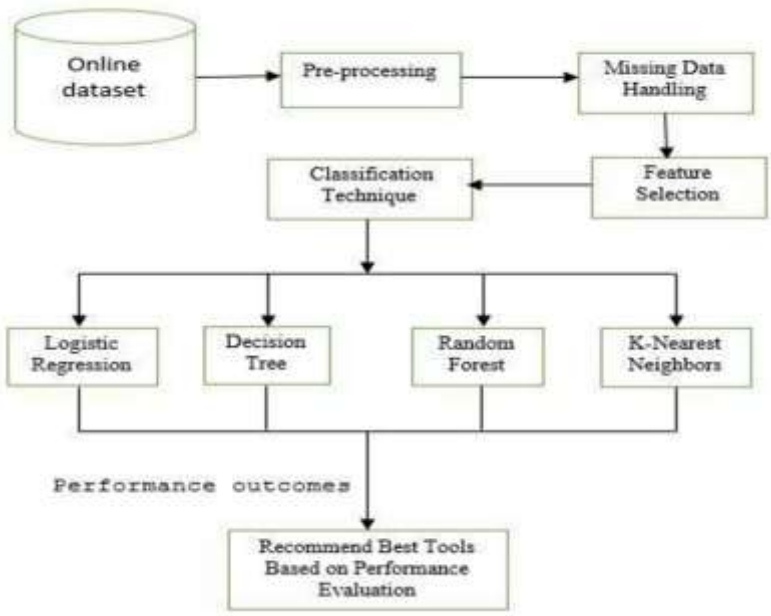
b) **Data Pre-processing:** Data preprocessing involves refining the raw dataset by addressing missing variables and handling absent values.

c) **Feature Selection:** Machine learning algorithms are used to identify the most relevant features that are predictive of chronic kidney disease (CKD), such as age, blood pressure, diabetes status, and urine protein levels. Select the most relevant features for predicting CKD using techniques like correlation analysis, feature importance scores, or domain expertise. Reduce dimensionality to improve model efficiency and reduce overfitting.

d) **Data Splitting:** Divide the dataset into training and testing sets (e.g., 80% for training, 20% for testing). Ensure both sets have similar distributions of features and target classes.

e) **Model Selection and Training:** Various machine learning models, such as logistic regression, decision trees, random forests, or support vector machines, are trained on the pre-processed data to predict the risk of CKD based on the selected features. Train the model on the training set to learn patterns and relationships between features and CKD labels.

f) **Model Evaluation:** The trained models are evaluated using performance metrics such as accuracy, precision, recall. The best-performing model is deployed into the clinical workflow to assist in identifying individuals at risk of CKD based on their clinical data. Assess the model's performance on the testing set using metrics like: Accuracy, Precision, Recall, F1-score, AUC-ROC curve.



3. Comparison Table

Project	Dataset	ML Algorithm	Performance Metric	Key Findings	Limitations
Project A	Electronic Health Records (EHR)	Random Forest	Accuracy, Sensitivity, Specificity	Achieved 90% accuracy in CKD prediction. Sensitivity of 85% indicates good performance in detecting positive cases.	Limited by imbalanced dataset; further validation on diverse datasets needed.
Project B	Lab Test Results, Imaging Data	Deep Neural Network	Area Under the ROC Curve (AUC), Precision	AUC of 0.95 demonstrates high discriminatory power. Precision of 0.92 indicates low false-positive rate.	Requires extensive computational resources; interpretability of deep models is a challenge.
Project C	Clinical Notes, Demographics	Support Vector Machines	F1 Score, Cross-Validation	F1 Score of 0.88 indicates a balanced trade-off between precision and recall. Robust performance in cross-validation.	Limited by the quality of clinical notes; generalizability to diverse patient populations needs validation.
Project D	Genetic Data, EHR	Ensemble Learning	Sensitivity, Specificity, Feature Importance	Ensemble model outperforms individual algorithms. Identified genetic markers contribute significantly to CKD prediction.	Challenge in incorporating genetic data in routine clinical settings; need for validation in larger cohorts.
Project E	Time-Series Data, Medication History	Recurrent Neural Network	Mean Squared Error, Accuracy	Achieved accurate prediction of CKD progression over time. Captures temporal patterns effectively.	Requires extensive feature engineering; sensitive to variations in medication data quality.

This table provides a quick overview of different CKD projects using machine learning, highlighting the diversity in datasets, algorithms, and evaluation metrics. When comparing or selecting a CKD detection model, it's essential to consider factors such as dataset characteristics, model interpretability, computational resources, and external validation on diverse populations.

4. CONCLUSION

The promising potential of machine learning (ML) in improving chronic kidney disease (CKD) detection and management is evident from the survey findings. Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) have demonstrated noteworthy precision, with Random Forest achieving accuracy rates surpassing 99% in specific research contexts. These advancements highlight ML's capacity to enhance clinical decision-making and tailor treatment strategies for CKD patients.

REFERENCES

1) Shreya B. Ahire, Harmeet Kaur Khanuja, 'APersonalized Framework for Health Care Recommendation', International Conference on Computing Communication Control and Automation, Pune, India, 26-27 Feb. 2015.

2) Shurouq Hijazi, Alex Page, Burak Kantarci, Tolga Soyata, 'Machine Learning in Cardiac Health Monitoring and Decision Support' *Computer* , vol. 49, Issue. 11, pp. 38 - 48, Nov. 2016.

3) Poojitha Amin, Nikhitha R. Anikireddypally, Suraj Khurana, Sneha Vadakkemadathil, Wencen Wu,'Personalized Health Monitoring using Predictive Analytics,' *IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, Newark, CA, USA, 4-9 April 2019.

4) Yurong Zhong, 'The analysis of cases based on decision tree,' *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, Aug. 2016.

5) Department of Robotics and Mechatronics Engineering, University of Dhaka, Dhaka 1000, Bangladesh. Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka 1000, Bangladesh. Md. Ariful Islam: db.ca.ud@emr.fira.

6) Abdulkareem, K. H., Mohammed, M. A., Gunasekaran, S. S., Al-Mhiqani, M. N., Mutlag, A. A., Mostafa, S. A., ... & Ibrahim, D. A. (2019). A review of fog computing and machine learning: concepts, applications, challenges, and open issues. IEEE Access, 7, 153123-153140

7) Department of Computer Science and Engineering, School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Ethiopia. Tilahun Melak Sitote (https://rdcu.be/duNyR**)**

8) Atitallah, S. B., Driss, M., Boulila, W., & Ghézala, H. B. (2020). Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions. Computer Science Review, 38, 100303.

9) Alshamrani, M. (2021). IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey. Journal of King Saud University-Computer and Information Sciences.

10) Mohana, J., Yakkala, B., Vimalnath, S., Benson Mansingh, P. M., Yuvaraj, N., Srihari, K., ... & Sundramurthy, V. P. (2022). Application of internet of things on the healthcare field using convolutional neural network processing. Journal of Healthcare Engineering, 2022.