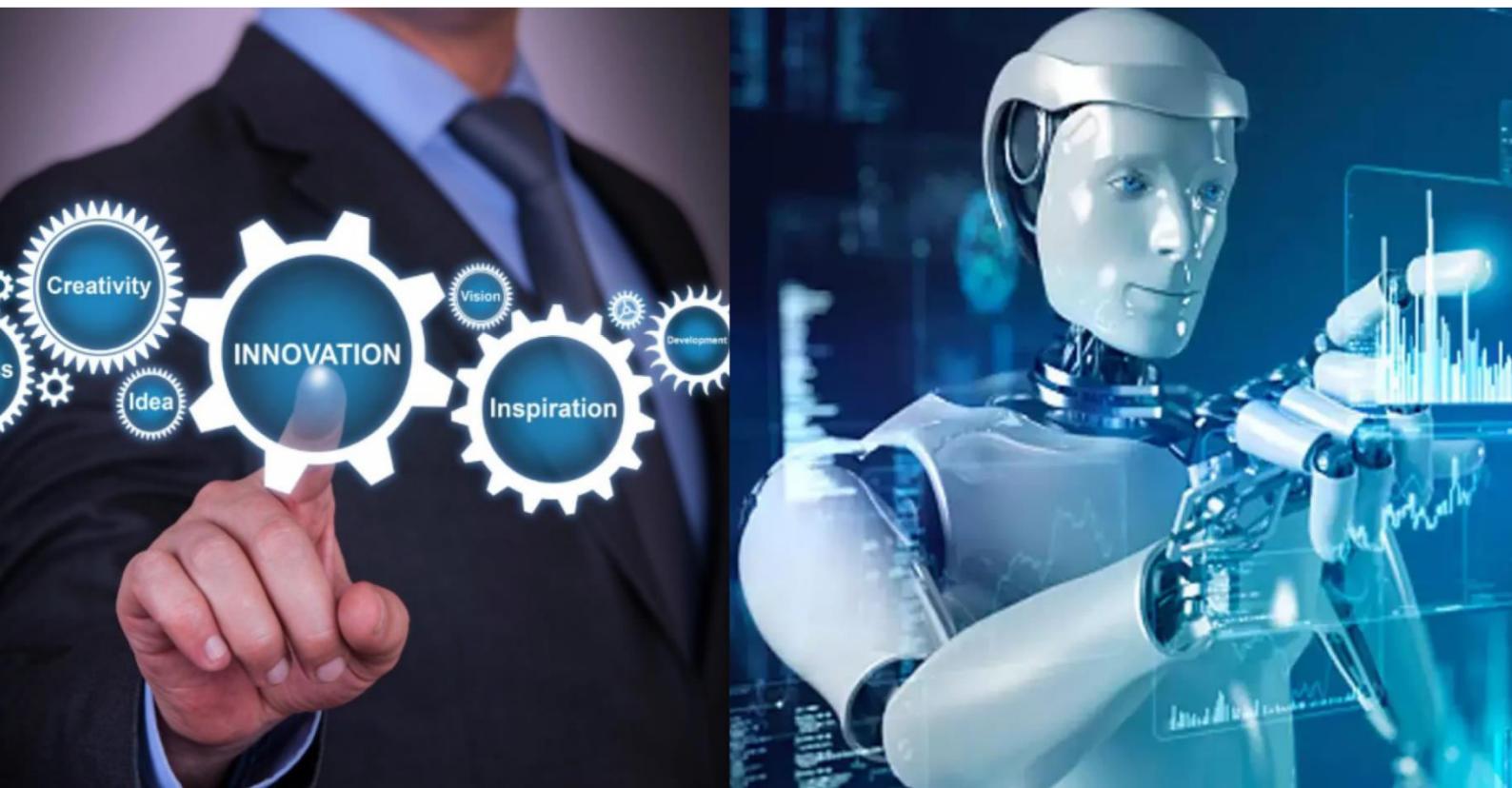




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 12, December 2025

A Machine Learning Methodology for Diagnosing Chronic Kidney Disease: A Review

Dr. Deepti Varshney, Aishwarya Ramesh Patil

Department of Computer Engineering, Ajeenkya Dr D.Y. Patil School of Engineering and Technology Lohegaon, Pune, India

ABSTRACT: Chronic Kidney Disease (CKD) is a serious health condition that often remains undetected in its early stages. This project presents a machine learning-based system to predict CKD using patient medical data. The system analyzes clinical parameters to classify patients accurately and efficiently. Various machine learning algorithms are used to improve prediction performance. This approach enables early diagnosis, reduces manual effort, and supports effective medical decision-making in healthcare systems.

KEYWORDS: Chronic Kidney Disease, Machine Learning, Medical Diagnosis, Classification, Data Mining.

I. INTRODUCTION

Chronic Kidney Disease is a long-term medical condition characterized by gradual loss of kidney function. It affects millions of people worldwide and can lead to severe complications such as cardiovascular disease, kidney failure, and death if not detected early. CKD often progresses without noticeable symptoms, making early diagnosis difficult using traditional clinical methods. Routine screening and laboratory tests generate large amounts of medical data that are underutilized. Hence, there is a strong need for intelligent systems capable of analyzing this data efficiently. Machine learning techniques provide powerful tools to identify hidden patterns in healthcare data. These techniques enable early and accurate prediction of CKD.

Recent advancements in healthcare informatics have encouraged the adoption of machine learning for disease prediction. Machine learning models can learn from historical patient data and assist doctors in making accurate decisions. In CKD diagnosis, multiple clinical parameters such as blood pressure, serum creatinine, hemoglobin, and glucose levels play an important role. However, manual analysis of these factors is time-consuming and error-prone. Automated prediction systems help overcome these limitations by offering faster and consistent results. Such systems can significantly reduce diagnosis time and healthcare costs. Therefore, applying machine learning to CKD prediction is both practical and beneficial.

Despite the availability of various diagnostic methods, CKD detection still faces challenges such as missing data, inconsistent measurements, and limited medical resources. Traditional systems rely heavily on expert interpretation, which may vary among physicians. Machine learning-based systems provide standardized decision support by using statistical and computational approaches. These systems can be integrated into clinical workflows to assist healthcare professionals. With proper preprocessing and feature selection, prediction accuracy can be greatly improved. Hence, this project focuses on designing a reliable CKD prediction system using machine learning techniques. The goal is to enhance early detection and patient care.

Motivation

The primary motivation of this project is the increasing prevalence of Chronic Kidney Disease and the lack of early diagnostic tools. Many patients are diagnosed only at advanced stages due to the absence of clear symptoms. Manual diagnosis requires expert knowledge and is prone to human error. Machine learning offers an automated, accurate, and scalable solution for early CKD detection. By utilizing patient data effectively, the system can assist doctors in making faster and better decisions. The project aims to reduce mortality rates and improve quality of life through early prediction.

II. LITERATURE SURVEY

Mohamed Mahyoub [1] - This paper discusses the importance of investigating Alzheimer's Disease using machine learning, the need to use both behavioural and biological markers data, and a computational method to rank Alzheimer's Disease risk factors by importance using different machine learning models on Alzheimer's Disease clinical assessment data from ADNI.

XIANCHUAN WANG [2] - The final established model, random forest-improved fruit fly optimization algorithm-kernel extreme learning machine, was used to effectively diagnose sepsis. Experimental results demonstrate that the proposed method obtains better results than other methods across four performance metrics. We screened out five biomarkers and performed statistical analysis on these five substances.

ZHENTAO GAO [3] - Diabetic retinopathy (DR) is a common eye disease and a significant cause of blindness in diabetic patients. Regular screening with fundus photography and timely intervention is the most effective way to manage the disease. The large population of diabetic patients and their massive screening requirements have generated interest in computer-aided and fully automatic diagnosis of DR. Deep neural networks, on the other hand, have brought many breakthroughs in various tasks in the recent years.

QuanZou [4] - In this study, we used decision tree, random forest and neural network to predict diabetes mellitus. The dataset is the hospital physical examination data in Luzhou, China. It contains 14 attributes. In this study, five-fold cross validation was used to examine the models. In order to verify the universal applicability of the methods, we chose some methods that have the better performance to conduct independent test experiments.

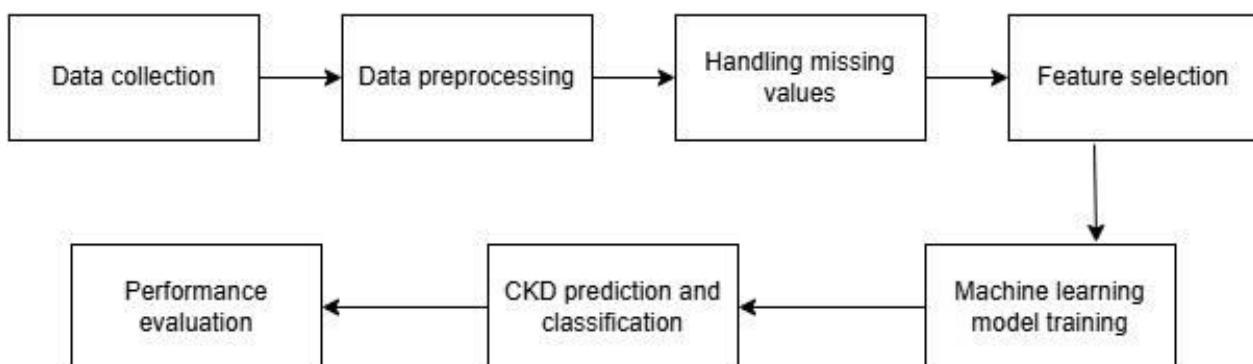
Miguel Patrício [5] , The goal of this exploratory study was to develop and assess a prediction model which can potentially be used as a biomarker of breast cancer, based on anthropometric data and parameters which can be gathered in routine blood analysis.

NamyongPark [6] - In this paper, we provide an AKI prediction model with a greater applicability, which relaxes the constraints of existing approaches, and fully utilizes irregular and heterogeneous data for learning the model. In a cohort of 21,022 cancer patients who were registered into Korea Central Cancer Registry (KCCR) in Seoul National University Hospital between January 1, 2004 and December 31, 2013, our method achieves 0.7892 precision, 0.7506 recall, and 0.7576 F-measure in predicting whether a patient will develop AKI during the next 14 days.

Gap Analysis

- Traditional CKD diagnosis is time-consuming and depends on expert interpretation
- Existing systems struggle with missing and inconsistent medical data
- Limited use of feature selection affects prediction accuracy
- Lack of automated decision support systems in clinical practice
- Need for cost-effective and scalable diagnostic solutions

III. SYSTEM ARCHITECTURE



IV. PROPOSED SYSTEM

The proposed system uses machine learning algorithms to predict Chronic Kidney Disease at an early stage. Medical data is first preprocessed to remove noise and handle missing values. Important features are selected to improve prediction accuracy. Multiple classifiers are trained and evaluated to determine the best performing model. The system automatically classifies patients as CKD or non-CKD based on input parameters. This approach reduces manual effort, increases accuracy, and provides fast diagnostic support. The system can be easily deployed in healthcare environments.

V. CONCLUSION

This project demonstrates the effectiveness of machine learning techniques in predicting Chronic Kidney Disease. By analyzing clinical data, the system accurately classifies patients and supports early diagnosis. Proper data preprocessing and feature selection significantly enhance model performance. The proposed system reduces dependency on manual diagnosis and helps healthcare professionals make informed decisions. Overall, the system provides a reliable, efficient, and cost-effective solution for CKD prediction. It contributes to improved patient outcomes and better healthcare management.

VI. FUTURE SCOPE

- Integration with real-time hospital databases
- Use of deep learning models for improved accuracy
- Prediction of CKD stages and severity levels
- Mobile and web-based healthcare applications
- Expansion to other chronic disease prediction systems

REFERENCES

1. M. Mahyoub, M. Randles, T. Baker, and P. Yang, "Comparison analysis of machine learning algorithms to rank alzheimer's disease risk factors by importance," in Proc. 11th Int. Conf. Develop. eSyst. Eng. (DeSE), Sep. 2018, pp. 1–11.
2. Q. Zou, "Predicting diabetes mellitus with machine learning techniques," *Frontiers Genet.*, vol. 9, p. 515, Nov. 2018.
3. Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, and J. Zhong, "Diagnosis of diabetic retinopathy using deep neural networks," *IEEE Access*, vol. 7, pp. 3360–3370, 2019.
4. N. Park, E. Kang, M. Park, H. Lee, H.-G. Kang, H.-J. Yoon, and U. Kang, "Predicting acute kidney injury in cancer patients using heterogeneous and irregular data," *PLoS ONE*, vol. 13, no. 7, Jul. 2018, Art. no. e0199839.
5. M. Patricio, "Using Resistin, glucose, age and BMI to predict the presence of breast cancer," *BMC Cancer*, vol. 18, p. 29, Jan. 2018.
6. X. Wang, Z. Wang, J. Weng, C. Wen, H. Chen, and X. Wang, "A new effective machine learning framework for sepsis diagnosis," *IEEE Access*, vol. 6, pp. 48300–48310, 2018.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com



Scan to save the contact details