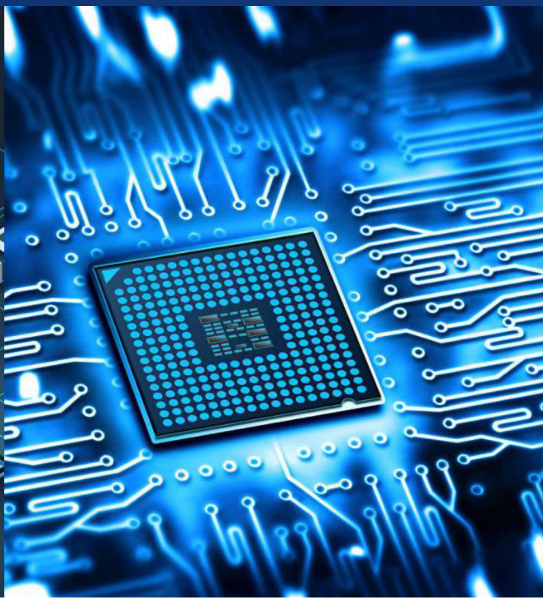
IJIRCCE

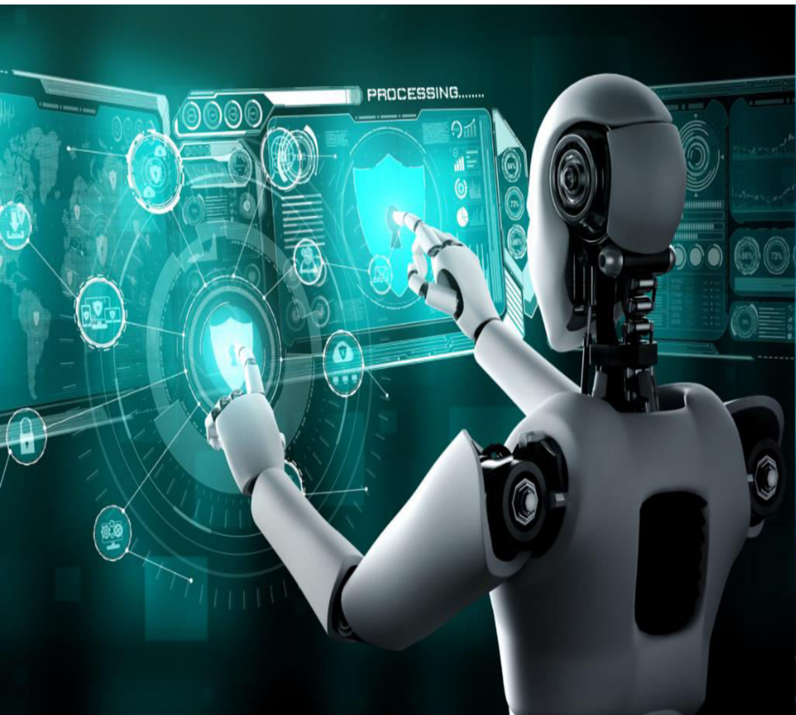
ISSN(O): 2320-9801

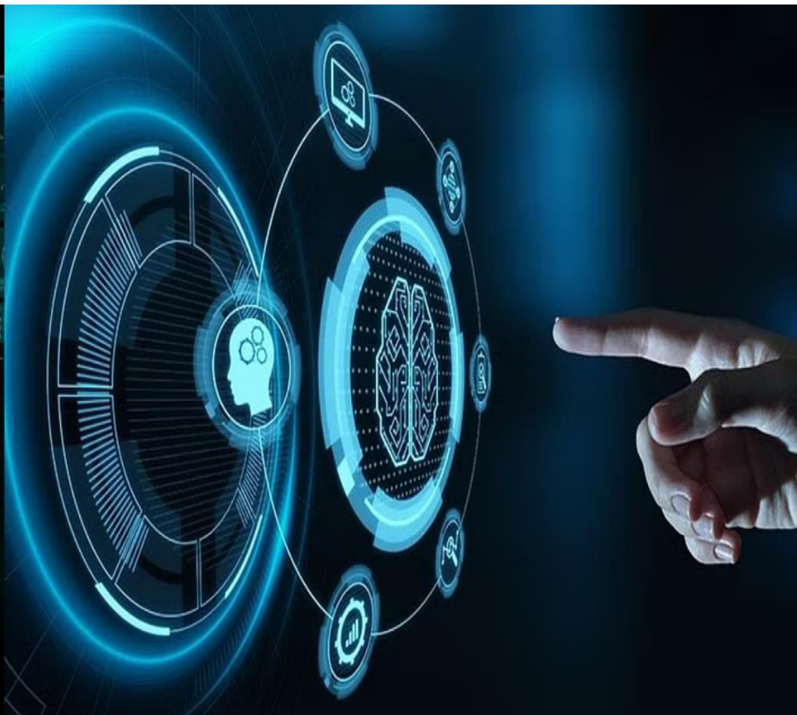
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

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





Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com | ijircce@gmail.com | +91-9940572462 | +91 63819 07438 www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Chronic Kidney Diseases Prediction using Machine Learning Techniques

Priyanka K, Dr. C. Anbarasa

III- B.Sc., Department of Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College, Coimbatore, India

Professor, Department of Computer Science with Data Analytics, Dr. N.G.P Arts and Science College, Coimbatore, India

ABSTRACT: Chronic Kidney Disease (CKD) is a growing global health concern, leading to high mortality rates and significant healthcare costs. Early detection and prediction of CKD can improve patient outcomes and reduce healthcare burdens. This study explores the use of machine learning (ML) techniques to predict the onset and progression of CKD. A variety of ML algorithms, including decision trees, support vector machines, random forests, and neural networks, were evaluated using patient data such as demographic information, medical history, and laboratory test results. The dataset was pre-processed to handle missing values, normalize features, and perform feature selection. The models were trained and tested on a dataset of CKD patients, and their predictive performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. Results demonstrate that machine learning approaches can offer a reliable and efficient method for predicting CKD, with certain models outperforming others in terms of prediction accuracy. This study emphasizes the potential of ML in clinical decision-making and highlights the importance of early diagnosis for better management of CKD, ultimately improving patient care and reducing the burden on healthcare systems.

KEYWORDS: Chronic Kidney Disease (CKD), Machine Learning (ML), Early Detection, Prediction Models, Decision Trees, Support Vector Machines (SVM), Random Forest.

I. INTRODUCTION

Chronic kidney disease (CKD) is a global public health problem affecting approximately10% of the world's population. The percentage of prevalence of CKD in China is 10.8%, and the range of prevalence is 10%-15% in the United States. This disease is characterized by a slow deterioration in renal function, which eventually causes a complete loss of renal function. CKD does not show obvious symptoms in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has high morbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. CKD is a progressive and irreversible pathologic syndrome. Hence, the prediction and diagnosis of CKD in its early stages is quite essential, it may be able to enable patients to receive timely treatment to ameliorate the progression of the disease.Kidney failure treatment targets to control the causes and decelerate the advance of therenal failure. If treatments are not enough, patient will be in the end-stage of renal failure and the last treatment is dialysis or renal transplant. At present, 4 out of every 1000 person in the United Kingdom are suffering from renal failure and more than 300,000 American patients in the end-stage of kidney disease survive with dialysis. Moreover, according to the National Health Service kidney disease is more frequent in South Asia, Africa, than in the other countries. Due to detecting the chronic kidney failure is not feasible until the kidney failure is completely progressed; thus, realizing the kidney failure in the first stage is extremely important. Through early diagnosis, the act of each kidney can be taken under control, which leads to decreasing the risk of irreversible consequences. For this reason, routine check-up and early diagnosis are crucial to the patients, for they can prevent vital risks of renal failure and related diseases. Blood test is one of the steps to detect CKD.

II. RELATED WORK

This study explored the application of machine learning algorithms, such as decision trees, support vector machines (SVM), and random forests, to predict CKD from a dataset containing clinical attributes like age, blood pressure,

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304055

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

specific gravity, and albumin levels. The authors found that SVM outperformed other models in terms of prediction accuracy, showing its potential for early CKD detection. The study highlighted the importance of data preprocessing and feature selection in improving model performance. This work underscores the utility of machine learning in clinical settings to identify at-risk patients earlier than traditional methods.

This paper compared several classification algorithms, including logistic regression, SVM, and artificial neural networks, to predict CKD. The researchers used a dataset containing demographic information, medical history, and laboratory data. The study showed that neural networks yielded the highest accuracy and provided insights into the relationships between different factors contributing to CKD. The paper also emphasized the need for handling imbalanced datasets through techniques like oversampling to improve the performance of machine learning models.

The reviewed studies demonstrate that machine learning techniques, particularly ensemble methods, neural networks, and support vector machines, show promise for predicting chronic kidney disease. Feature selection and data preprocessing are critical in enhancing the accuracy and reliability of these models.The combination of clinical data with demographic and medical history is often emphasized, as it can provide a more holistic view of a patient's health, leading to better prediction outcomes. Moreover, the integration of advanced deep learning models and hybrid systems is paving the way for more sophisticated and accurate CKD prediction systems. These approaches hold great potential for assisting clinicians in early diagnosis and treatment planning for CKD patients, ultimately improving patient outcomes.

III. METHODOLOGY

In this study, the K-Nearest Neighbour classification algorithm and Naïve Bayes classifier algorithm were used to diagnose Chronic Kidney Disease. To diagnose Chronic Kidney Disease, two essential types of feature selection methods namely, wrapper and filter approaches were chosen to reduce the dimension of the Chronic Kidney Disease dataset. After selecting features from our dataset, we used a variety of machine learning models to determine the best classification models.

This paper investigates how CKD can be diagnosed by using machine learning (ML)techniques. ML algorithms have been a driving force in detection of abnormalities in different physiological data, and are, with a great success, employed in different classification tasks. In the present study, a number of different ML classifiers are experimentally validated to a real data set, taken from the Machine Learning Repository, and our findings are compared with the findings reported in the recent literature.

The results are quantitatively and qualitatively discussed and our findings reveal that the Hybrid algorithm Named as Bagged tree classifier and Random Forest classifier achieves the near-optimal performances on the identification of CKD subjects.

For this project, our primary packages are going to be Pandas to work with data, NumPy to work with arrays, scikit-learn for data split, building and evaluating the classification models. Let's import all of our primary packages into our python environment.

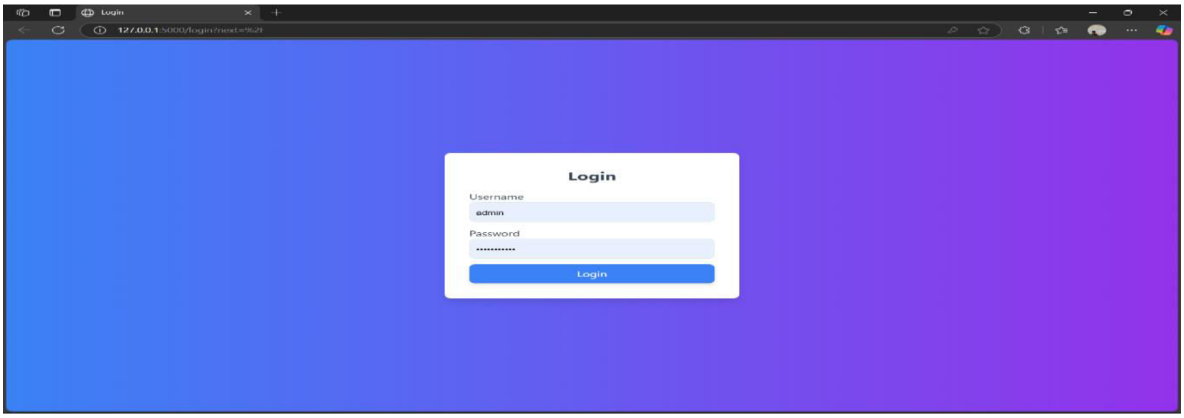
IV. EXPERIMENTAL RESULT

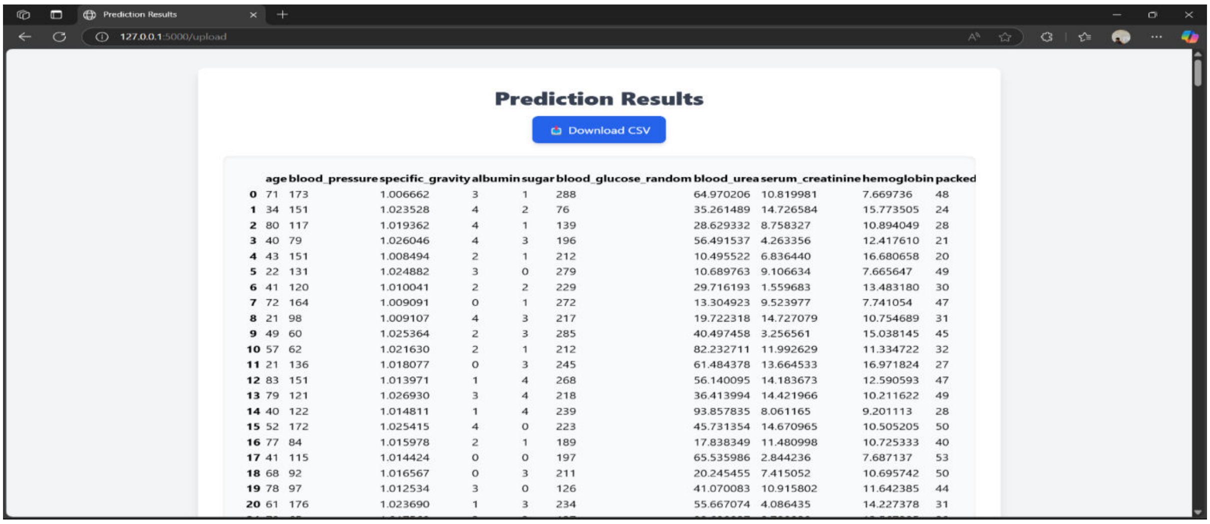
This testing involves evaluating whether the model's predictions align with medical standards and whether healthcare professionals can interpret and trust the results. It includes functional validation, ensuring that the model correctly classifies CKD and non-CKD cases, and usability testing to confirm that the interface and outputs are clear and actionable for doctors and clinicians. Performance benchmarks such as accuracy, precision, and recall are assessed to verify the model's reliability. Additionally, real patient data or expert reviews may be used to validate the system's effectiveness in clinical scenarios.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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





	age	blood_pressure	specific_gravity	albumin	sugar	blood_glucose_random	blood_urea	serum_creatinine	hemoglobin	packed
0	71	173	1.006662	3	1	288	64.970206	10.819981	7.669736	48
1	34	151	1.023528	4	2	76	35.261489	14.726584	15.773505	24
2	80	117	1.019362	4	1	139	28.629332	8.758327	10.894049	28
3	40	79	1.026046	4	3	196	56.491537	4.263356	12.417610	21
4	43	151	1.008494	2	1	212	10.495522	6.836440	16.680658	20
5	22	131	1.024882	3	0	279	10.689763	9.106634	7.665647	49
6	41	120	1.010041	2	2	229	29.716193	1.559683	13.483180	30
7	72	164	1.009091	0	1	272	13.304923	9.523977	7.741054	47
8	21	98	1.009107	4	3	217	19.722318	14.727079	10.754689	31
9	49	60	1.025364	2	3	285	40.497458	3.256561	15.038145	45
10	57	62	1.021630	2	1	212	82.232711	11.992629	11.334722	32
11	21	136	1.018077	0	3	245	61.484378	13.664533	16.971824	27
12	83	151	1.013971	1	4	268	56.140095	14.183673	12.590593	47
13	79	121	1.026930	3	4	218	36.413994	14.421966	10.211622	49
14	40	122	1.014811	1	4	239	93.857835	8.061165	9.201113	28
15	52	172	1.025415	4	0	223	45.731354	14.670965	10.505205	50
16	77	84	1.015978	2	1	189	17.838349	11.480998	10.725333	40
17	41	115	1.014424	0	0	197	65.535986	2.844236	7.687137	53
18	68	92	1.016567	0	3	211	20.245455	7.415052	10.695742	50
19	78	97	1.012534	3	0	126	41.070083	10.915802	11.642385	44
20	61	176	1.023690	1	3	234	55.667074	4.086435	14.227378	31

V. CONCLUSION

The CKD diagnosis is a difficult challenge. In our literature, we have presented a predictive model using different machine learning algorithms, including NN, RF, SVM, RT and BTM, to predict CKD earlier. We mainly focus on the empirical comparisons of those mentioned ML algorithms. Base on the empirical results of the applied algorithm, the NN, RF and SVM have given the highest accuracy on the full dataset, and Random Forest Classifier has given the highest performance on the dataset.

REFERENCES

1. M. M. Hossain, R. K. Detwiler, E. H. Chang, M. C. Caughey, M. W. Fisher,T. C. Nichols,E. P. Merricks, R. A. Raymer, M. Whitford, D. A. Bellinger,L. E. Wimsey, and C. M. Gallippi,"Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement:Preclinical validation and pilot *in vivo* clinical results in kidney allografts," *IEEE Trans.Ultrason.,Ferroelectr., Freq. Control* , vol. 66, no. 3, pp. 551–562, Mar. 2019
2. H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on supportvector machine by feature selection methods,"*J. Med. Syst.*, vol. 41, no. 4, p. 55, Apr. 2017.
3. N. R. Hill, "Global prevalence of chronic kidney disease - A systematic review and metaanalysis," *PLoS ONE* , vol. 11, no. 7, Jul. 2016, Art. no. e0158765.
4. K. M. Z. Hasan and M. Z. Hasan, "Performance evaluation of ensemble based machinelearning techniques for prediction of chronic kidney disease," in Emerging Research in Computing, Information, Communication and Applications. Singapore: Springer, 2019.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com