



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: VI    Month of publication: June 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.44178>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# CKD Stage Prediction and Suitable Diet Recommendation: A Survey

Prof. Girijamba D L<sup>1</sup>, Kavitha N<sup>2</sup>, N Samanvitha<sup>3</sup>, Rohit Patil<sup>4</sup>, Aishwarya Gowda C<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup>Department of Electronics and Communication Engineering , Vidyavardhaka College of Engineering Mysuru, India

**Abstract:** *Chronic kidney disease is one of the world's most complicated illnesses. Renal failure occurs when the kidneys have been afflicted for a long time. Filtration of waste and surplus fluid from the blood is the job of the kidneys. Because waste builds up as kidneys fail, early detection of chronic renal disease and proper treatment are critical in today's medical industry. On the subject of chronic renal disease prediction, there are numerous research studies available. Kidney disease prediction employs a variety of effective approaches and algorithms.*

*A large number of research publications apply data science techniques for renal disease prediction, with promising outcomes. Some studies gathered training datasets from hospitals, which were then trained using efficient machine learning algorithms, yielding positive results.*

*One of the most difficult challenges in the present medical field is identifying this complex kidney illness. Doctors have a difficult time manually identifying the condition.*

*ML algorithms have been used in research to generate appropriate models, and the best model is evaluated to provide better outcomes.*

**Keywords:** *CKD, Naïve Bayes, Support Vector Machine, Logistic Regression, KNN, CNN.*

## I. INTRODUCTION

Chronic Kidney Disease is damages our kidney, this leads to failure of the kidney to filter blood, and failure of kidney. Chronic Kidney Disease is a disease that goes unnoticed.

The signs and symptoms of CKD, if present, are typically non-specific, and unlike many other chronic conditions (such as congestive heart failure and chronic obstructive pulmonary disease), they do not indicate the disease's diagnosis or severity. As you become older, you're more prone to get kidney disease. If you've had diabetes, high blood pressure, or heart disease for a long period, you're more likely to develop kidney disease.

CKD is more common in African Americans, Hispanics, and American Indians. The increased risk is primarily owing to these groups' higher incidence of diabetes and high blood pressure. Other probable causes of this elevated risk are being investigated by scientists. Machine learning and Data Discovery approaches can assist identify CKD risk early in the course of a patient's medical history and diagnostic data.

We want to create a machine learning model that can predict the risk of CKD in a new instance. Several binary classification techniques, such as Logistic Regression and Support Vector Machines, have been tested (SVM). The National Kidney Foundation and other national institutes suggest glomerular filtration rate (GFR) estimations for the definition, categorization, screening, and monitoring of CKD through their Kidney Disease Quality Outcome Initiative (K/DOQI).

The Ma and MacIsaac equations had better bias and accuracy for CKD patients than the MDRD, and the Ma formula's mean findings were closer to mGFR than the other equations in CKD stages 2 to 5. In CKD stages 2–4, the discrepancies between Macisaac and mGFR were much fewer than in CKD stages 1–5. When compared to the MDRD equation, Stevens and Rule's equations showed similar bias and accuracy. When compared to the results in other stages, the MDRD formula demonstrated a higher accuracy in CKD stages 3 - 5.

Another study used electronic health records to obtain only one CKD attribute from year-long temporal data (EHRs). Patients without electronic health records, on the other hand, would be unable to use this strategy. All of the prior studies, however, used the results and selected features from the black-box nature classification approach to model creation, and failing to interpret the diagnostic model conclusion can have bad or even life-threatening consequences. Furthermore, there are insufficient rationales in the current studies for selecting specific traits for model decision-making. As a result, one hot area of research is the creation of interpretable machine learning models for use with computer-assisted diagnostic systems, which would allow doctors to better analyse model decisions and establish the role of specific model attributes.

## II. LITERATURE REVIEW

Machine learning algorithms for CKD prediction were investigated by Bilal Khan, Rashid Khan, and Ghulam Abbas [1]. NBTree, NB, SVM, J48, MLP, and LR are among the seven algorithms employed in addition to CHIRP. The data was taken from the University of California, Irvine Repository. These data suggest that NB, LR, MLP, J48, SVM, and NBTree may be unable to correctly predict CKD, and that CHIRP is a potential technique for CKD prediction based on experimental results. According to the findings of this study, CHIRP is the ideal method for practitioners to follow in order to avoid diagnostic and treatment errors.

Veenita Kunwar and Khushboo Chande [2] used data mining classification techniques to perform CKD analysis. The clinical data for 400 records used in the study was obtained from the UCI Machine Learning Repository. After cleaning and deleting missing values, the data totals 220. Rapid Miner was used to put the data together. The dataset has 25 attributes. Data mining classifiers such as ANN and Naive Bayes have been used to predict and diagnose Chronic Kidney Disease. The Rapidminer tool is used to compare the performance of these methods. The results of the ANN and Naive Bayes classifiers are compared, with Naive Bayes achieving 100% accuracy compared to 72.73 percent for ANN.

Mubarik Ahmed and Peny Amalia [3] used SVM to create a CKD support system. The goal of this system's development is to provide a decision support system for detecting kidney illness and forecasting whether or not patients with renal impairment have progressed to chronic kidney disease. The system was created using the Supported Vector Machine approach. P. Soundrarapandian's 'chronic kidney disease' dataset was downloaded from the UCI repository. For system development, the R programming language is employed.

Nikhila [4] developed a machine learning algorithm. When compared to Bagging and Gradient Boosting, AdaBoost and Random Forest were the best classifiers based on seven performance measures, including Accuracy, Sensitivity, Specificity, F1-Score, and Correlation Coefficient. Gradient Boosting has a 98.33 percent accuracy, whereas Bagging has a 99.166 percent accuracy. AdaBoost and Random Forest are both 100 percent accurate. The datasets were retrieved from the UCI repository, and the model was trained using a 10-fold cross validation procedure with the training set.

Devika R, Sai Vaishnavi Avilala, and V.Subramaniaswamy did a comparative study of classifiers for chronic kidney disease prediction using Naive Bayes, KNN, and Random Forest [5]. We discovered that the Random Forest classifier outperformed the other options, outperforming Naive Bayes in particular (99.635 percent), Random Forest (99.844 percent), and KNN (87.78 percent). Benchmarks from the UCI device mastering repository were used to create this dataset. The development of a new selection assistance device for the early detection of CKD aids in the timely treatment of CKD patients and the prevention of the disease from worsening.

A Machine Learning Analysis of a Chronic Kidney Disease Dataset was undertaken by Shahriar Shamiluulu and Azamat Serek [6]. The data was acquired via UCI's machine learning repository. It covers a total of 400 cases, with 250 of them being CDK patients and the other 150 being non-CDK patients. The support vector machines method was used to identify people with chronic renal illness. The confusion matrix is a common technique for assessing the classifier and estimating the quality of the classification process. The study was carried out using Python and the Scikit-learn machine learning framework, which includes a variety of data pre-processing techniques. The trials demonstrated that a suitably developed classifier can obtain a 94.602 percent overall performance. Such computer studies will be critical for improving people's health and detecting diseases at an early stage.

MD. rashed-al-mahfuz, abedul haque, akm azad, salem A. alyami, julian M. W. quinn, and mohammad ali moni published Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening [7]. They found a reliable strategy for CKD categorization and attribution selection that is both simple and cost-effective. The Apollo Hospitals data set is accessible in the UCI machine learning library and is from Tamil Nadu, India. Using more test features than necessary has a large budgetary impact, which makes CKD screening more difficult. As a result, they used the recently established SHAP approach to determine relevant features for CKD detection classifiers. Random forest, gradient boosting, extreme gradient boosting, logistic regression, and support vector machine were the five ML classifier methods utilised, with random forest (RF) providing the best results. As a result, the proposed RF classifier and decreased test characteristics could be used to lower diagnosis expenses and improve early treatment plan management.

J.Snegha, V.Tharani, and S.Dhivya Preetha [8] used several data mining techniques to detect kidney-related disorders, with the overall goal being to indulge the solid diagnosis rather than to locate the optimal solution. For the prediction of chronic kidney disease, they used two data mining algorithms: the Random Forest algorithm and the Back Propagation Neural Network. They used the UCI repository to get roughly 24 characteristics and 400 records. They employed two data mining techniques to estimate the probability of predicting chronic kidney disease. The output of the Random Forest Algorithm analysis yields a confidence of 88.7% and a Receiver operating characteristics Area under the curve of 90.2%.

A new sensing technique for the automated identification of renal illness was proposed by Navaneeth Bhaskar and Suchetha M [9]. To execute the classification operation, an SVM classifier is paired with a CNN network. They combined a 1-Dimensional (1-D) deep learning Convolutional Neural Network (CNN) algorithm with a Support Vector Machine (SVM) classifier to create a 1-Dimensional (1-D) deep learning Convolutional Neural Network (CNN) method. To detect the disease, the concentration of urea in the saliva is measured. The samples for this investigation were taken from 102 people, 40 of whom were healthy volunteers and 62 of whom had kidney disease.

A machine learning algorithm for diagnosing CKD was presented by Jiongming QIN and Lin Chen [10]. Logistic regression and random forest are two machine learning methods that were employed. The CKD diagnostic methodology recommended is realistic in terms of data imputation and sample diagnosis. The CKD data set came from the University of California, Irvine (UCI) machine learning repository, which features a lot of missing information. This is an integrated model that uses perceptron to combine logistic regression and random forest, with an average accuracy of 99.83 percent. We hypothesised that with more complex clinical data, our approach may be utilised to diagnose illnesses. In real-life medical diagnosis, this technology could be utilised to analyse clinical data from various disorders.

Lee Au-Yeung, Xianghua Xie, James Chess, and Timothy Scale [11] employed machine learning algorithms to refer patients with chronic renal disease to secondary care. The dataset was taken from the repository of the University of California, Irvine. The dataset contains 400 patients, with 250 of them having been diagnosed with CKD. The methods used were logistic regression, artificial neural networks (ANN), and support vector machines (SVM). We were able to get an overall accuracy of 88.48 percent, 87.12 percent, and 85.29 percent using logistic regression, ANN, and SVM, respectively. ANNs achieved the highest sensitivity of 89.74 percent, compared to 86.67 percent for logistic regression and 85.51 percent for SVM. Logistic regression models have been found to be extremely stable in tests.

Dr. Uma N Dulhare and Mohammad Ayesha [12] used the Nave Bayes Classifier to study chronic renal disease. They employed the oneR method, which is also known as the Nave Bayes Classifier. The dataset comes from the UCI repository and has 25 attributes, including 1 class attribute, 13 nominal attributes, and 11 numerical attributes. GFR is calculated from the provided attributes using the MDRD algorithm to forecast different stages. The system recommends action guidelines for different phases of CKD so that the appropriate treatment can be administered. When compared to the previous system, this enhanced accuracy by 12.5 percent.

Machine learning methodologies were employed by Anusorn Charleonnann, Thipwan Fufaung, and Tippawan Niyomwong [13] to implement predictive analysis for Chronic Kidney Disease. K-nearest neighbours, Support Vector Machine (SVM), Logistic regression, and Decision tree Classifiers were used to predict CKD. The dataset contains 400 instances, 24 attributes, and two classes and comes from the UCI Repository. There are two components to the dataset: one for training and one for testing. According to the data, the SVM classifier is appropriate for diagnosing CKD with 98.3 percent accuracy.

### III. CONCLUSION

As it is very difficult to predict the chronic kidney disease in real time. Some efficient research works has been done to predict chronic kidney disease in less time. Some of the efficient algorithms used for kidney disease prediction are SVM, decision tree, J48, logistic regression algorithms, Naive Bayes, KNN. Even though many works done on this chronic kidney disease, still some efficient and proper research works required to predict chronic kidney disease using efficient data science algorithms. Developing a real time system helps to predict chronic kidney disease in less time and this helps doctors can handle the patients in better way and give treatment at the early stage.

### REFERENCES

- [1] Bilal Khan, Rashid Khan, Ghulam Abbas “ An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy ”, IEEE international conference supported by the National Research Foundation of Korea through the Research Program under Grant NRF-2019R1A2C1005920.
- [2] Veenita Kunwar, Khushboo Chande “chronic kidney disease analysis using data mining classification techniques”, Amity University Uttar Pradesh Noida, India.
- [3] Mubarik Ahmad, Vitri Tundjungsari, Dini Widianti, Peny Amalia, Ummi Azizah Rachmawati “Diagnostic Decision Support System of Chronic Kidney Disease Using Support Vector Machine”, Information Technology and Faculty of Medicine YARSI University Jakarta, Indonesia.
- [4] Nikhila “Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm”, 2021International Conference on Computing, Communication, and Intelligent Systems(ICCCIS).
- [5] Devika R, Sai Vaishnavi Avilala, V.Subramaniaswamy “Comparative Study of Classifier for Chronic Kidney Disease prediction using Naive Bayes, KNN and Random Forest”, Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019).
- [6] Shahriar Shamiluulu, Azamat Serek “Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods”, Yedilkhan Amirgaliyev Institute of Information and Computing Technologies (IICT), Almaty, Kazakhstan.



- [7] MD. rashed-al-mahfuz , abedul haque , akm azad, salem A. alyami , julian M. W. quinn , and mohammad ali moni “Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening”, date of current version 26 April 2021.
- [8] J.Snegha, V.Tharani, S.Dhivya Preetha “Chronic Kidney Disease Prediction Using Data Mining”, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE).
- [9] Navaneeth Bhaskar and Suchetha M “A Deep Learning-based System for Automated Sensing of Chronic Kidney Disease”, This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LENS.2019.2942145, IEEE Sensors Letters.
- [10] Jiongming QIN, Lin Chen “A Machine Learning Methodology for Diagnosing Chronic Kidney Disease”, This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2963053, IEEE Access
- [11] Lee Au-Yeung, Xianghua Xie, James Chess and Timothy Scale “ Using Machine Learning to Refer Patients with Chronic Kidney Disease to Secondary Care”, 2020 25th International Conference on Pattern Recognition (ICPR) Milan, Italy, Jan 10-15, 2021.
- [12] Dr. Uma N Dulhare, Mohammad Ayesha “ Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier”, CSED, MJCET Hyderabad, India.
- [13] Anusorn Charleonnann, Thipwan Fufaung, Tippawan Niyomwong “ Predictive Analytics for Chronic Kidney Disease Using Machine Learning Techniques”, The 2016 Management and Innovation Technology International Conference (MITiCON-2016).



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)