STM JOURNALS

Journal of
Computer Technology & Applications

https://journals.stmjournals.com/jocta

ISSN: 2229-6964 (Online)
ISSN: 2347-7229 (Print)
Volume 15, Issue 3, 2024
September–December
DOI (Journal): 10.37591/JoCTA

Review JoCTA

Literature Review and Discussion of Machine Learning Algorithms for Predicting Chronic Kidney Disease

Rajesh Yadav[1],*

Abstract

Being one of the most serious and most occurring diseases in our era, chronic kidney disease requires a fast and correct diagnosis. The usage of machine learning in medicine has now grown to such a level that it could be a means of diagnosis. The doctor can be the first one to get the ailment by using machine learning classifier algorithms. This has been the data science sector's new horizons, just as a lot of other industries have been covered, healthcare in the first place. Now, most chronic diseases are no longer a mystery to healthcare workers who can foretell them using data analytics. Multiple machine learning (ML) algorithms are likely to be the future of healthcare. Not only do they identify chronic diseases, but they also forecast them at an early stage. This literature survey will focus on the effectiveness of various ML algorithms, such as K-nearest neighbor (KNN), Support Vector Machine (SVM), Random Forest, Recurrent Neural Network (RNN), etc., in predicting chronic kidney disease (CKD).

Keywords: Data analytics, machine learning algorithms, chronic kidney disease (CKD), disease prediction

INTRODUCTION

Life on Earth has been developing over millions of years, and nature has helped humans to improve their survival rate. We created and improved every field on Earth. However, one of the most significant industries is healthcare, which has undergone various improvements. With the progress of technology, a large amount of healthcare data has accumulated in recent years. This data indicates so much about a patient's health history, presence, and future [1, 2]. Tons of methods and algorithms are produced to make this data audible, and hence, diagnose patients variously in time. However, not all these aspects have yet to be addressed. Scientists are working to discover the best methods to predict and prevent chronic diseases in patients. For instance, machine learning algorithms can identify heart disease, kidney disease, etc., providing a correct diagnosis.

Chronic kidney disease (CKD) is a lifelong condition in which the performance of human kidneys suffers over time. When it is severe enough to lead to complete renal failure, dialysis for life, and sometimes death, it becomes a serious disease. CKD is a kidney killer if it is a silent disease, as the majority of its patients remain silent and do not have any symptoms until their kidney function has dropped to almost zero. CKD has various causes, including urinary infections, high blood pressure, diabetes, cholesterol, and enlarged prostate. CKD is a kidney ailment classified into stages. However, diagnosing this disease at the start is a significant issue for a patient who has a better chance of being cured [3, 4]. Creatinine levels in the blood are a key determinant of the glomerular filtration rate (eGFR), which is the most reliable measurement of renal function. The estimated glomerular filtration rate was milliliters per minute per 1.73 square meters.

*Author for Correspondence
Rajesh Yadav
E-mail: ry280888@gmail.com

[1]Assistant Professor, Department of Computer Science, South Indian Education Society College of Arts, Science and Commerce (Empowered Autonomous), Mumbai, Maharashtra, India

Received Date: September 08, 2024
Accepted Date: September 12, 2024
Published Date: October 07, 2024

Citation: Rajesh Yadav. Literature Review and Discussion of Machine Learning Algorithms for Predicting Chronic Kidney Disease. Journal of Computer Technology & Applications. 2024; 15(3): 34–39p.

The estimated glomerular filtration rate (eGFR) is a test that checks the level of kidney function in an individual and reports the corresponding stage of kidney disease.

If the eGFR figure is low, the kidneys are not functioning correctly and, as a result, CKD may develop from the apex to the feet.

An eGFR less than 60 for three or more consecutive months or a glomerular filtration rate (GFR) greater than 60 in the presence of signs of kidney damage, such as increased albuminuria, indicates chronic kidney disease (CKD). Under this assumption, Figure 1 presents the five stages of CKD with GFR values.

As a result, I compare and identify the best machine learning (ML) algorithms for predicting a range of fatal chronic diseases in this literature review.

The following algorithms were tested to predict CKD: K-nearest neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), logistic regression, artificial neural networks (ANN), Decision Tree (DT), and Naïve Bayes (NB).

LITERATURE REVIEW

The present study focused on the machine learning used for CKD prediction in the literature review (Table 1). The primary research questions are as follows: What are the most effective machine learning methods for the detection of CKD? "What kinds of machine learning methods are now being utilized in CKD diagnosis, and what are they?" In the process of diagnosing CKD, what characteristics are now considered by machine learning techniques? "What machine learning method has shown the highest degree of performance in the diagnosis of CKD in general applications?"

STAGES OF CHRONIC KIDNEY DISEASE		GFR*	% OF KIDNEY FUNCTION
Stage 1	Kidney damage with **normal** kidney function	90 or higher	90-100%
Stage 2	Kidney damage with **mild loss** of kidney function	89 to 60	89-60%
Stage 3a	**Mild to moderate** loss of kidney function	59 to 45	59-45%
Stage 3b	**Moderate to severe** loss of kidney function	44 to 30	44-30%
Stage 4	**Severe** loss of kidney function	29 to 15	29-15%
Stage 5	Kidney **failure**	Less than 15	Less than 15%

*** Your GFR number tells you how much kidney function you have. As kidney disease gets worse, the GFR number goes down.**

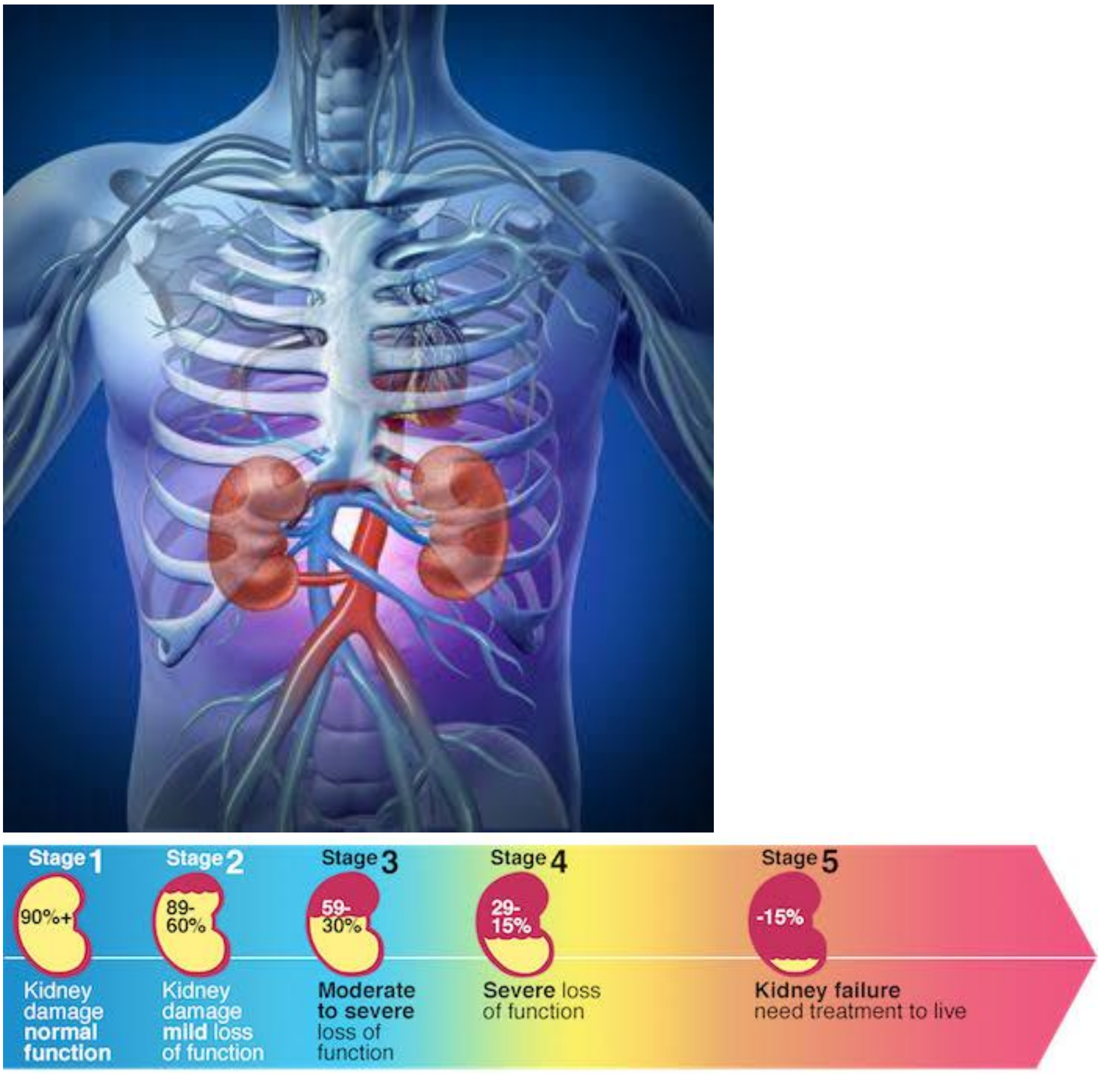

Figure 1. Stages of chronic kidney disease with GFR [5].
Source: National Kidney Foundation. (2024). Kidney Failure Risk Factor: Estimated Glomerular Filtration Rate (eGFR). [online] Available from: https://www.kidney.org/kidney-failure-risk-factor-estimated-glomerular-filtration-rate-egfr

Table 1. Representing year-wise literature review and statistical modeling method used.

References	Year	Statistical modeling method
Islam M.A. et al. [1]	2023	SVM, KNN, LGBM, CatBoost, Ada, hybrid, XGBoost, Decision Tree
Emon M.U. et al. [2]	2021	Naïve Bayes (NB), Multilayer Perceptron (MLP), Random Forest, Logistic Regression (LG), Decision Tree (DT), Adaptive Boosting (AdaBoost), Stochastic Gradient Descent (SGD)
Chittora P. et al. [6]	2021	Artificial neural network, C5.0, logistic regression, the linear support vector machine with penalty L1 and L2 as well as the Random Tree
Sathiya Priya S. et al. [7]	2018	logistic regression, lasso regression, support vector machine, Random Forest, ridge regression, neural network, Elastic Net, and K-nearest neighbor
Almansour N.A. et al. [8]	2019	SVM, ANN
Devika R. et al. [9]	2019	KNN, Naïve Bayes, and Random Forest
Anandajayam P. et al. [10]	2019	SVM, KNN, Logistic regression, Recurrent Neural Network
Sinha et al. [11]	2015	SVM, KNN
Vijayarani D.S. and Dhayanand M.S. [12]	2015	SVM, ANN

This study was designed to provide answers to these questions. In general, the questions seek expounding. In this area, many researchers have been doing research, and I will now describe some of the selected ones.

Islam Ma et al. [1] discussed a study in which initially 25 variables in addition to the class property were included in the study; however, finalizing the list of the top 30 percent of those parameters as the most identifiable subset to the CN was left. Twelve different machine learning-based classifiers were tested in a supervised learning environment. A total of twelve different machine learning-based classifiers were evaluated in a supervised learning environment, and the highest performance indicators were an accuracy of 0.98, a precision of 0.98, a recall of 0.98, and an F1-score of 0.98 for the XGBoost classifier have been achieved.

In 2021, Emon et al. [2] investigated and verified classifiers that could accurately recognize CKD at the primary stage to form a typical model for each classifier that can be used for this purpose. They are based on Naïve Bayes, Multilayer Perceptron (MLP), Random Forest, Logistic Regression (LG), Bagging, Decision Tree, Adaptive Boosting (AdaBoost), Stochastic Gradient Descent (SGD), and bagging. The application identified the crucial components by the principal component analysis (PCA) method.

Chittora et al. [6] conducted a study based on the application of seven major classifier algorithms: ANN, C5.0, logistic regression, SVM, and the Random Tree. In their work, 551 patients' data were collected and analyzed by nine classical machine learning methods, such as XGBoost, logistic regression, lasso regression, SVM, Random Forest, ridge regression, neural network, Elastic Net, and KNN, for their precision, recall, receiver operating characteristic(ROC) curve, and accuracy; it was found that the linear model had the maximum accuracy.

According to the study of Sathiya Priya et al. [7], in their feature selection method, the CKD Dataset was the main dataset around which their final tests revolved. In other words, it was the basis for the formulation of the entire binary feature set problem. ACO stands for Ant Colony Optimization and was used to perform the feature selection task. According to feature selection, the ACO method is the best part of the wrapper approach. Twenty-four different attributes were assigned to the receiver to choose from in their dataset. After the feature selection process, the model was constructed with only 12 features. The model was constructed by applying SVMs to the classification method.

In 2019, Almansour et al. [8] conducted an identical study for early diagnosis of CKD. Nonetheless, they explored SVM and ANN algorithms and later compared the effectiveness of these two algorithms, which gave welcoming results that showed more than 90% accuracy for each algorithm; however, ANN was slightly better than SVM.

Similarly, in 2019, Devika et al. [9] conducted a multifold validation experiment to verify the comparative efficiencies of various classifiers for CKD detection. Three machine learning algorithms were used: KNN, Naïve Bayes, and Random Forest. Several key performance measures, such as classification accuracy, precision, recall, and F-measure, were utilized. The confusion matrix and ROC curve were used to examine the algorithms and compare them in terms of their ability to predict the early onset of CKD. They revealed that the correct diagnostic tumor results of the three algorithms were >85%. However, the Random Forest model outperformed the other models in forecasting CKD in the early stages.

The authors of a study in 2019 Anandajayam et al. [10] also employed multiple classifiers to improve the results of the CKD diagnosis. The dataset contained both structured and unstructured data; therefore, different methodologies were used, such as SVM, KNN, Logical regression, Natural Language Processing (NLP), and Recurrent Neural Network (RNN). In a trial of multiple parameters, it was concluded that RNN wins in terms of the accuracy of CKD detection by using the feed-forward loop, which is a control group for a gap.

In 2015, Sinha et al. [11] performed a comparative study to predict CKD using SVM and KNN. The data were collected from reliable medical institutions. The dataset included 400 samples and more than 25 variables, including basic information, vital signs, complete blood count, creatinine levels, pedal edema, and other clinical parameters. The two classifiers were compared in terms of four evaluation parameters: accuracy, precision, recall, and measurement. The study was conducted using matrix laboratory(MATLAB) to eliminate the need for manual and time-consuming plotting of functions and implementation algorithms. In the analysis of the first four parameters, KNN was determined to be more accurate than SVM. Moreover, the study of KNN was more efficient in the early detection of CKD as the precision index of KNN was higher than that of SVM at the end of the study or in the early stages of the study.

In 2015, Vijayarani and Dhayanand [12] compared SVM and ANN algorithms to predict CKD. Prediction accuracy and execution time were the two most important factors in this study.

CONCLUSION

All participants were worried about chronic sickness. However, it is becoming simpler for doctors to diagnose/predict these life-threatening illnesses during the preliminary phases using state-of-the-art technology. Technology has shown that machine learning algorithms can ultimately guide the prediction of CKD and therefore protect human lives. Many studies have used a variety of ML algorithms that were studied in this literature review. Most provided results with sufficient satisfaction. Over 85% of the time, machine learning algorithms are correct and precise in their analysis.

REFERENCES

1. Islam MA, Majumder MZH, Hussein MA. Chronic kidney disease prediction based on machine learning algorithms. J Pathol Inform. 2023;14:100189. DOI: 10.1016/j.jpi.2023.100189. PubMed: 36714452.
2. Emon MU, Islam R, Keya MS, Zannat R. Performance analysis of chronic kidney disease through machine learning approaches. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT); 2021. IEEE Publications; 2021. p. 713–9. DOI: 10.1109/ICICT50816.2021.9358491.
3. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019;17:119. DOI: 10.1186/s12967-019-1860-0. PubMed: 30971285.
4. Ahmed MR, Hasan Mahmud SM, Hossin MA, Jahan H, Haider Noori SR. A cloud-based four-tier architecture for early detection of heart disease with machine learning algorithms. In: 2018 4th International conference on computer and communications (ICCC); 2018. IEEE Publications; 2018. p. 1951–5. DOI: 10.1109/CompComm.2018.8781022.
5. NHS. Overview – Chronic kidney disease. NHS Choices. 2024. Available from: https://www.nhs.uk/conditions/kidney-disease/
6. Chittora P, Chaurasia S, Chakrabarti P, Kumawat G, Chakrabarti T, Leonowicz Z, et al. Prediction of chronic kidney disease – A machine learning perspective. IEEE Access. 2021;9:17312–34. DOI: 10.1109/ACCESS.2021.3053763.
7. Sathiya Priya S, Suresh Kumar M. Chronic kidney disease prediction using machine learning. Int J Comput Sci Inf Secur. 2018;16:308–11.
8. Almansour NA, Syed HF, Khayat NR, Altheeb RK, Juri RE, Alhiyafi J, et al. Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. Comput Biol Med. 2019;109:101–11. DOI: 10.1016/j.compbiomed.2019.04.017. PubMed: 31054385.
9. Devika R, Avilala SV, Subramaniyaswamy V. Comparative study of classifier for chronic kidney disease prediction using Naïve Bayes, KNN and random forest. In: 2019 3rd International conference on computing methodologies and communication (ICCMC); 2019. IEEE Publications; 2019. p. 679–84. DOI: 10.1109/ICCMC.2019.8819654.

10. Anandajayam P, Aravindkumar S, Arun P, Ajith A. Prediction of chronic disease by machine learning. In: 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN); 2019. IEEE Publications; 2019. p. 1–6. DOI: 10.1109/ICSCAN.2019.8878724.
11. Sinha P, Sinha P. Comparative study of chronic kidney disease prediction using KNN and SVM. Int J Eng Res Technol. 2015;4:608–12.
12. Vijayarani S, Dhayanand S, Phil M. Kidney disease prediction using SVM and ANN algorithms. Int J Comput Bus Res. 2015;6:1–2.

