

Chronic Kidney Disease Prediction Based On Machine Learning Algorithms

Hizam Ahmed Al-Saeedi

Table of contents:

- Summery research paper's
- Critical Analysis
- Connection
- Conflicts between the findings
- Synthesis Matrix for Literature Review
- References

Summery research paper's

Paper 1: Chronic Kidney Disease Prediction Based On Machine Learning Algorithms

This paper analyses the suitability of several machine learning algorithms for CKD prediction in consideration of health-related parameters like age, blood pressure, and blood glucose levels. It suggests that decision trees, possibly combined with ensemble techniques like Random Forests, might be suitable for application in clinical environments because they provide insights which are useful for medical practitioners (Kethineni et al., 2025).

Paper 2: Chronic kidney disease prediction using machine learning techniques

In this study, both binary and multi classification for stage prediction have been carried out. The prediction models used include Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT). Analysis of variance and recursive feature elimination using cross validation have been applied for feature selection. Evaluation of the models was done using tenfold cross-validation. The results from the experiments indicated that RF based on recursive feature elimination with cross validation has better performance than SVM and DT (Debal & Sitote, 2022).

Paper 3: Machine learning to predict end stage kidney disease in chronic kidney disease

The purpose of this study was to assess the feasibility of machine learning (ML) in predicting the risk of end-stage kidney disease (ESKD) from patients with chronic kidney disease (CKD). Data

were obtained from a longitudinal CKD cohort. Predictor variables included patients' baseline characteristics and routine blood test results (Bai et al., 2022).

Paper 4: Early Prediction of Chronic Kidney Disease Using Deep Belief Network

This Paper proposes an intelligent classification and prediction model. It uses modified Deep Belief Network (DBN) as classification algorithm to predict the kidney related diseases and the Softmax as activation function and the Categorical Cross-entropy as a loss function (Elkholy et al., 2021).

Paper 5: Prediction of Chronic Kidney Disease - A Machine Learning Perspective

This article objects to predict Chronic Kidney Disease based on full features and important features of CKD dataset. For feature selection three different techniques have been applied: correlation-based feature selection, Wrapper method and LASSO regression. In this perception, seven classifiers algorithm were applied viz. artificial neural network, C5.0, logistic regression, CHAID, linear support vector machine (LSVM), K-Nearest neighbors and random tree. For each classifier, the results were computed based on full features, selected features by CFS, selected features by Wrapper, selected features by LASSO regression, SMOTE with selected features by LASSO, SMOTE with full features. It was observed that LSVM achieved the highest accuracy of 98.86% in SMOTE with full features.(Chittora et al., 2021).

Paper 6: Prediction of Chronic Kidney Disease Using Machine Learning Algorithm

In today's era everyone is trying to be conscious about health although due to workload and busy schedule one gives attention to the health when it shows any symptoms of some kind. But CKD is a disease which doesn't shows symptoms at all or in some cases it doesn't show any disease specific symptoms it is hard to predict, detect and prevent such a disease and this could be lead to permanently health damage, but machine learning can be hope in this problem it is best in prediction and analysis. By using data of CKD patients with 14 attributes and 400 record we are going to use various machine learning techniques like Decision Tree, SVM, etc. To build a model with maximum accuracy of predicting whether CKD or not and if yes then its Severity (Tekale et al., 2018).

Critical Analysis:

Chronic Kidney Disease (CKD) is a progressive condition that affects kidney function and can lead to severe complications if not diagnosed early. Machine Learning (ML) algorithms have been increasingly utilized for CKD prediction, offering advantages such as early detection, automated diagnosis, and improved decision-making. However, despite its potential, ML-based CKD prediction faces several methodological, technical, and practical challenges that need to be addressed.

Conflicts between the findings:

- Algorithm Choice:** Papers 1 and 6 emphasize Decision Trees for interpretability, but Paper 5 finds LSVM to be the most accurate, and Paper 2 favors Random Forest for performance.
- Feature Selection:** Paper 5 explores multiple feature selection techniques (e.g., LASSO, correlation-based), while Paper 2 focuses on Recursive Feature Elimination, indicating a lack of consensus on the best method.
- Evaluation Metrics:** Different evaluation criteria (e.g., tenfold cross-validation in Paper 2, accuracy in Paper 5) and datasets might lead to varying results, complicating direct comparisons.
- Model Complexity:** Paper 4 advocates for a Deep Learning approach (DBN), potentially at odds with Papers 1 and 6, which prioritize simpler, interpretable models.

Synthesis Matrix for Literature Review:

Study	Authors	Year	Research Design	Key Findings/Results	Methodology	Limitations
Chronic Kidney Disease Prediction Based On Machine Learning Algorithms	Kethineni, Likitha. Nithinchandra. Kumar, Narendra. Sk, Sajida Sultana.	2025	Experimental	High accuracy of the Decision Tree model demonstrates that simple and interpretable	A.Data Collection A. Data Gathering The data set was obtained from the UCI Machine Learning	More complex models, like CNN and ANN, require large amounts of data and may

				models can be very effective for the task of CKD detection when relevant features are well selected.	Repository .used in this study, which consisted of the medical history of 400 patients. The dataset had been determined for 24 relevant health attributes that will be useful to predict chronic kidney disease.	not be the best choice for smaller structured datasets.
Chronic kidney disease prediction using machine learning techniques	Debal, Dibaba Adebaba Sitote, Tilahun Melak	2022	Experimental	Applying the models on the original dataset, we have got the highest accuracy with RF, SVM, and XGBoost. The accuracy was 99.8% for the binary class and 82.56% for five-class. DT produced lowest performance compared to RF. RF also produced the highest f1_score values. SVM and RF with RFECV produced the highest accuracy of 99.8% for binary class. XGBoost has 82.56%	This study focuses on chronic kidney disease prediction using machine learning models based on the dataset with big size and recent than online available dataset collected from St. Paulo's Hospital in Ethiopia with five classes: notckd, mild, moderate, severe, and ESRD and binary classes: ckd and notckd by applying machine-learning models.	There were very limited works in the context of Ethiopia. The focuses of global studies were mainly on binary classification. There were very limited attempt on multiclass classification. Additionally, the size of the datasets used were small which makes them susceptible for overfitting and difficult for comparison.

				accuracy for five-class datasets which is the highest		
Machine learning to predict end stage kidney disease in chronic kidney disease	Bai, Qiong. Su, Chunyan. Tang, Wen. Li, Yike.	2022	Experimental	Details of the five imputed sets are provided in the supplemental materials. There was no significant difference between the imputed sets and the original dataset in each variable where missing data were replaced by imputed values. The hyperparameter settings for each classifier are displayed in Table 2. The best overall performance, as measured by the AUC score, was achieved by the random forest algorithm (0.81, see Table 3). Nonetheless, this score and its 95% confidence interval had overlap with those of the	data analysis, and drafting the manuscript. C.S. was involved in data collection. W.T. conceptualized the idea, interpreted the results and wrote part of the draft. Y.L. conceptualized the idea, analyzed the data, performed all coding, evaluated all machine learning models, drafted and edited the manuscript.	Several limitations should be noted. First, this cohort consisted of less than 1000 subjects and ESKD only occurred in a small portion of them, both of which might have affected model performance as discussed earlier. Second, although this study aimed to assess the feasibility of a prediction model for ESKD without any urine variables, this was partially due to the lack of quantitative urine tests at our institute when this cohort was established. A

				other three models, including the logistic regression, naïve Bayes, and the KFRE (Fig. 1). Interestingly, the KFRE model that was based on 3 simple variables, demonstrated not only a comparable AUC score but also the highest accuracy, specificity, and precision. At the default threshold, however, the KFRE was one of the least sensitive models (47%)		
Early Prediction of Chronic Kidney Disease Using Deep Belief Network	Elkholy, Shahinda Mohamed Mostafa. Rezk, Amira. Saleh, Ahmed Abo El Fetoh.	2021	Experimental	The efficiency of the proposed model is evaluated, and also made a comparison with the existing models. The proposed model achieves better performance,	The methodology of our research consists of sequence steps including data collection, pre-processing, deep learning-based feature engineering and classification followed by evaluation	The used dataset has missing values, which reduce the efficiency, so it must be removed before analyzing data.

				comparing with the existing models, with accuracy 98.52%. So, the proposed model presents proper predictor and classifier for CKD.		
Prediction of Chronic Kidney Disease - A Machine Learning Perspective	Chittora, Pankaj. Chaurasia, Sandeep. Chakrabarti, Prasun. Kumawat, Gaurav.	2021	Experimental	As per the result, it is concluded that SMOTE is a best technique for balancing a dataset. It is noted that SMOTE gave better results with selected features by LASSO regression as compare to without SMOTE on LASSO regression model. LSVM achieved the highest accuracy in all experiments as compared to other classifiers algorithms.	In this perception, seven classifiers algorithm were applied viz. artificial neural network, C5.0, logistic regression, CHAID, linear support vector machine (LSVM), K-Nearest neighbors and random tree. In this research, we have developed a model to predict CKD disease in patients. The performance of the model was tested on both all attributes and selected features.	It contains missing value
Prediction of Chronic Kidney Disease Using Machine Learning Algorithm	Tekale, Siddheshwar. Shingavi, Pranjal. Wandhekar, Sukanya.	2018	Experimental	In this paper we have studied different machine learning algorithms.	Date , Data Prepossessing, Cleaning dataset, Training data, Testing data, ML	To build a machine learning model targeting chronic

				We have analysed 14 different attributes related to CKD patients and predicted accuracy for different machine learning algorithms like Decision tree and Support Vector Machine. From the results analysis, it is observed that the decision tree algorithms gives the accuracy of 91.75% and SVM gives accuracy of 96.75%. When considering the decision tree algorithm it builds the tree based on the entire dataset by using all the features of the dataset. The advantage of this system is that, the prediction process is less time consuming.	algorithm, Predication.	kidney disease with overall accuracy of 99.99%, will need millions of records with zero missing values
--	--	--	--	---	--------------------------------	---

References:

- Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-12316-z>
- Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., Jasinski, M., Jasinski, L., Gono, R., Jasinska, E., & Bolshev, V. (2021). Prediction of Chronic Kidney Disease - A Machine Learning Perspective. In *IEEE Access* (Vol. 9, pp. 17312–17334). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2021.3053763>
- Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, 9(1). <https://doi.org/10.1186/s40537-022-00657-5>
- Elkholy, S. M. M., Rezk, A., & Saleh, A. A. E. F. (2021). Early Prediction of Chronic Kidney Disease Using Deep Belief Network. *IEEE Access*, 9, 135542–135549. <https://doi.org/10.1109/ACCESS.2021.3114306>
- Kethineni, Likitha., Nithinchandra, Kumar, N., & Sk, S. Sultana. (2025). Chronic Kidney Disease Prediction Based On Machine Learning Algorithms. *ITM Web of Conferences*, 74, 01004. <https://doi.org/10.1051/itmconf/20257401004>
- Tekale, S., Shingavi, P., & Wandhekar, S. (2018). Prediction of Chronic Kidney Disease Using Machine Learning Algorithm. *IJARCCCE*, 7(10), 92–96. <https://doi.org/10.17148/IJARCCCE.2018.71021>