Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

Chronic Kidney Disease Risk Prediction Using Machine Learning Techniques

Baswaraj D

Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, Telangana, India. E-mail: braj.d@staff.vce.ac.in

Chatrapathy K

School of Computing and Information Technology, REVA University, Bangalore (North), Karnataka, India. Email: pathykc@gmail.com

Mudarakola Lakshmi Prasad

Computing Science and Engineering, Institute of Aeronautical Engineering, Hyderabad, Telangana, India. E-mail: prasad.hinduniv@gmail.com

Pughazendi N

Computer Science and Engineering, Panimalar Engineering College, Chennai, Tamil Nadu, India. E-mail: pughazendi@gmail.com

Ajmeera Kiran

Computer Science and Engineering MLR Institute of Technology, Dundigal, Hyderabad, Telangana, India. E-mail: kiranphd.jntuh@gmail.com

Partheeban N

School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India. E-mail: n.partheeban@galgotiasuniversity.edu.in

Pundru Chandra Shaker Reddy*

*Corresponding Author, School of Computing Science and Artificial Intelligence, SR University, Warangal-506371, Telangana, India. E-mail: chandu.pundru@gmail.com

Abstract

In healthcare, a diagnosis is reached after a thorough physical assessment and analysis of the patient's medicinal history, as well as the utilization of appropriate diagnostic tests and procedures. 1.7 million People worldwide lose their lives every year due to complications from chronic kidney disease (CKD). Despite the availability of other diagnostic approaches, this investigation relies on machine learning because of its superior accuracy. Patients with chronic kidney disease (CKD) who experience health complications like high blood pressure, anemia, mineral-bone disorder, poor nutrition, acid abnormalities, and neurological-complications may benefit from timely and exact recognition of the disease's levels so that they can begin treatment with the most effective medications as soon as possible. Several works have been investigated on the early recognition of CKD utilizing machine-learning (ML) strategies. The accuracy of stage anticipations was not their primary concern. Both binary and multiclass classification methods have been used for stage anticipation in this investigation. Random-Forest (RF), Support-Vector-Machine (SVM), and Decision-Tree (DT) are the prediction models employed. Feature-selection has been carried out through scrutiny of variation and recursive feature elimination utilizing cross-validation (CV). 10-flod CV was utilized to assess the models. Experiments showed that RF utilizing recursive feature removal with CV outperformed SVM and DT.

Keywords: Machine Learning, CKD, Prediction, SVM, RF, Data Analysis

Journal of Information Technology Management, 2024, Vol. 16, Issue 1, pp. 118-134
Published by University of Tehran, Faculty of Management
doi: https://doi.org/ 10.22059/jitm.2024.96378
Article Type: Research Paper

Received: December 17, 2023
Received in revised form: January 06, 2024
Accepted: February 02, 2024
Published online: February 29, 2024

Introduction

Two kidneys, situated in the rear of the peritoneal cavity, perform a variety of crucial functions for the human body. The kidneys are responsible for controlling the levels of salt, water, and ions and trace elements throughout the body. Vitamin D controls calcium and phosphorus levels, bone-structure, and many other functions, whereas erythropoietin promotes red blood cell synthesis and maturation in the bone marrow (Raj et al., 2023). Hormones that control BP, fluid-balance, bone-metabolism, and vascular calcifications all act on the kidneys. The kidneys are responsible for getting rid of any excess metabolic waste, medications, or toxins in the body (Teju et al., 2021).

CKD is typically brought on by either diabetes or high-BP. High blood sugar levels are a hallmark of diabetes, which can lead to complications affecting the kidneys, heart, blood-vessels, and eyes. Inadequate management of hypertension is also a risk factor for

cardiovascular disease, cerebrovascular accident, and kidney failure. Most people with CKD have no signs until their renal-function declines to between 15 and 20% of normal (Ashreetha et al., 2022). Fatigue, weakness, difficulty concentrating, loss of appetite, trouble sleeping, nighttime muscle cramps, swelling of the legs &ankles, puffiness across the eyes, dry-skin with deep itching, and increased urination, particularly at night, are the most prominent symptoms of CKD's advanced stage (Hui et al., 2023). Damage to both kidneys leads to the most frequent form of kidney illness, called chronic kidney disease (CKD), which causes long-term suffering for those who are affected. In this context, the word "kidney damage" refers to any disorder affecting the kidneys that can lead to impaired renal function. Any disease or deficiency, such as a lowered GFR, could be to blame. Our suggested prediction model utilizes a stacking classifier built upon the random forest method to make predictions based on clinical symptoms (LK et al., 2021).

Due to its ability to do complex analysis, machine learning is gaining importance in healthcare diagnostics. This helps reduce the likelihood of human mistake and improve the accuracy of forecasts. Diseases like heart disease, diabetes, cancers, and liver disease may now be reliably predicted using machine learning algorithms and classifiers (Deo et al., 2023). Nave Bayes, support vector machines (SVMs), and decision trees were utilized for classification in medical prediction using machine learning techniques; regression was handled using random forests, logistic regression, and linear regression. The death rate can be reduced with early diagnosis and prompt treatment with the use of these algorithms. Patients with chronic renal disease should preserve their clinical symptoms by engaging in regular physical activity (Collo et al., 2022). They need to get some exercise, hydrate, and stay away from the junk stuff. Chronic renal disease's typical signs are depicted in Figure1 below.

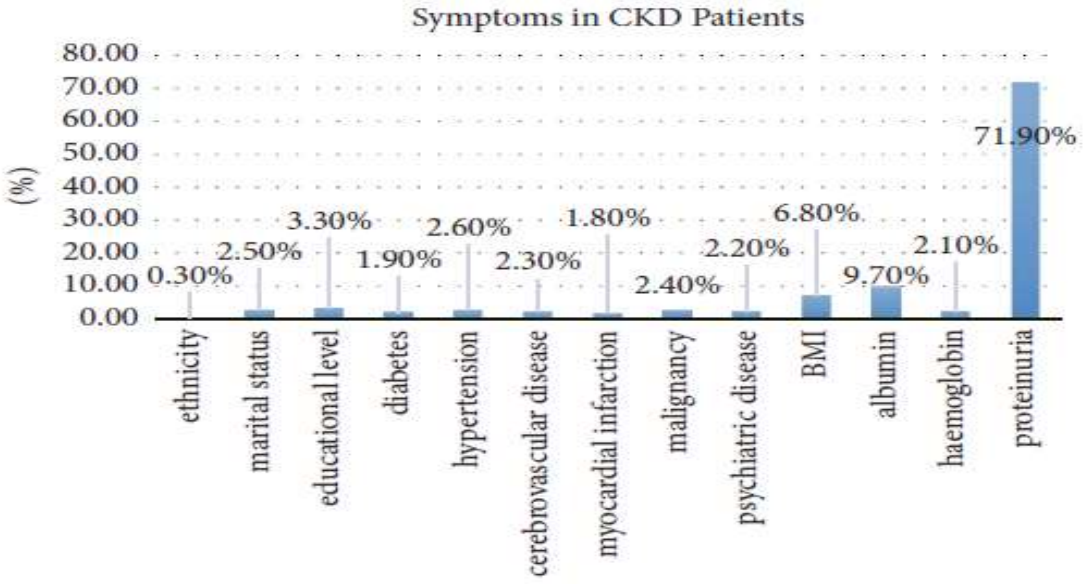

Figure 1. General symptoms of CKD patients

This study first provides a background and analysis of the ML-classifiers utilized in CKD-diagnosis, and then proceeds to implement and evaluate those classifiers. The significance of machine learning classifiers in healthcare is also discussed, along with an explanation of how these might improve prediction accuracy (Matsushita et al., 2023). An accurate ML-anticipation strategy for CKD is the primary focus of this article's research and proposed implementation. In this article, we examine the performance of various ML-strategies on the similar dataset in terms of their exactness. By combining popular ML-classifiers like the DT, SVM & RF, we aim to construct the most accurate hybrid model possible while also solving the overfitting problem (Major et al., 2022). The end goal is to provide CKD patients with a precise and successful treatment at a lower cost. We necessitate brushing up on our knowledge of common renal illnesses before moving on (Muthappa et al., 2023). The medical community faces a significant problem in attempting to diagnose and treat CKD at an early stage. The nephrologist's goals are twofold: first, to prevent the disease from progressing to more severe stages, and second, to treat the aforementioned systemic symptoms. The following are some of the most important results from using this methodology:

- To effectively develop classification techniques to anticipate the risk for CKD incident, it is necessary to do data preprocessing to guarantee that the dataset instances are scattered uniformly.
- A comparative estimation of the performance of diverse models is presented taking into account the metrics, including Precision, Recall, F-Measure, Accuracy, and F1-score.
- The results of a performance assessment are shown, revealing that all models produced impressively high outcomes, with RF producing the best results across all measures and so constituting the main recommendation of this study.

Here are how the remaining parts of the task are organized. In Section 2, we showcase further works that make use of ML in the context of CKD. In addition, the dataset and methods are analyzed and discussed in Section 3. In addition, we report and discuss the findings in Section 4. In Section 5, we wrap up the paper and lay out plans for what comes next.

Literature Review

Academicians and scientists are currently fascinated in designing tools and methods for supervising and anticipating various diseases, particularly those that are widespread in human life. Here, we'll go through some of the most up-to-date research on using ML strategies to anticipate CKD-risk, as well as some techniques for handling very small datasets. In (Kumar et al., 2023), authors used machine learning to construct a self-learning knowledge-based system for analysis and cure. They found that SVM had the greatest classification accuracy (98.3%) and sensitivity (0.99) of any method they tested. Classifiers like K-NN, RF, & ANN

were evaluated on a 400-sample dataset (Wang et al., 2022). 5features were chosen for model development based on the study's implementation of a wrapper feature selector. By using RF, we get an RMSE of 0.11 and a classification precision of 98%. A dataset of 400cases and 14features was used in the "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm" (Mondol et al., 2022). They have made use of decision trees and SVMs. Preprocessing has been applied to the dataset, and the original 25 features have been reduced to 14. With a claimed accuracy of 96.75 percent, SVM is touted as the superior model.

Prediction of chronic renal disease using naïve bayes was performed (Barua et al., 2020). They analyzed the past medical records of 551 people who had proteinuria and utilized 18 criteria to categorize the severity of the condition into mild, moderate, and severe. They determined that Logistic regression was superior due to its AUC of 0.873, Se of 83%, and Sp of 82%. For the first three stages of chronic renal disease, Mohammed and This study makes use of a limited amount of data, but the researchers have created a prototype that lets patients query KBS to monitor the process of advice distribution. The rules were generated using a decision tree. It has been reported that the prototype is 91% effective in its current form (Rao et al., 2023).

The goal of the research conducted in (Lambert et al., 2022) was to estimate the performance of ML strategies in the anticipation of chronic renal illness based on a small number of features. Predictive features were chosen using statistical &ML methods, to accomplish some modeling. They found that Gradient Boosting achieved an F-measure of 99.1—the highest of any method they tested. They evaluate the data of CKD patients to forecast the risk of CKD utilizing ML-strategies. There has been some use of Random Forest and an ANN. Twenty of the possible 25 features have been retrieved and RF and ANN have been applied (Aswathy et al., 2022). The greatest percentage of correct RF identifications to date is 97.12%. Algorithms for predicting renal disease stages were compared using probabilistic neural networks (PNNs), multilayer-perceptrons (MLPs), SVMs, and radial-basis-functions (RBFs). The study relied on a short dataset with a limited number of attributes. Probabilistic Neural Networks, according to the results of this paper, achieve better results. Various ML-strategies are just some of the supervised-ML strategies that (Baskar et al., 2023) evaluated to resolve which one would be the best for BCD anticipation. In conclusion, the results demonstrated that k-NN achieves the highest accuracy (97%) on the BCD dataset.

According to these reviews, there have been multiple attempts to forecast chronic renal disease with machine learning approaches. Dataset size, quality, and collection time are just a few of the many variables that can significantly affect a model's effectiveness. In this study, we apply ML strategies to the problem of CKD anticipation using a large, recently collected dataset. Most past studies have only looked at two groups, which makes it hard to develop therapy recommendations based on the stages involved.

Methodology

Data source and description

St. Paul's Hospital provided the data for this analysis. It's the second-biggest public hospital in Ethiopia, and it sees a lot of patients with long-term conditions. The facility offers dialysis services and is equipped to perform kidney transplants. Patients with CKD who were hospitalized to the renal ward in 2018 and 2019 comprise the dataset for this study. Some of them came from the same patient history information, gathered at various points in time Sampath et al., 2023). Interviews with subject matter experts have been undertaken to help prepare the dataset and get insight into its properties. There are a total of 1718 cases in the dataset, and the dataset has 19 features (12 numerical and 7 nominal). When broken down into their respective stages, the distribution shows that 441 (25.67%) cases represent end-stage renal disease, 399 (23.22%) represent severe stage4, 354 (20.61%) represent moderate stage3, 248 (14.44%) represent mild stage2, and 276 (16.07%) represent no CKD. There are 1,442 ckds (stages 1–5) in the binary class and 272 notcckds. There is a discrepancy between the two categories. Data resampling methods that employ oversampling have been utilized to achieve this fair distribution of resources. The resampling method increased the size of the binaryclass dataset to 2888.

Preprocessing

Inconsistencies in real-world data can have an impact on how well models work. An essential aspect of building a machine-learning model is preprocessing the data before feeding it into classifiers. The dataset used in this investigation also has missing values that must be dealt with properly. It also needs to be in a form that can be easily modeled. As can be seen in Figure 2, preparatory work has therefore been completed (Ediaredoh et al., 2022).

Preprocessing includes cleaning noisy data by removing outliers and smoothing the data. Values that are significantly outside the norm are called outliers. The inherent variability of clinical data might give birth to outliers. Equipment failure might cause data to be unavailable (or missed), and inconsistencies can lead to deletions or data not being entered into the database because of misunderstandings or because it was not deemed significant at the time (Sucharitha et al., 2023).

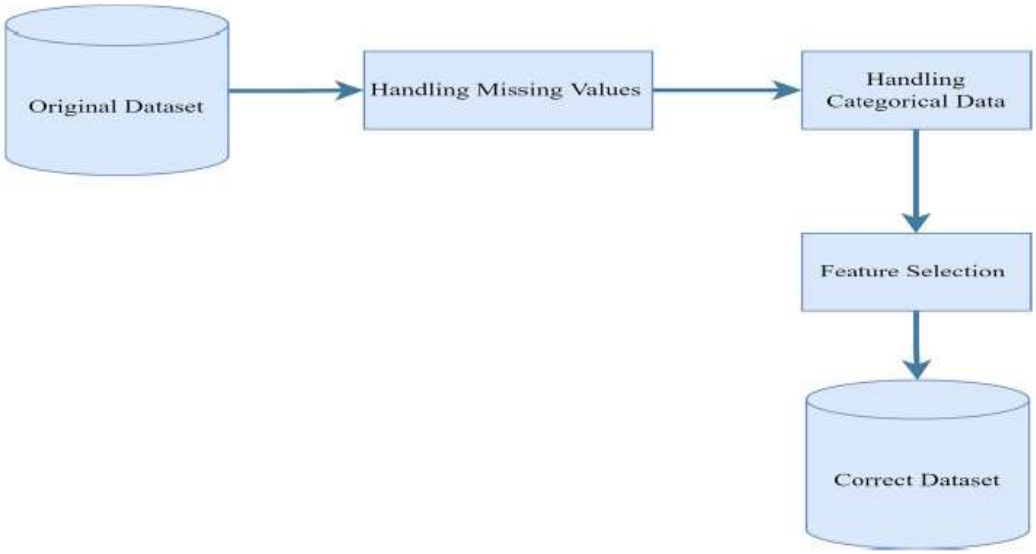

Figure 2. Preprocessing steps of the proposed model

There is a lack of diagnostic test findings in patient data that may be used to forecast the likelihood of diagnoses or the efficacy of treatment (Lee et al., 2022). Prediction accuracy is negatively impacted by missing data. Dropping & filling in missing values are just two methods for dealing with them. If the proportion of missing data is low, say 10% or less, it may be disregarded. However, this is not ideal for the model's health because the missing number may be a crucial factor in the evolution of the model. In some cases, the model does not change even if missing variables are substituted with 0. Since the missing features in this research are numbers, and since mean imputation performs better for numerical missing values, it was utilized to manage the missing data. This stage involves converting data into the necessary format. The nominal information was transformed into binary digits (0s and 1s). Normalization is performed by taking the mean and standard deviation of a feature's values. This is how it goes:

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

Where z is Z-score, x is feature-value, μ is mean-value and σ is SD.

Feature selection

For a high-quality outcome, it is crucial to isolate a small collection of useful predictive features (Kumar et al., 2023). The goal of feature-selection (FS) is to decide the most informative and useful features to nosh into models. To solve the high-dimensionality crisis, this preprocessing step is crucial. Therefore, the primary goal of FS is to pick the relevant and decoupled subset of characteristics for training the model (Sucharitha et al., 2023). Feature selection is also significant when designing a predictive model for CKD. After the dataset was

built, relevant characteristics were chosen using feature selection techniques. Due to the relative simplicity of the method, it has found widespread application. The wrapper technique uses a classification algorithm to identify the most pertinent information. In terms of precision, it outperforms the filter FS approach. However, additional time for processing is needed (Suneel aet al., 2024). The filter methods used to pick the univariate features used in this study were chosen for their speed, efficiency, and scalability. When it comes to selecting features, the wrapper technique has relied on recursive feature elimination with cross-validation (RFECV).

Univariate-Feature-Selection (UFS): This technique is widely used since it is fast, easy, and effective for selecting features from healthcare datasets. It analyzes the correlation between each attribute and the dependent variable independently. It may be used with any classifier and is quick to implement. Pearson-correlation, information=gain, chi-square, and Analysis-of-Variance (ANOVA) are only a few examples of available univariate algorithms. In this research, we used the ANOVA equation presented in Eq. 2 to choose features.

$$F = \frac{MST}{MSE} \tag{2}$$

Mean-sum-of-squares of treatment (MST), mean-sum-of-squares error (MSE), and the analysis of variance (ANOVA) coefficient (F).

Recursive-Feature-Elimination with Cross-Validation (RFECV): A method for optimizing the development of a trained machine-learning model by iteratively excluding non-essential characteristics. In order to find the optimal subset of features, iteratively creating the design, setting aside the lowest performing attribute at every iteration, and then building the next form with the remaining features is used (Bai et al., 2022). To boost the model's generalization performance, it gets rid of redundant and weak features that won't hurt training but preserves independent and strong ones (Saha et al., 2022). To determine which qualities have been deemed most significant, this technique employs an iterative process for rating them. Since these methods function best in tandem with a machine learning model, the latter is constructed using the full collection of features and then prioritized.

Machine Learning Models

The models that will be taken into account in the CKD-risk anticipation framework are briefly described in this section. This is done by employing a wide range of classifiers and gauging how well they make predictions. Classification strategies such as SVM, DT, and RF will be detailed. The idea of this research was to examine the efficacy of ML in predicting chronic renal disease. SVM, DT &RF, three ML strategies, were utilized in this research. The methods were chosen based on their prevalence in CRD anticipation and their classification performance in prior works of study (Latha et al., 2023). Both multiclass and binary classification ML models were constructed. From the three algorithms used for each

categorization, the one with the best overall performance was chosen as the best machine learning model.

Random-Forest: The ensemble learning framework called RF uses a number of diverse groups of DTs. It can be put to work in both regression and classification settings. There are multiple decision trees in this model, and the class target that obtains the most votes is the one that gets produced (Sharma et al., 2023). RF creates a forest of trees that are autonomous of one another through the use of bagging and random feature selection. Collectively, they were able to make a more precise prediction than any one tree could have. After the forest is constructed, test examples are propagated to all of the individual trees, and from there, each tree makes a classification prediction (Deivasigamani, et al., 2023).

Support vector machine (SVM): When it comes to supervised machine learning algorithms, SVM is among the most well-known and practical applications. To properly categorize the high-dimensional data at hand, a collection of hyperplanes must be constructed. Compounds are categorized according to which side of a discrete hyperplane that is constructed in the signifier liberty of the train data. Separating the dots in the data are decision boundaries called hyperplanes (Abdel-Fattah et al., 2022).

Decision Tree (DT): It is widely used since it is one of the best supervised ML-classification methods. To address the challenge of machine learning, Decision Tree employs a tree representation of the data achieved by ordered feature values. Nodes in a decision tree represent features of an instance, while leaf nodes reflect the classes to which those examples belong (Mamatha et al., 2023).

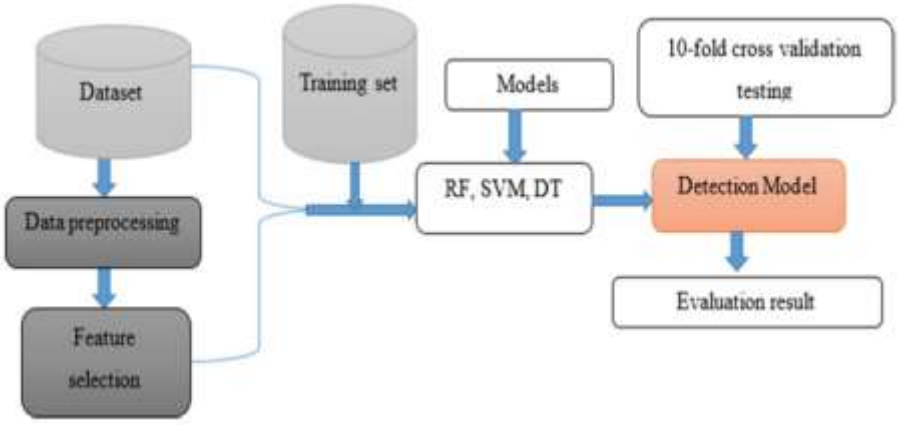

Figure 3. Workflow diagram of the proposed model

Prediction model evaluation

In order to create a reliable ML strategy, it is essential to regularly assess its performance. The effectiveness of a prediction model on both known and unknown data requires testing. Estimating how well a model performs on data it has never seen is the goal of performance evaluation. In order to evaluate and compare models, Cross-Validation (CV) is one performance estimation that uses data partitioning. Nine folds from the original dataset were utilized to train a model, and one fold was utilized to test or verify the model. The results of k iterations of this procedure will be averaged. In this research, ten-fold cross-validation was used. Equations (2) through (7) have been used to calculate a variety of performance appraisal metrics, including as accuracy (Acc), precision (Pe), recall (Re), f1-score, and specificity (Sp).

The proposed method for breast cancer categorization has substantial implications for enhancing classification accuracy and speed. It was suggested that future studies use a broader range of ML-models as baselines, incorporate additional types of medical images, and investigate further meta-learning methodologies.

$$\text{Acc} = \frac{Tr_{Positive} + Tr_{Negative}}{Tr_{Positive} + Tr_{Negative} + Fa_{Positive} + Fa_{Negative}} \quad (3)$$

$$\text{Pe} = \frac{Tr_{Positive}}{Tr_{Positive} + Fa_{Negative}} \quad (4)$$

$$\text{Sp} = \frac{Tr_{Negative}}{Tr_{Negative} + Fa_{Negative}} \quad (5)$$

$$\text{Re} = \frac{Tr_{Positive}}{Tr_{Positive} + Fa_{Postive}} \quad (6)$$

$$\text{F1-score} = 2 \times \left(\frac{Pe \times Re}{Pe + Re}\right) \quad (7)$$

Results and Discussion

Two distinct sets of features emerged from the features selection procedure utilizing UFS and RFECV, respectively. The resulting feature subsets have been implemented in RF, SVM, and DT training. Because RFECV can automatically delete less extrapolative elements iteratively based on the model, the preferred features vary between the five-class and binary-class scenarios. Table 1 displays the number of binary &five-class features chosen to determine the dataset size.

Table 1. FS outcomes

FS Method	ML-approach	Binary	5-class
RFECV	SVM	9	10
	RF	8	9
	DT	16	7
UFS	SVM	12	14
	RF	14	16
	DT	16	18

Two FS strategies and 3 classifiers were used to test 18 models for binary & five-class classification, respectively. These models have been trained and tested using 10 fold CV. It randomly divides the dataset into 10sets of the same size. After that, we used 10-1 folds for model training and the final fold for testing. For each fold, the procedure is repeated. The findings are shown in both two- and five-class classification structures. Binary &multi-class classification models were initially performed on a preprocessed dataset without employing FS techniques. The 2FS approaches are then used to the modeling experiments described below.

Binary classification models evaluation results

The binaryclass dataset derived from the original multi-class dataset was used to develop these classification models. notckd and ckd are the intended recipient classes. Tenfold CV and other CV performance evaluation criteria are used to train and test the models. As has been said, the dataset was preprocessed prior to any modeling being done on it, and no feature selection techniques were used. Then, the most predictive traits have been picked using FS methods. UFS and RFECV were utilized to implement feature selection. Table 2 displays the results of each test for the RF, SVM, & DT techniques with and without the FS approach used.

Table 2. Binary classification with & without FS

Model	Acc (%)	Pe (%)	Re (%)	F1 (%)	Sp (%)
SVM	96.8	99.1	99.7	98.24	99.2
RF	99.6	99.9	99.5	98.19	99.35
DT	98.2	99.4	97.56	98.38	99.6
SVM with RFECV	97.3	98.5	93	97.39	99.4
RF with RFECV	99.9	99.8	99.7	99.8	99.9
DT with RFECV	98.5	99.5	98.25	98.27	99.7
SVM with UFS	98.1	99.4	97.67	97.24	99.5
RF with UFS	99.8	99.7	98.7	99.7	99.9
DT with UFS	98.6	99.3	96.1	98.34	99.6

The RF with RFECV design with the chosen 8features achieved the best accuracy of 99.9%. Model output without feature selection is depicted graphically in Figure 4.

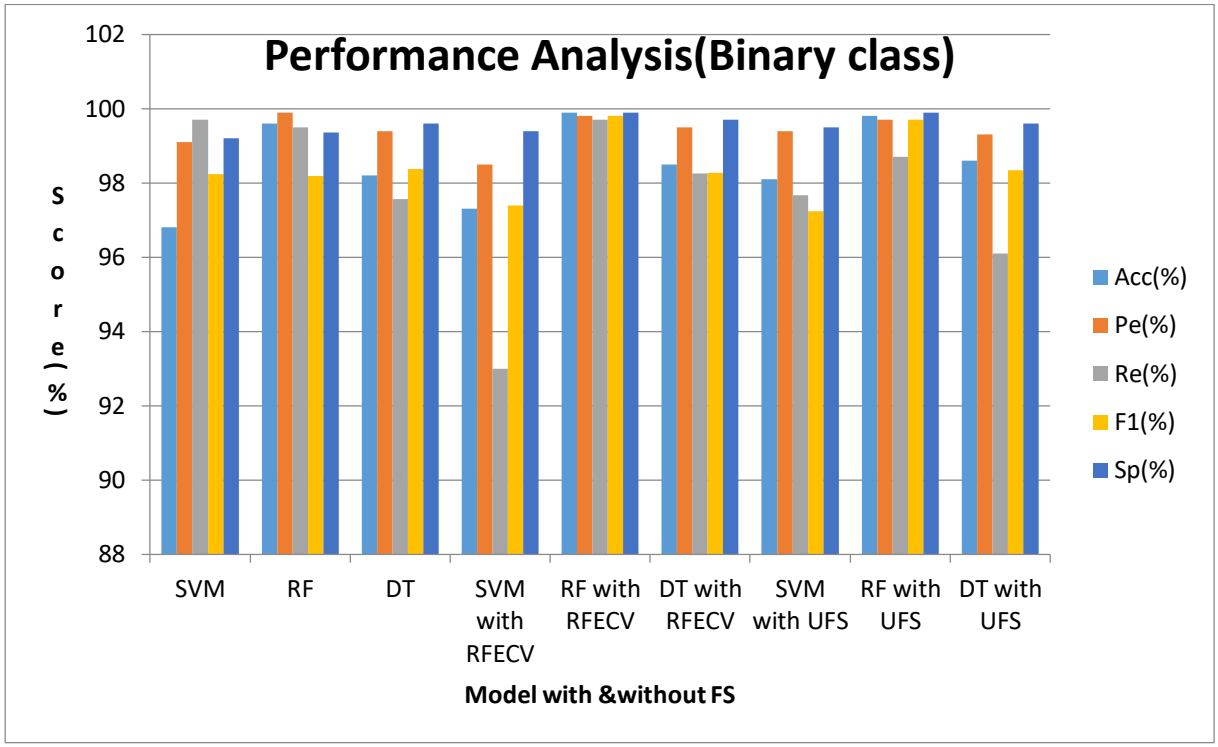

Figure 4. Binaryclass classification results with &without FS

Multiclass classification models evaluation results

The five-class dataset was preprocessed, and then the multiclass models were constructed in a similar fashion. Tenfold CV is used for training and testing the models, and along with other metrics, the models are evaluated for their performance. Table 3 displays the results of the performance measures for both the no-feature-selection and feature-selection-applied versions of each trained model (RF, SVM, & DT). We start by training and testing our models using all available features, without using any feature selection techniques.

Table 3 displays the before and after feature selection results for CV performance measures for three classifiers of the multiclass dataset. Using RFECV with the chosen 9 features, RF achieves an accuracy of 79%. Figure 5 is a visual representation of the outcomes of model training using feature selection. In a similar vein, we have optimized the hyperparameters of a support vector machine for a multiclass dataset by means of a grid search informed by cross validation. No feature selection was used to do this. Even if the performance increased to 78.78%, it is still not as good as the best-performing model RF with RFECV.

Table 3. Multi-class classification with & without FS

Model	Acc (%)	Pe (%)	Re (%)	F1 (%)	Sp (%)
SVM	63	61	61.5	60	90.5
SVM with RFECV	62	59	60.8	56.5	90.4
SVM with UFS	61.5	59.2	60.5	58.4	90.2
DT	77.9	80	76.4	76.2	94.5
DT with RFECV	78	81	76.9	76.6	94.6
DT with UFS	77.9	79.8	76.7	76.4	94.4
RF	78	80	77.5	77.2	95
RF with RFECV	79	81	78	78	94.7
RF with UFS	77.5	79	76.5	76.7	94.3

We used performance appraisal metrics such accuracy, recall, f1, specificity to assess the effectiveness of the suggested models in a ten-fold cross-validation. Previously to the application of feature selection, the machine-learning models RF, SVM, & DT were implemented. The maximum binary accuracy was achieved by RF, at 99.6%. For five different categories, its accuracy is 78%. The accuracy of the SVM was 96.8% for the binaryclass and 63% for the multiclass. The accuracy of DT's binary class results was 98.2%, but its five-class results were just 77.9%. The two datasets were then subjected to the feature selection procedures. The maximum accuracy was achieved for the binary class by both SVM and RF with RFECV. The results look good, and we think they can be used to help doctors make quicker, more accurate diagnoses. Based on the results of our analysis, we conclude that SVM, RF with RFECV is the best method for performing binary class and five-class classifications.

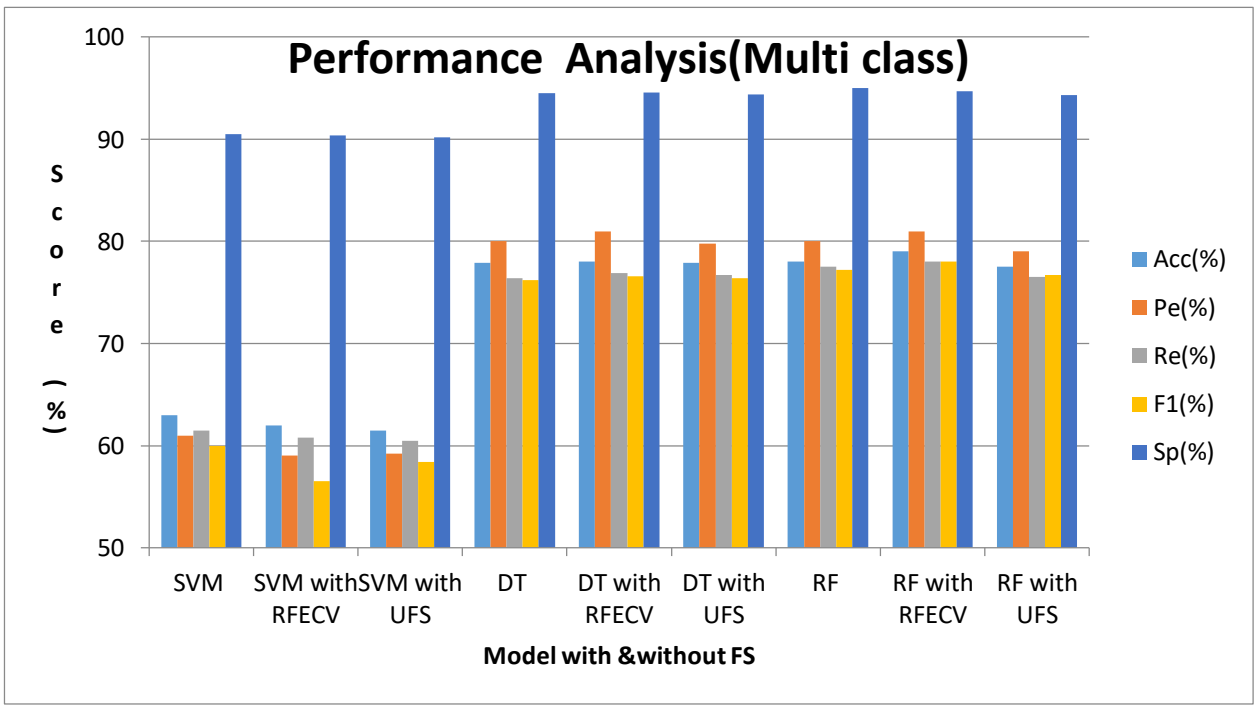

Figure 5. Multiclass classification results with &without FS

Conclusion

Experts and patients alike must have access to reliable early prediction tools if they are to reduce the risk of CKD progressing to kidney collapse. In this research, we employed three different ML startegies (RF, SVM, & DT) and two different FS techniques (RFECV & UFS) to construct our final models. Ten-fold CV was utilized for the model assessments. To begin, datasets containing all 19 features were fed into each of the four machine learning methods. We found that RF and SVM performed best when applied to the raw data. For the binary class, accuracy was at 99.9%, but it was only at 79% for the five-class. When compared to RF, DT's output was significantly lower. The highest f1_score values were also generated via RF. The best accuracy (99.8%) for the binary class was achieved using SVM and RF with RFECV. Therefore, we think our multi categorization work was crucial for understanding the disease's progression and recommending appropriate treatments for individuals.

The models were developed using a supervised machine-learning algorithm and feature selection techniques. Using models built with unsupervised or deep learning algorithms allows for a clearer comparison of performance outcomes. It is preferable to create the proposed model a mobile-based system that allows specialists to track the state of patients and assist individuals in using the system to learn about their condition.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Ahmed, T. I., Bhola, J., Shabaz, M., Singla, J., Rakhra, M., More, S., & Samori, I. A. (2022). Fuzzy logic-based systems for the diagnosis of chronic kidney disease. BioMed Research International, 2022.

Ashreetha, B., Devi, M. R., Kumar, U. P., Mani, M. K., Sahu, D. N., & Reddy, P. C. S. (2022). Soft optimization techniques for automatic liver cancer detection in abdominal liver images. *International journal of health sciences*, *6*.

Aswathy, R. H., Suresh, P., Sikkandar, M. Y., Abdel-Khalek, S., Alhumyani, H., Saeed, R. A., & Mansour, R. F. (2022). Optimized tuned deep learning model for chronic kidney disease classification. *Comput. Mater. Contin*, *70*, 2097-2111.

Barua, N. (2020). *Computational Study on Interfacial Properties of Boron Nitride Nanosheets and Its Relevance for Application in Ethylene Glycol Removal from Water* (Doctoral dissertation, North Carolina Agricultural and Technical State University).

Baskar, S., Nandhini, I., Prasad, M. L., Katale, T., Sharma, N., & Reddy, P. C. S. (2023, November). An Accurate Prediction and Diagnosis of Alzheimer's Disease using Deep Learning. In 2023 IEEE North Karnataka Subsection Flagship International Conference (NKCon) (pp. 1-7). IEEE.

Colli, V. A., González-Rocha, A., Canales, D., Hernández-Alcáraz, C., Pedroza, A., Pérez-Chan, M. & Denova-Gutierrez, E. (2022). Chronic kidney disease risk prediction scores assessment and development in Mexican adult population. *Frontiers in Medicine*, *9*, 903090.

Deivasigamani, S., Rani, A. J. M., Natchadalingam, R., Vijayakarthik, P., Kumar, G. B. S., & Reddy, P. C. S. (2023, August). Crop Yield Prediction Using Deep Reinforcement Learning. In 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON) (pp. 137-142). IEEE.

Deo, R., Dubin, R. F., Ren, Y., Murthy, A. C., Wang, J., Zheng, H. & Ganz, P. (2023). Proteomic cardiovascular risk assessment in chronic kidney disease. *European Heart Journal*, *44*(23), 2095-2110.

Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. *Bioengineering*, *9*(8), 350.

Hui, M., Ma, J., Yang, H., Gao, B., Wang, F., Wang, J. & Zhao, M. (2023). ESKD Risk Prediction Model in a Multicenter Chronic Kidney Disease Cohort in China: A Derivation, Validation, and Comparison Study. *Journal of Clinical Medicine*, *12*(4), 1504.

Kumar, A., Satheesha, T. Y., Salvador, B. B. L., Mithileysh, S., & Ahmed, S. T. (2023). Augmented Intelligence enabled Deep Neural Networking (AuDNN) framework for skin cancer classification and prediction using multi-dimensional datasets on industrial IoT standards. *Microprocessors and Microsystems*, *97*, 104755.

Kumar, G. R., Reddy, R. V., Jayarathna, M., Pughazendi, N., Vidyullatha, S., & Reddy, P. C. S. (2023, May). Web application based Diabetes prediction using Machine Learning. In *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-7). IEEE.

Lambert, J. R., & Perumal, E. (2022). Oppositional firefly optimization based optimal feature selection in chronic kidney disease classification using deep neural network. *Journal of Ambient Intelligence and Humanized Computing*, *13*(4), 1799-1810.

Latha, S. B., Dastagiraiah, C., Kiran, A., Asif, S., Elangovan, D., & Reddy, P. C. S. (2023, August). An Adaptive Machine Learning model for Walmart sales prediction. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 988-992). IEEE.

Lee, C. L., Liu, W. J., & Tsai, S. F. (2022). Development and validation of an insulin resistance model for a population with chronic kidney disease using a machine learning approach. *Nutrients*, *14*(14), 2832.

LK, S. S., Ahmed, S. T., Anitha, K., & Pushpa, M. K. (2021, November). COVID-19 outbreak based coronary heart diseases (CHD) prediction using SVM and risk factor validation. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-5). IEEE.

Major, R. W., Cockwell, P., Nitsch, D., & Tangri, N. (2022). The next step in chronic kidney disease staging: individualized risk prediction. *Kidney international*, *102*(3), 456-459.

Mamatha, B., Rashmi, D., Tiwari, K. S., Sikrant, P. A., Jovith, A. A., & Reddy, P. C. S. (2023, August). Lung Cancer Prediction from CT Images and using Deep Learning Techniques. In 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON) (pp. 263-267). IEEE.

Matsushita, K., Kaptoge, S., Hageman, S. H., Sang, Y., Ballew, S. H., Grams, M. E., ... & Coresh, J. (2023). Including measures of chronic kidney disease to improve cardiovascular risk prediction by SCORE2 and SCORE2-OP. *European journal of preventive cardiology*, *30*(1), 8-16.

Mondol, C., Shamrat, F. J. M., Hasan, M. R., Alam, S., Ghosh, P., Tasnim, Z. & Ibrahim, S. M. (2022). Early prediction of chronic kidney disease: A comprehensive performance analysis of deep learning models. *Algorithms*, *15*(9), 308.

Muthappa, K. A., Nisha, A. S. A., Shastri, R., Avasthi, V. & Reddy, P. C. S. (2023). Design of high-speed, low-power non-volatile master slave flip flop (NVMSFF) for memory registers designs. *Applied Nanoscience*, 1-10.

Raj, A., Tollens, F., Caroli, A., Nörenberg, D., & Zöllner, F. G. (2023). Automated prognosis of renal function decline in ADPKD patients using deep learning. *Zeitschrift für Medizinische Physik*.

Rao, K. R., Prasad, M. L., Kumar, G. R., Natchadalingam, R., Hussain, M. M., & Reddy, P. C. S. (2023, August). Time-Series Cryptocurrency Forecasting Using Ensemble Deep Learning. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 1446-1451). IEEE.

Saha, I., Gourisaria, M. K. & Harshvardhan, G. M. (2022). Classification System for Prediction of Chronic Kidney Disease Using Data Mining Techniques. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2021* (pp. 429-443). Singapore: Springer Singapore.

Sampath, S., Parameswari, R., Prasad, M. L., Kumar, D. A., Hussain, M. M. & Reddy, P. C. S. (2023, December). Ensemble Nonlinear Machine Learning Model for Chronic Kidney Diseases Prediction. In 2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon) (pp. 1-6). IEEE.

Sharma, R., Shrivastava, S., Singh, S. K., Kumar, A., Singh, A. K. & Saxena, S. (2023). EnDL-HemoLyt: Ensemble Deep Learning-based Tool for Identifying Therapeutic Peptides with Low Hemolytic Activity. *IEEE Journal of Biomedical and Health Informatics*.

Sucharitha, Y., Reddy, P. C. S. & Chitti, T. N. (2023, July). Deep learning based framework for crop yield prediction. In *AIP Conference Proceedings*,2548 (1). AIP Publishing.

Sucharitha, Y., Reddy, P. C. S., & Suryanarayana, G. (2023). Network Intrusion Detection of Drones Using Recurrent Neural Networks. *Drone Technology: Future Trends and Practical Applications*, 375-392.

Suneel, S., Balaram, A., Amina Begum, M., Umapathy, K., Reddy, P. C. S., & Talasila, V. (2024). Quantum mesh neural network model in precise image diagnosing. Optical and Quantum Electronics, 56(4), 559.

Teju, V., Sowmya, K. V., Yuvanika, C., Saikumar, K., & Krishna, T. B. D. S. (2021, December). Detection of diabetes melittus, kidney disease with ML. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 217-222). IEEE.

Wang, H., Bowe, B., Cui, Z., Yang, H., Swamidass, S. J., Xie, Y., & Al-Aly, Z. (2022). A deep learning approach for the estimation of glomerular filtration rate. *IEEE Transactions on NanoBioscience*, *21*(4), 560-569.

Bibliographic information of this paper for citing:

D, Baswaraj; K, Chatrapathy; Mudarakola; Lakshmi Prasad, N, Pughazendi; Kiran, Ajmeera; N, Partheeban & Shaker Reddy, Pundru Chandra (2024). Chronic Kidney Disease Risk Prediction Using Machine Learning Techniques. *Journal of Information Technology Management,* 16 (1), 118-134. https://doi.org/10.22059/jitm.2024.96378