Journal of Computer Science

Research Article

Comparative Analysis on Prediction of Chronic Kidney Disease by various Machine Learning Models

[1]Nikhil Verma, [1]Tripti Sharma, [1]Bobbinpreet Kaur and [2]Ayush Dogra

[1]Department of Electronics & Communication, Chandigarh University, Mohali, Punjab, India
[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India

Article history
Received: 30-12-2024
Revised: 26-03-2025
Accepted: 07-04-2025

Corresponding Author:
Nikhil Verma
Department of Electronics & Communication, Chandigarh University, Mohali, Punjab, India
Email: nvvermanikhil@gmail.com

Abstract: Chronic Kidney Disease (CKD) is a prevalent global health concern, which is affecting a number of people in the whole world. Early detection & precise checking of CKD are really important for essential supervision & intervention to prevent disease progression and associated complications. Machine learning (ML) models have emerged as favourable tools for CKD detection due to their ability to analyse complex data patterns and provide predictive insights. This abstract explores the application of various ML techniques in CKD detection, encompassing diverse data sources such as demographic information, clinical biomarkers, medical imaging, and genetic data. We review the performance of different ML algorithms including SVM, Decision Tree, Random Forest, Neural Networks, & ensemble methods in CKD prediction tasks. In recent years, machine learning techniques have garnered considerable attention for their efficacy in medical diagnosis and prognosis, with Random Forest (RF) emerging as a prominent algorithm due to its robustness and versatility. This abstract investigates the application of Random Forest in CKD detection, utilising diverse datasets encompassing demographic information, clinical biomarkers, medical imaging, and genetic profiles. We explore the process of feature selection, model training, and evaluation within the RF framework, highlighting the ability to manage high-dimensional data and nonlinear relationships effectively. Moreover, we review studies showcasing the performance of RF models in predicting CKD onset, progression, and risk stratification, comparing its efficacy with other machine learning approaches. Challenges such as data imbalance, interpretability, and model calibration are also discussed, along with potential strategies to address these issues & enhance the clinical utility of RF-based CKD detection systems. Ultimately, this abstract underscores the promising role of Random Forest as a valuable method in detection and management of Chronic Kidney Disease, offering insights that can potentially improve patient outcomes and healthcare delivery.

Keywords: Machine Learning, Decision Tree, Random Forest, Recursive Feature Selection, AdaBoost

Introduction

Chronic Kidney Disease (CKD) presents a major global health burden, which affects millions of individuals and leads to substantial morbidity and mortality if left undiagnosed or untreated. Early detection and intervention are crucial for mitigating CKD progression and its associated complications. In recent times, Machine Learning (ML) models have been picked as favourable tools for enhancing CKD detection and management by leveraging advanced computational techniques to analyze complex data patterns and make predictive assessments (Saumya *et al.*, 2023). This introduction provides an overview of the implementation of ML models in the context of CKD detection, outlining their potential benefits, challenges and current research trends. It explores the various types of ML algorithms used, the data sources utilized and the implications for improving patient outcomes and healthcare delivery. The application in CKD detection represents a paradigm shift in healthcare, offering the potential to augment traditional diagnostic approaches with data-driven insights that can facilitate early intervention and personalized treatment strategies. Subashini and

Venkatesh (2023) By leveraging the power of ML, healthcare providers can harness vast amounts of patient data, including demographic information, clinical biomarkers, medical imaging and genetic profiles, to develop predictive models capable of identifying individuals at risk of CKD onset, progression, or complications. Furthermore, these models in CKD offer the advantage of adaptability and scalability, enabling continuous refinement and optimization based on evolving datasets and clinical insights (Gracious *et al.*, 2023). This iterative process can ultimately guide the development of increasingly accurate and reliable predictive models tailored to the specific characteristics and needs of CKD populations. This introduction sets the stage for exploring the role of ML models in CKD detection, highlighting their potential benefits, challenges and implications for improving healthcare outcomes. Subsequent sections will delve into the application of models to CKD problems, data sources, model evaluation techniques and practical considerations (Rubini, 2015). The aim of this study is to exercise the use of different Machine learning models such as Decision Tree, Random Forest, AdaBoost, Stochastic Gradient Boosting and the proposed model, namely, Random Forest with Recursive Feature Selection to predict CKD at early stages (Salekin & Stankovic, 2016).

The primary aim of the investigation is to observe the foremost and possible method so that CKD can be predicted with greater accuracy and efficiency (Lichman (n.d)). Early detection will not only help provide sufficient treatment for the patients but also help cut down on personal expenses. The approach to predicting CKD is done by identifying the correct metrics and finding the relevant attributes that can be used to increase the model's accuracy (Haralick *et al.*, 1973). The priority of this study lies in the fact that a detailed analysis of attributes and variables can be done to predict Chronic Kidney Disease. The comparison of all the machine learning models will be assessed, and detailed information needs to be discussed in this study. These models will outline the CKD problem with a detailed analysis of different parameters and values that will ultimately lead to a precise prediction and better outcome (Park *et al.*, 2021). This study clearly describes the different sections that need to be addressed. Section 2 discusses the previous work that has been done in the area of Chronic Kidney disease prediction or simply the literature review that has been done till now in the field of CKD with the assistance of machine learning models (Saba, 2020). The problems and viable solutions are being discussed in the paper. The next section covers the process, with the assistance of the usage of machine learning models on the dataset. The later section outlines the results and proposed model that can be highlighted to predict chronic kidney disease. Section 5 compares the metrics by applying the different machine learning models. The last section describes the conclusion and the work that needs to be done in the future.

Literature Review

It is the section where the work around the prediction and detection of chronic kidney disease has been listed by the different authors. Taylor *et al.* (2018) conducted a single-centre, multi-centre, retrospective investigation of 80387 adults with the symptoms of Urinary Tract Infection (UTI). Here, the authors try to find a solution to the problem of which artificial intelligence algorithm will give the highest specificity and sensitivity in diagnosing UTIs with the help of the clinical manifestations with the availability of urine samples. This study highlights both laboratory and clinical data to develop a model for his UTI prediction using the 6 ML algorithms, namely, Extreme Gradient, Adaptive Boosting, RF, SVM, Elastic Net, and ANN. The model used a dataset of 211 factors and a set of 10 variables, namely, categorized in age, sex, nitrite, UA bacteria, UA blood, UA WBC, epithelial cells, UTI history, and dysuria. It was developed as UTI prediction, which was differentiated from the earlier documentation of urinary tract infection diagnosis and administration of antibiotics. The algorithm that was best achieved for both models was Extreme Gradient Boost, with a value of AUC of 0.904. The limitation of this research is basically the small set of variables that can be expanded to get more accurate results.

Burton *et al.* (2019) attempted to use AI to lessen the diagnostic workload without impacting the detection of urinary tract infections. The aim of this study is to find out which parameters in the urine samples were the sensitive ones and then specify them to assist in the diagnosis of urinary tract infections. It uses two classification methods: heuristic models and machine learning approaches. It tested three different algorithms, namely, Random Forests, Neural Networks, and Extreme Gradient Boosting Reinforcement. In the study, the deduction was made that combinations of WBC and bacterial count are most strongly correlated. Elhoseny *et al.* (2019) in this study, the author compares 3 different classification algorithms with the new algorithm which is known as the Density-based feature selection ACO algorithm, Olex GA, which results in the lowest accuracy among all, i.e. 75% with the implication of the degraded classifier outcomes.

Both the ACO and PCO algorithms indicated an accuracy of about 87.5-85%, respectively, but their performance is much less as compared to the Density-based-Ant Colony Optimisation algorithm. An accuracy of almost 95% is obtained by the application of the D-ACO algorithm, which not only indicates the effective performance but also gives a higher accuracy factor on the dataset on which the algorithm is being applied. Ultimately, the results of different classification algorithms on the CKD dataset in terms of various performance measures proposed that the mentioned D-ACO algorithm be more efficient in the classification of

the dataset. The reason is that it has an advantage over DFS as well as the nature of wrapper methods, which steadily implement the DFS & ACO algorithms repeatedly. This method has an overall accuracy of 95%, but it can be improved by an alternative method or application of other deep learning models.

In this study by Bai *et al.* (2022), the author considers the dataset with 748 chronic kidney disease patients. In this study, 3 Machine Learning models, namely Logistic regression, Naïve Bayes, and random Forest, are applied to the dataset. After applying this dataset, it almost showed similar predictability and greater sensitivity as compared to the Kidney Failure Risk Equation. This equation shows the highest accuracy, specificity, and precision. This study also becomes important to showcase the application of Machine learning for the evaluation of the detection of chronic kidney disease with relevant parameters.

In the study by Debal and Sitote (2022), the author lists the crucial aspects of detecting chronic kidney disease so as to avoid the risk of kidney failure. In this study, the author applied 3 Machine Learning models, namely, Random Forest, SVM, and Decision Tree, with the application of 2 feature selection methods. The evaluation of the models is performed by the 10-fold cross-validation. A total of 19 features are mentioned in this study, with the Random Forest which shows the highest accuracy on the overall basis. The Decision tree shows lower accuracy when compared to random forest. The Support Vector Machine and Random Forest with the method of RFECV give the highest accuracy of about 99.8% for the binary class. Saha *et al.* (2019) highlight the study on the significance of Chronic Kidney Disease and its prediction for early detection and ultimately saving lives. Based on the research, the author applied the Machine Learning models to perform CKD analysis. The models like Random Forest, Naïve Bayes, MLP, LR, and NN are applied. The observation process is done with the help of the CKD dataset obtained from the National Kidney Foundation, Bangladesh, with various Machine Learning models. The result shows that the optimized NN performs better than the other models like Random Forest, NB, MLP, and LR. The study compares the identification of CKD by the eight different ML models. Out of all the machine learning models, Gradient boosting is the one which achieves 98% accuracy, whereas the KNN and Extra Tree classifier both obtain 99% accuracy. The dataset on which the machine learning models are applied is the publicly available UCI dataset. This dataset has 25 attributes, including the target variable. This dataset is also categorized into two broad categories, namely, Yes and No. No work has been found to detect CKD based on the least number of predictors. This dataset has the problem of missing values and imbalanced classes. Preprocessing of data, missing value management, classification methods, and extraction are all steps performed in the process to obtain the desired accuracy. The accuracy obtained with a new performance tuning layered strategy is 98.75%, with 100% sensitivity, having 96.55% for specificity and 99.03% f1-score. Another limitation is the integration of ML models into clinical practice. The major problems, such as data quality, interpretability, transparency of the model, and ethical considerations, need to be highlighted to certify the responsible and productive implementation of ML-based CKD detection systems (Verma & Verma, 2022).

Moreover, the translation of research findings into real-world clinical settings requires robust validation, regulatory approval, and integration with existing healthcare workflows. Despite these challenges, the possible effect of ML models on CKD detection and management is profound (Baidya *et al.*, 2022). By providing clinicians with actionable insights derived from comprehensive data analysis, ML models have the potential to revolutionize CKD care delivery, leading to earlier detection, more targeted interventions and improved patient outcomes (Srivastava *et al.*, 2022). To address these limitations and improve the accuracy of predicting CKD at an early stage, the author is proposing a new method with a combination of Machine learning model and Feature Selection method. In this study, the author put forward the application of Random Forest with Recursive Feature Selection, which paved the way for improved accuracy. After analyzing all the previous work, the random forest model was selected as it shows much-improved accuracy over other models.

Materials and Methods

The dataset used for the study is University of California, Irvine (UCI) Chronic Kidney Dataset, which is accessible on the public domain. This dataset clearly identifies 400 instances with 25 unique attributes. The class labels are identified as ckd, notckd. The evaluation metrics have been obtained on the Anaconda Navigator 2.6.5 by using Jupyter Version 7.0.8 notebook, which is an interactive computing notebook environment. The final code has been written on the Jupyter notebook with the help of Python 3, a high level, general purpose programming language.

This section will outline the methodology that will be used to predict chronic kidney disease by applying the different machine-learning models and the proposed method. The process to achieve a better outcome has certain steps, which include access to relevant and publicly available datasets. This dataset will be preprocessed by removing the missing values using mode imputation. When the preprocessing step is performed, different selected machine learning models according to previous work are applied and compared with the proposed model. This proposed model is further compared with the existing models on the basis of evaluation metrics.

Data Collection

Data collection is the first step in the methodology, which is shown in Figure (1). It is really important to select the relevant dataset for the detection and prediction of CKD. The University of California, Irvine (UCI) CKD dataset, a publicly available dataset, was used for this study, which has a total of 400 instances with 25 attributes, including the target variable. Out of the total 25 attributes, 11 attributes are numeric, and 14 attributes are nominal features. The only purpose is to predict the presence of CKD or not by applying the ML Models (Moreno-Ibarraet, 2013). The different attributes consist of red blood cells, blood glucose, haemoglobin, white blood cells, etc. It also signifies the classes with the presence of CKD and Not CKD. The attributes have a mixture of numerical and categorical variables.

Fig. 1: Methodology

Data Pre Processing

UCI CKD dataset, which is being used for this study, has the problem of missing values and encoding of categorical attributes. The missing values problem is addressed by the usage of mode computation. All missing values are adjusted by the mode formula, i.e., replacing the missing value with the value with the highest frequency of occurrence. The encoding of categorical attributes is solved by applying one hot encoding. These 14 categorical attributes are converted into binary form so that evaluation can be deducted to achieve the desired outcome. This method is performed by adding import OneEncoder in the coding console. Both methods help to complete the preprocessing step and further share the dataset for the application of machine learning models.

Machine Learning Models

Machine learning makes use of statistical models and techniques to make predictions and analyses with the result format. The system is able to find certain deductions on the basis of the datasets (Maier *et al.*, 2019; Segal *et al.*, 2020). The process follows when a large amount of data and dataset is provided to the model so that the accuracy of the models can be judged and some relevant results can be deduced. The dataset provides a lot of new information that can definitely be used to create new information about medical diseases. The models can be analyzed when the various machine-leading models are applied to the UCI dataset (Ifraz *et al.*, 2021). The machine learning models that are applied are Decision Trees, Random Forest, AdaBoost, Stochastic Gradient Boosting, etc. These machine-learning models will understand the different features and obtain different performance metrics (Islam *et al.*, 2020).

Decision Tree

Decision trees are versatile and interpretable machine learning models widely used in various domains for classification and regression tasks. Their intuitive structure, with specialization to handle both categorical and numerical data and straightforward interpretation, makes them valuable tools for data analysis and decision-making. The article provides a detailed overview of decision trees, covering their theoretical foundations, practical applications, model evaluation techniques and recent advancements (Almasoud & Ward, 2019). Through a detailed examination of decision tree algorithms, splitting criteria, pruning techniques and ensemble methods, readers will gain a thorough understanding of how decision trees work and how they can be utilized in real-world scenarios.

Random Forest

This model is a powerful ensemble learning algorithm which has gained widespread presence in various domains of ML because of its versatility, robustness and superior performance. This comprehensive article provides an in-depth exploration of Random Forest, covering its theoretical foundations, practical applications, model evaluation techniques and recent advancements (Bennett, 1992). Through a detailed examination of Random Forest's underlying principles, ensemble learning concepts, decision tree aggregation methods and parameter tuning strategies, readers will gain a profound understanding of how Random Forest operates and how it can be effectively utilized in real-world scenarios.

AdaBoost

Adaptive Boosting is among the machine learning algorithms that are applied for the application of regression and classification tasks. The main objective behind AdaBoost is to focus on the instances that are hard to classify and give them more weight so that subsequent weak learners pay more attention to them. In this technique, each training instance is initially given an equal weight. It assists in classifying the data better than random guessing. The instances that are misclassified by the weak learner are given higher weight in the next iteration so that subsequent weak learners focus more on these difficult instances (Noshad *et al.*, 2019). The final prediction is made by joining the predictions of every weak learner, which is weighted by their individual accuracy. The steps are repeated for a predefined unit of iterations or till an expected level of accuracy is reached. The outcome given by AdaBoost is a weighted combination of every weak learner. This model is effective when the weak learners are simple and have low complexity, namely decision stumps. This simplicity makes AdaBoost less prone to overfitting. However, it can be sensitive to noisy data and outliers (Pandey & Prabhakar, 2016).

Stochastic Gradient Boosting

Stochastic Gradient Boosting, also known as Gradient Boosted Trees or simply Gradient Boosting Machines (GBM), is an add-on to traditional Gradient Boosting that introduces randomness into the training process. It's a popular ensemble learning technique used for regression and classification tasks. In traditional Gradient Boosting, the algorithm starts with an initial prediction and aims to improve upon it with subsequent iterations. The weak learners are added sequentially to the ensemble (Zhang *et al.*, 2019). At each iteration, a new tree is trained to predict the residuals. Stochastic Gradient Boosting introduces randomness in two main ways. Rather than using the entire training dataset to train every tree, a random subset of training data is sampled. This introduces variability into the training process. Then, at each node of each tree, a random subset of features is considered for splitting. This helps prevent overfitting and increases model diversity. Stochastic Gradient Boosting is widely used in practice due to its effectiveness and robustness (Aziz *et al.*, 2020). It often achieves state-of-the-art performance in various machine-learning competitions and real-world applications.

Proposed Work

In this study, the author will be using the Recursive Feature Selection method with Random Forest. This feature selection is used in accordance with the Random Forest model to achieve better accuracy than the other Machine learning models. First, we understand the Recursive Feature Selection method. Recursive Feature Selection is selected as it reduces the problem of overfitting while improving model performance, interpretability and efficiency. Obviously, it will identify the key parameters that will have an impact on the model, which ultimately helps to reduce computational power and, in the future, can be crucial when working with limited hardware or in large-scale applications.

Recursive Feature Selection

Recursive Feature Selection (RFE) is an important technique used for feature selection in machine learning, aimed at identifying the most relevant subset of features that contribute to model performance while discarding irrelevant or redundant attributes (Karkare, 2024; Senan *et al.*, 2021). This comprehensive article provides an in-depth exploration of Recursive Feature Selection, covering its theoretical foundations, practical applications, implementation strategies and recent advancements. Through a detailed examination of RFE's underlying principles, iterative feature elimination process, selection criteria and model evaluation techniques, readers will gain a profound understanding of how RFE operates and how it can be effectively utilized to enhance model interpretability, reduce overfitting and improve predictive performance. Furthermore, this article discusses the challenges and future directions in Recursive Feature Selection research, including approaches for handling high-dimensional datasets, integrating with different machine learning algorithms and addressing computational complexity. By unravelling the intricacies of Recursive Feature Selection, this article aims to empower researchers, practitioners and enthusiasts to leverage this powerful dimensionality reduction technique in their data analysis and modelling endeavours (Munasinghe & Karunanayake, 2021; Priyatno & Widiyaningtyas, 2024). Now, RFE is applied with the Random Forest method to obtain the best attributes that will impact all the parameters like accuracy, F1-score, Recall, etc.

Model Training and Performance

The performance of various ML models on the basis of accuracy, F1-score, precision and recall are discussed in this section. Before comparing all the evaluation metrics, we need to address the issue of class imbalance, which was highlighted in the previous work. Now, this issue is solved by the use of K-Fold cross-validation. In our coding part, we use 5-fold cross-validation, which divides all five equal-sized folds. Out of 5 folds, four folds are used for the training part, and the remaining part is the validation part.

The hyperparameter tuning was optimized by using the import GridSearchCV. GridSearchCV will return the best set of hyperparameters and allow you to assess the model's performance. The hypertuning is done by using the n_estimators = 100, max_depth = 10, min_samples_leaf = 1.

	Features	VIF
5	pus_cell	1.62
2	albumin	1.54
7	diabetes_mellitus	1.49
6	pus_cell_clumps	1.43
1	specific_gravity	1.39
3	sugar	1.24
9	appetite	1.21
4	red_blood_cells	1.16
8	coronary_artery_disease	1.15
10	aanemia	1.14

Fig. 2: Variables obtained by RFE

The attributes obtained by applying the RFE method are presented in Figure (2). These attributes are decided with the value of VIF. The variance inflation factor is used to detect the multicollinearity. These 10 attributes out of the existing 24 attributes are highly important and clearly become a high priority so that a clear decision on CKD detection can be made. It clearly highlights that the proposed model is showing better results than the existing ML models. Random Forest with RFE achieved the highest accuracy of 97.5% among all tested machine-learning models. The detailed accuracy among all the models is shown in Figure (3).

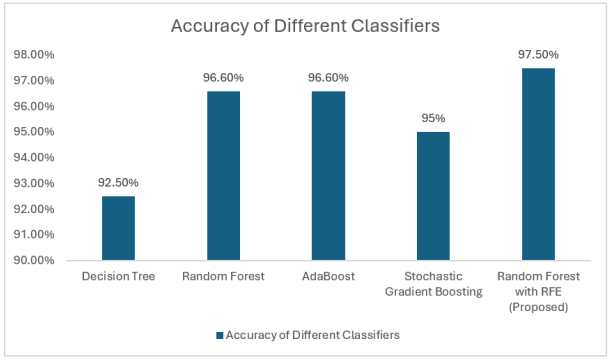

Fig. 3: Accuracy parameter for all models

Precision is a critical metric in machine learning, especially in situations where false positives can lead to significant consequences. Precision is defined by the ratio of True Positives (TP) by the sum of True Positive and False Positives. The proposed methods show a precision of 97.6% as indicated in Figure (4).

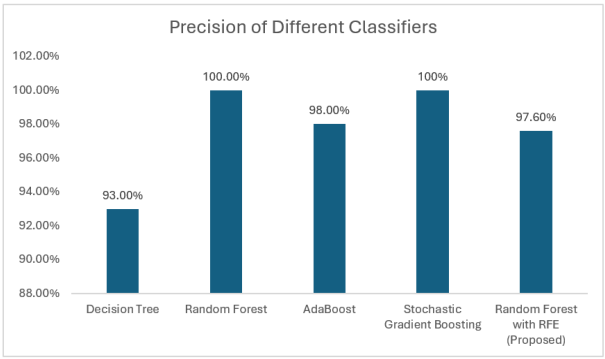

Fig. 4: Precision parameter for all models

Recall, which is also known as sensitivity, is another key metric used to evaluate the performance of classification models, especially when the cost of false negatives is high. Recall is the ratio of True Positives (TP) by the sum of True Positive and False Negatives. The proposed methods show a recall of 97.5% as indicated in Figure (5).

Figure (6) shows the F1-score for the Random Forest with the RFE model is 97.4% and is the highest among all the machine learning models. The rest of the models show a lesser F1 score. F1-score is the harmonic mean of Precision and recall. The precision and recall of the various machine learning models with proposed models have been depicted in Figs. (4-5). Now, we discussed all the evaluation parameters that were of utmost importance for the desired outcome to predict CKD at an early stage.

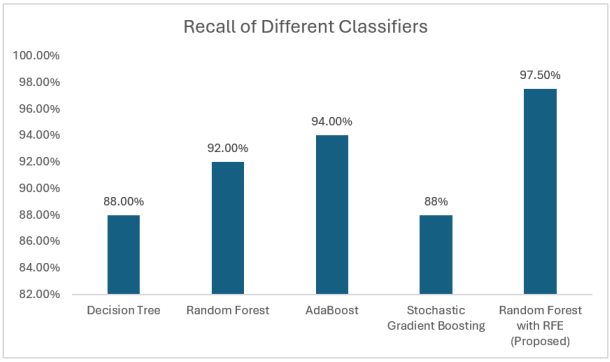

Fig. 5: Recall parameter

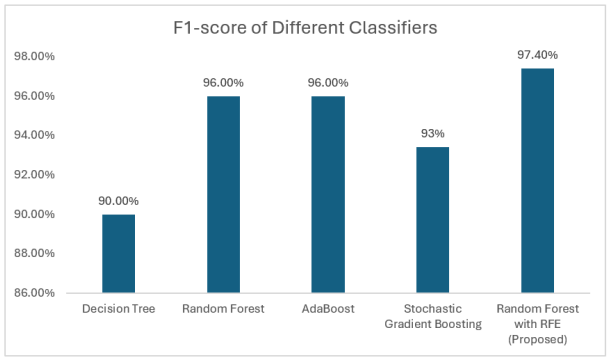

Fig. 6: F1-score parameter for all models

Results and Discussion

The machine learning models and proposed model were applied to the available dataset. It clearly shows that random forest with RFE shows better accuracy than the rest of the machine learning models. Table (1) compares all the relevant parameters which helps us to define the fact that the proposed model shows 97.5% accuracy, which is the highest among all the other machine learning models. The precision stood at 97.6% for the proposed model, with recall showing 97.5% which is highest among all the other models applied on the dataset. The F1-score shows 97.4% which is greater than all the other models.

Table 1: Comparison of various methods with different parameters

Type of Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	92.5	93	88	90
Random Forest	96.6	100	92	96
Ada Boost	96.6	98	94	96
Stochastic Gradient Boosting	95	100	88	93
Random forest with RFE (Proposed)	97.5	97.6	97.5	97.4

Conclusion and Future Work

The various ML algorithms that have been applied to the dataset are Decision Tree, Random Forest, AdaBoost, Stochastic Gradient Boosting, and Random Forest with Recursive Feature Selection. Every model is applied to obtain maximum accuracy and achieve higher values of parameters like recall, precision and F1-Score. Future work will include applying more machine learning models with improved mechanisms so that greater accuracy in the prediction and detection of CKD can be obtained. The future scope is also to apply more feature selection methods with machine learning models. The RFE method can also be compared with other feature selection methods, but computational challenges need to be taken into consideration.

Acknowledgement

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author's Contributions

Nikhil Verma, Tripti Sharma and Bobbinpreet Kaur: Participated in material preparation, data collection and analysis.

Ayush Dogra: Checked final draft and proof reading.

All authors read and approved the final manuscript.

Ethics

This material is the authors' own original work, which has not been previously published elsewhere.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

Almasoud, M., & Ward, T. E. (2019). Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors. *International Journal of Advanced Computer Science and Applications*, *10*(8), 89-96. https://doi.org/10.14569/ijacsa.2019.0100813

Aziz, N., Akhir, E. A. P., Aziz, I. A., Jaafar, J., Hasan, M. H., & Abas, A. N. C. (2020). A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems. *2020 International Conference on Computational Intelligence (ICCI)*, 11-16. https://doi.org/10.1109/icci51257.2020.9247843

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), 8377. https://doi.org/10.1038/s41598-022-12316-z

Baidya, D., Umaima, U., Islam, M. N., Shamrat, F. M. J. M., Pramanik, A., & Rahman, Md. S. (2022). A Deep Prediction of Chronic Kidney Disease by Employing Machine Learning Method. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1305-1310. https://doi.org/10.1109/icoei53556.2022.9776876

Bennett, K. P. (1992). Decision tree construction via linear pro-gramming. *Proceedings of the 4th Midwest Artificial In-Telligence and Cognitive Science Society*, 97-101.

Burton, R. J., Albur, M., Eberl, M., & Cuff, S. M. (2019). Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. *BMC Medical Informatics and Decision Making*, *19*(1), 171. https://doi.org/10.1186/s12911-019-0878-9

Debal, D. A., & Sitote, T. M. (2022). Chronic kidney Disease Prediction Using Machine Learning Techniques. *Journal of Big Data*, *9*(1), 109. https://doi.org/10.1186/s40537-022-00657-5

Elhoseny, M., Shankar, K., & Uthayakumar, J. (2019). Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease. *Scientific Reports*, *9*(1), 9583. https://doi.org/10.1038/s41598-019-46074-2

Gracious, L. A. A., Jasmine, R. M., Pooja, E., Anish, T. P., Johncy, G., & Subramanian, R. S. (2023). Machine Learning and Deep Learning Transforming Healthcare: An Extensive Exploration of Applications, Algorithms, and Prospects. *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, 1-6. https://doi.org/10.1109/gcat59970.2023.10353476

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610-621. https://doi.org/10.1109/tsmc.1973.4309314

Ifraz, G. M., Rashid, M. H., Tazin, T., Bourouis, S., & Khan, M. M. (2021). Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods. *Computational and Mathematical Methods in Medicine*, *2021*, 1-10. https://doi.org/10.1155/2021/6141470

Islam, Md. A., Akter, S., Hossen, Md. S., Keya, S. A., Tisha, S. A., & Hossain, S. (2020). Risk Factor Prediction of Chronic Kidney Disease based on

Machine Learning Algorithms. *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 952-957. https://doi.org/10.1109/iciss49785.2020.9315878

Karkare, P. (2024). The Artistry of Stochastic Gradient Boosting and Gradient Boosting Classifiers in Chronic Renal Disease Classification. *Journal of Electrical Systems*, *20*(3), 2405-2415. https://doi.org/10.52783/jes.4086

Kumar, N., Narayan Das, N., Gupta, D., Gupta, K., & Bindra, J. (2021). Efficient Automated Disease Diagnosis Using Machine Learning Models. *Journal of Healthcare Engineering*, *2021*, 1-13. https://doi.org/10.1155/2021/9983652

Lichman, M. (n.d.). UCI Machine Learning Repository. *UCI Machine Learning Repository*. https://archive.ics.uci.edu/

Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift Für Medizinische Physik*, *29*(2), 86-101. https://doi.org/10.1016/j.zemedi.2018.12.003

Moreno-Ibarra, M.-A., Villuendas-Rey, Y., Lytras, M. D., Yáñez-Márquez, C., & Salgado-Ramírez, J.-C. (2021). Classification of Diseases Using Machine Learning Algorithms: A Comparative Study. *Mathematics*, *9*(15), 1817. https://doi.org/10.3390/math9151817

Munasinghe, K., & Karunanayake, P. (2021). Recursive Feature Elimination for Machine Learning-based Landslide Prediction Models. *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 126-129. https://doi.org/10.1109/icaiic51459.2021.9415232

Noshad, Z., Javaid, N., Saba, T., Wadud, Z., Saleem, M. Q., Alzahrani, M. E., & Sheta, O. E. (2019). Fault Detection in Wireless Sensor Networks through the Random Forest Classifier. *Sensors*, *19*(7), 1568. https://doi.org/10.3390/s19071568

Pandey, P., & Prabhakar, R. (2016). An analysis of machine learning techniques (J48 & AdaBoost)-for classification. *2016 1st India International Conference on Information Processing (IICIP)*, 1-6. https://doi.org/10.1109/iicip.2016.7975394

Park, D. J., Park, M. W., Lee, H., Kim, Y.-J., Kim, Y., & Park, Y. H. (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Scientific Reports*, *11*(1), 7567. https://doi.org/10.1038/s41598-021-87171-5

Priyatno, A. M., & Widiyaningtyas, T. (2024). A Systematic Literature Review: Recursive Feature Elimination Algorithms. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *9*(2), 196-207. https://doi.org/10.33480/jitk.v9i2.5015

Rubini, L. (2015). Early stage of chronic kidney disease UCI machine learning repository. *Chronic Kidney Disease*.

Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. *Journal of Infection and Public Health*, *13*(9), 1274-1289. https://doi.org/10.1016/j.jiph.2020.06.033

Saha, A., Saha, A., & Mittra, T. (2019). Performance Measurements of Machine Learning Approaches for Prediction and Diagnosis of Chronic Kidney Disease (CKD). *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, 200-204. https://doi.org/10.1145/3348445.3348462

Salekin, A., & Stankovic, J. (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes. *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 262-270. https://doi.org/10.1109/ichi.2016.36

Saumya, L., Manimozhi, I., & Swathi. (2023). A Survey on Machine Learning Algorithms for the Detection of Chronic Kidney Disease. *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 1832-1837. https://doi.org/10.1109/icacrs58579.2023.10404425

Segal, Z., Kalifa, D., Radinsky, K., Ehrenberg, B., Elad, G., Maor, G., Lewis, M., Tibi, M., Korn, L., & Koren, G. (2020). Machine learning algorithm for early detection of end-stage renal disease. *BMC Nephrology*, *21*(1), 518. https://doi.org/10.1186/s12882-020-02093-0

Senan, E. M., Al-Adhaileh, M. H., Alsaade, F. W., Aldhyani, T. H. H., Alqarni, A. A., Alsharif, N., Uddin, M. I., Alahmadi, A. H., Jadhav, M. E., & Alzahrani, M. Y. (2021). Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques. *Journal of Healthcare Engineering*, *2021*, 1-10. https://doi.org/10.1155/2021/1004767

Srivastava, S., Yadav, R. K., Narayan, V., & Mall, P. K. (2022). An ensemble learning approach for chronic kidney disease classification. *Journal of Pharmaceutical Negative Results*, *10*(13), 2401-2409.

Subashini, N. J., & Venkatesh, K. (2023). CKD Prediction using Mixed Dataset Concept and Deep Learning: A Combined Approach for Accurate Diagnosis. *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, 1-6. https://doi.org/10.1109/rmkmate59243.2023.10369225

Taylor, R. A., Moore, C. L., Cheung, K.-H., & Brandt, C. (2018). Predicting urinary tract infections in the emergency department with machine learning. *PLOS ONE*, *13*(3), e0194085. https://doi.org/10.1371/journal.pone.0194085

Verma, V. K., & Verma, S. (2022). Machine learning applications in healthcare sector: An overview. *Materials Today: Proceedings*, *57*, 2144-2147.

https://doi.org/10.1016/j.matpr.2021.12.101

Zhang, Y., Ni, M., Zhang, C., Liang, S., Fang, S., Li, R., & Tan, Z. (2019). Research and Application of AdaBoost Algorithm Based on SVM. *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 662-666. https://doi.org/10.1109/itaic.2019.8785556