Original article

Development and optimization of predictive machine learning models for patients with chronic kidney disease

DOI: 10.5377/alerta.v8i4.21163

David A. Tejada

National Health Observatory, National Institute of Health, San Salvador, El Salvador.

*Correspondence
david.tejada@salud.gob.sv

1. 0000-0003-2502-1433

OPEN ACCESS

Desarrollo y optimización de modelos predictivos de *machine learning* para pacientes con enfermedad renal crónica

Suggested citation:
Tejada DA. Development and optimization of predictive machine learning models for patients with chronic kidney disease.
Alerta. 2025;8(4):354-365.
DOI: 10.5377/alerta.v8i4.21163

Editor:
David Rivera.

Received:
June 7, 2024.

Accepted:
August 15, 2025.

Published:
October 31, 2025.

Author contribution:
DAT: study conception, manuscript design, literature search, data collection, data or software management, data analysis, writing, revising, and editing.

Conflicts of interest:
No conflicts of interest.

Abstract
Introduction. Artificial intelligence has evolved, becoming an essential tool for analyzing information in medicine. **Objective.** Develop, optimize, and compare the performance of different machine learning models for analyzing factors associated with chronic kidney disease. **Methodology.** Machine learning models were developed to predict chronic kidney disease. We constructed logistic regression models, support vector machine, random forest and decision trees; then we evaluated performance metrics.. The model with the lowest performance was selected and optimized using conventional machine learning techniques, hyperparameter tuning, and advanced approaches. We evaluated performance using accuracy, area under the curve, sensitivity, specificity, 95 % confidence intervals, and p-values < 0.05. **Results.** Logistic regression stood out for its accuracy (85.29 %) and sensitivity (95.65 %), and support vector machines for the area under the curve (92.09 %). Random forest achieved a balance between accuracy (82.35%) and area under the curve (90.32 %). The tree showed high specificity (90.91 %) and positive predictive value (90 %). After hyperparameter tuning, the decision tree achieved an accuracy of 80.39%. **Conclusion.** Logistic regression, support vector machines, and random forest performed best with conventional training. Machine learning techniques allowed the performance of the models to be adjusted and optimized. Male sex, high blood pressure, and exposure to pesticides were identified as key factors associated with chronic kidney disease.

Keywords
Artificial Intelligence, Predictive Models, Machine Learning, Chronic Kidney Disease.

Resumen
Introducción. La inteligencia artificial ha evolucionado, convirtiéndose así en una herramienta esencial para el análisis de información en medicina. **Objetivo.** Desarrollar, optimizar y comparar el desempeño de distintos modelos de *machine learning* para analizar factores asociados a la enfermedad renal crónica. **Metodología.** Se desarrollaron modelos de *machine learning* para predecir la enfermedad renal crónica. Se construyeron modelos de regresión logística, máquinas de soporte vectorial, bosque aleatorio y árboles de decisión, y se evaluaron las métricas de rendimiento. Se seleccionó el modelo con menor desempeño y se optimizó mediante técnicas convencionales de aprendizaje automático, ajuste de hiperparámetros y enfoques avanzados. El desempeño se evaluó utilizando la exactitud, área bajo la curva, sensibilidad, especificidad, intervalos de confianza al 95 %, y valores de p < 0,05. **Resultados.** La regresión logística se destacó por su exactitud (85,29 %) y sensibilidad (95,65 %) y las máquinas de soporte vectorial por el área bajo la curva (92,09 %). El bosque aleatorio logró un equilibrio entre exactitud (82,35 %) y área bajo la curva (90,32 %). El árbol de decisión presentó una alta especificidad (90,91 %) y valor predictivo positivo (90 %). Tras el ajuste de hiperparámetros, el árbol de decisión alcanzó una exactitud de 80,39 %. **Conclusión.** La regresión logística, las máquinas de soporte vectorial y el bosque aleatorio presentaron el mejor desempeño con entrenamiento convencional. Las técnicas de *machine learning* permitieron ajustar y optimizar el desempeño de los modelos. Se identificó al sexo masculino, la hipertensión arterial y la exposición a plaguicidas como factores determinantes de la enfermedad renal crónica.

Palabras clave
Inteligencia Artificial, Modelos Predictivos, Aprendizaje Automático, Enfermedad Renal Crónica.

Introduction

Over the years, artificial intelligence (AI) has evolved significantly, becoming an essential tool for enhancing information analysis across disciplines such as medicine.[i] Since its beginnings in the 1950s, AI has advanced from basic rule-based systems to complex machine learning (ML) and deep learning (DL) algorithms,[ii] enabling the analysis of large databases with greater accuracy and speed, while facilitating research, and optimizing processes in fields such as medicine.[iii]

In health sciences, the application of ML and DL models has transformed the way diseases are diagnosed, treated, and predicted.[iv] These models allow for deeper exploration of medical data, improve early detection of disease, allow for personalized treatments, and improve the efficiency of medical interventions.[v]

Machine learning algorithms, such as regression and decision trees, provide robust tools for the classification and prediction of medical data.[vi] Meanwhile, DL models, such as convolutional and recurrent neural networks, have proven to be excellent at processing images and sequences (text, audio, or time series), achieving significant advances in the interpretation of medical images and the analysis of clinical records.[vii]

In the field of chronic diseases, the use of these models is becoming increasingly indispensable for advancing research and improving medical management. They also allow researchers to more accurately identify complex patterns and predict the progression of priority conditions such as chronic kidney disease (CKD), thereby providing a stronger foundation for future research and clinical advances.[viii-x]

In this context, the use of ML models in the analysis of secondary data, such as databases, electronic records, and previous studies, enables AI algorithms to extract valuable information and patterns that cannot be detected with conventional models. This has facilitated not only the identification of factors associated with disease but also the monitoring of disease progression and the personalization of treatment.[v,xi]

Therefore, in this study, supervised ML models were applied to process patient data with factors associated CKD. The advantages and accuracy of each model were analyzed, as well as different optimization techniques, which allowed the establishment of a complementary methodology for the analysis and investigation of this disease. The objectives of this research were to develop, optimize, and compare the performance of ML models in patients with CKD.

Methodology

Machine learning models were developed to predict chronic kidney disease using data from the study "Risk factors associated with chronic kidney disease in Chalatenango, El Salvador."[xii] The database comprised 174 records, including 58 CKD cases and 116 non-CKD cases. The baseline included 26 independent variables and one dependent variable, "chronic kidney disease." Logistic regression (LR), support vector machines (SVM), random forest (RF), and decision tree (DT) models were applied.

A correlation matrix with a threshold of ± 0.7 was applied to identify and remove highly correlated predictors, thereby reducing multicollinearity in linear models.

The variables considered included: area, age, sex, formal education, educational level, marital status, work activity, farmer condition, harmful tobacco and alcohol use, unhealthy intake of salt, water, fruits and vegetables, sedentary lifestyle, exposure to high temperatures, diabetes *mellitus*, hypertension, obesity, recurrent urinary tract infections, chronic pesticide exposure, and harmful consumption of non-steroidal anti-inflammatory drugs (NSAIDs).

Predictor variables were normalized to ensure a homogeneous scale, facilitate comparisons, and improve ML algorithm convergence. Also, the same variables were included in all models to ensure similar and comparable conditions between algorithms.

The data was split into 80 % training and 20 % testing. Class balance was evaluated based on the distribution of the outcome variable, comparing proportions of CKD and non-CKD cases with distribution plots and Chi-Square tests.

To address class imbalance, oversampling was applied to the training set with the "ROSE" package (Random OverSampling Examples) in RStudio with the "ovun.sample" function and the "over" method, increasing minority-class samples to balance the training dataset.

To train the ML models, the specific parameters of each algorithm were adjusted. Logistic regression was configured using the standard method with the "binomial" family, this addressing the binary classification problem to determine the presence of ERC. Support vector machines with "kernel radial" were implemented using the "svmRadial" method.

For the RF, the "rf" method was used, which integrates multiple decision trees to improve predictive accuracy. Likewise, an DT model was trained using the "rpart" method. In all cases, ten-fold cross-validation was applied to ensure accurate and standardized evaluation, in order to prevent model overfitting.

The performance of the models was evaluated using model accuracy, which measures the proportion of correct predictions made by each model, as well as 95 % confidence intervals for accuracy. We also analyzed the non information rate (NIR), defined as the proportion of observations belonging to the majority class, which represents the accuracy that would be obtained by classifying all cases as belonging to that class.

This metric was used as a reference to evaluate the actual gain of the model. The accuracy obtained by the model was statistically compared with the NIR using a hypothesis test, considering a value of $p < 0.05$ considered the threshold for statistical significance.

The Kappa coefficient was used to measure the agreement between the model predictions and the actual values, considering the possibility that some of the agreement could be due to chance, which allows for a more accurate assessment of the model's performance. We used the McNemar test to evaluate the error patterns in the classification of each model compared to the NIR. This test contrasts cases classified correctly by the model and incorrectly by the NIR, and vice versa. A p-value < 0.05 was interpreted as evidence of systematic differences in model performance compared to classification by the majority model, indicating systematic biases or worse performance compared to the reference model.

In addition, sensitivity and specificity were estimated, along with positive and negative predictive values. Finally, the overall performance of the models was determined using balanced metrics that evaluated different aspects of performance. Balanced accuracy averages sensitivity and specificity to consider both classes in unbalanced situations equally. The F1 score combines precision and sensitivity using the harmonic mean to balance between false positives and false negatives. The area under the curve (AUC) measures the overall ability of the model to discriminate between classes across all possible decision thresholds, where a value closer to unity indicates better discriminatory performance.

The model with the lowest initial performance was selected to highlight different methodological possibilities for model optimization and improve the quality of the analysis. An DT developed with the "rpart" package was chosen, which was configured with the default hyperparameters of the RStudio package and presents a base training, which had the lowest performance among the ML models evaluated. Different optimization techniques and methodologies were applied to this model to improve its performance. We developed three additional DT models, the first one considered the threshold for statistical significance "Train-Test Split," dividing the data into 70 % for training and 30 % for testing. The hyperparameters were left at their default values, meaning that no specific adjustments were made to the tree structure.

For the second decision tree, we used a model with hyperparameter optimization, employing the "Grid Search" technique to adjust the hyperparameters, focusing on the complexity parameter (CP), which automatically evaluated a range of CP values from 0.001 to 0.1 in increments of 0.005, selecting the value that found the best balance between the performance metrics. Five-fold cross-validation was applied to ensure robustness under different configurations. CP tuning and cross-validation both help prevent overfitting: CP controls tree pruning, limiting complexity and reducing overfitting to the training data, while cross-validation evaluates performance across multiple subsets, improving generalizability.

In the third model, we implemented a tree with advanced and combined techniques (advance techniques model), using a "Train-Test Split" with a 70 % split for training and 30 % for testing. Then, cross-validation was applied with ten folds, dividing the data into ten parts to validate the model. Probability Estimation was also enabled to generate class probabilities, and binary classification metrics were used to evaluate the model's performance.

Finally, the optimized model was selected by adjusting hyperparameters that showed the best balance between performance and lower risk of overfitting, and this was used to make new predictions about CKD.

There was no missing data in this research, so no imputation techniques were applied. Performance was evaluated using the metrics described above. The processing and analysis of the information was carried out using RStudio version 4.3.2.

The research was conducted in accordance with Good Clinical Practices. The database was coded to preserve participant confidentiality, and the study protocol was approved by the Ethics Committee of the National Institute of Health under registration number CEINS/2024/006.

Results

Performance of the machine learning models

e carried out a comparative analysis of the performance of four ML models. (Table 1). Regarding their overall performance, LR achieved the highest accuracy at 0.853 (95 % CI = 0.689-0.950), correctly classifying 85.29 % of CKD cases. RF and SVM also performed well, with accuracies of 0.823 (95 % CI = 0.655-0.932) and 0.794 (95 % CI = 0.621-0.913), respectively.

Table 1. Comparative chart of the performance of different machine learning models

Metric	Logistic Regression	Support Vector Machines	Random Forest	Decision Tree
Overall Performance				
Accuracy	0.853	0.794	0.823	0.559
95 % Confidence Interval*	0.689 - 0.950	0.621 - 0.913	0.6547 - 0.932	0.379 - 0.728
NIR	0.676	0.676	0.676	0.676
p-value (Accuracy > NIR)**	0.017	0.097	0.044	0.949
Kappa	0.638	0.518	0.577	0.230
McNemar's Test p-value	0.371	1	0.683	0.002
Sensitivity and Specificity				
Sensitivity	0.956	0.870	0.913	0.391
Specificity	0.636	0.636	0.636	0.909
Predictive Values				
Positive Predictive Value	0.846	0.833	0.840	0.900
Negative Predictive Value	0.875	0.700	0.778	0.417
Prevalence and Detection				
Prevalence	0.676	0.676	0.676	0.676
Detection Rate	0.647	0.588	0.618	0.265
Detection Prevalence	0.765	0.706	0,735	0.294
Balanced Metrics				
Balanced Accuracy	0.796	0.753	0.775	0.650
Area Under the Curve (AUC)	0.852	0.921	0.903	0.682
F1 Score	0.897	0.851	0.876	0.545

Source: Based on data processing across different models.

* Confidence interval for accuracy.

** p-value [Accuracy > NIR]: p-value associated with the hypothesis that the model's accuracy is greater than the No Information Rate (NIR).

On the other hand, DT showed the lowest accuracy of the four models, with a value of 0.559 (95 % CI = 0.379-0.728), correctly classifying 55.88 % of the cases, which is lower than the no-information rate (67.65 %).

When analyzing the statistical significance of the models and testing the equality hypothesis against NIR, LR, and RF proved superior and statistically significant ($p < 0.05$).

In terms of sensitivity and specificity, LR and RF achieved the highest sensitivity, 0.956 and 0.91, respectively, indicating that they have the greatest ability to correctly identify CKD patients. In contrast, DT showed the lowest sensitivity (0.391) but the highest specificity (0.909), indicating better performance in identifying patients without CKD.

Figure 1 shows the ROC curve for each model, along with the corresponding AUC. In addition, a comparison between the models is included to evaluate their relative performance. When balanced metrics were analyzed, SVM and RF obtained the best results, with balanced accuracies of 0.75 and 0.775 and AUC values of 0.921 and 0.903, respectively, suggesting that these models achieve a good balance between sensitivity and specificity. Although DT had the worst overall accuracy, its balanced accuracy and AUC surpassed LR, indicating that this model might be more suitable in contexts requiring better balance between positive and negative case identification.

McNemar's test results showed that LR, SVM, and RF did not present significant differences in their classification error patterns compared with NIR ($p > 0.05$), indicating no systematic bias in their predictions.

Optimization of decision tree models

We performed a comparative analysis of the performance of four DT models with varying adjustment levels, from a baseline model to one developed with advanced techniques (Table 2). Overall, accuracy increased progressively as the adjustment level improved (Figure 2).

The advanced adjustment model achieved the highest performance, with an accuracy of 1.0 (95 % CI = 0.930-1), correctly classifying 100 % of cases. The conventional adjustment and hyperparameter optimization models achieved accuracies of 0.725 (95 % CI = 0.583-0.841) and 0.804 (95 % CI = 0.669-0.902), respectively, both were statistically significant when tested against NIR ($p < 0.05$).

Regarding sensitivity and specificity, the advanced adjustment model scored 1.0 in both metrics. The conventional adjustment and hyperparameter optimization models presented sensitivities of 0.676 and 0.794, and specificities of 0.823 respectively. In contrast, the baseline model had the lowest sensitivity (0.391) but the highest specificity (0.909).

When analyzing balanced metrics, the advanced adjustment model obtained 1.0 for both balanced accuracy and AUC. The conventional adjustment and hyperparameter-optimized models presented balanced accuracies of 0.750 and 0.809 and AUC values of 0.702 and 0.777, respectively. The baseline model showed lower values: 0.650 for balanced accuracy and 0.682 for AUC.

The McNemar test shows that the base model presents statistically significant differences in the pattern of classification errors compared to the NIR, $p = 0.002$. On the other hand, the models with conventional adjustment, hyperparameter optimization, and advanced adjustment techniques did not present significant differences ($p > 0.05$). This suggests that these models do not have systematic biases in their predictions and classification errors.

Decision tree model performance

The hyperparameter-optimized DT achieved an accuracy of 80.8 %, sensitivity of 79.4 %, and an AUC of 0.777. This model mainly used sex and hypertension as main factors for classification (Figure 3). It began by identifying patients without CKD as the predominant category; it then classified individuals based on sex, hypertension, fruit and vegetable consumption, and chronic pesticide exposure.

For women, the DT distinguished between those without hypertension or other risk factors and those with hypertension, also considering fruit and vegetable intake. For men, classification was based on hypertension and chronic pesticide exposure. The analysis determined that hypertension and, in men, chronic pesticide exposure were important predictors of CKD.

Importance of predictor variables in tree decision-making

The importance percentages of variables in a DT indicate how relevant each variable is to model prediction, reflecting their relative contribution to overall performance. These percentages are calculated according to the frequency and quality of splits made by each variable in the tree nodes. The more frequent and effective a variable is in improving predictions, the higher its importance.

In the CKD prediction analysis, importance percentages highlighted the most relevant variables for identifying the disease. Male sex (21.62 %) was the most important predictor, followed by chronic pesticide exposure (17.99 %) and high-temperature activities (17.62 %).

These factors most strongly contributed to the model's performance in identifying CKD cases and understanding key elements associated with the disease.

Discussion

This study applied different supervised learning models to analyze patient data with factors associated with CKD in the department of Chalatenango, El Salvador. The findings demonstrate the potential of these models to enhance analytical accuracy and optimize health system management, as has been reported in previous studies.[x-xi]

Overall, the logistic regression model showed the best performance, with higher accuracy and F1 score than the other models. These findings are consistent with previous research highlighting the usefulness of logistic regression in predicting chronic diseases such as CKD.[xii,xiii] This can be atributed to its ability to model the relationship between a binary response variable and one or more predictor variables, demonstrating its effectiveness when applied to biometric data with similarly adjusted models.[xi]

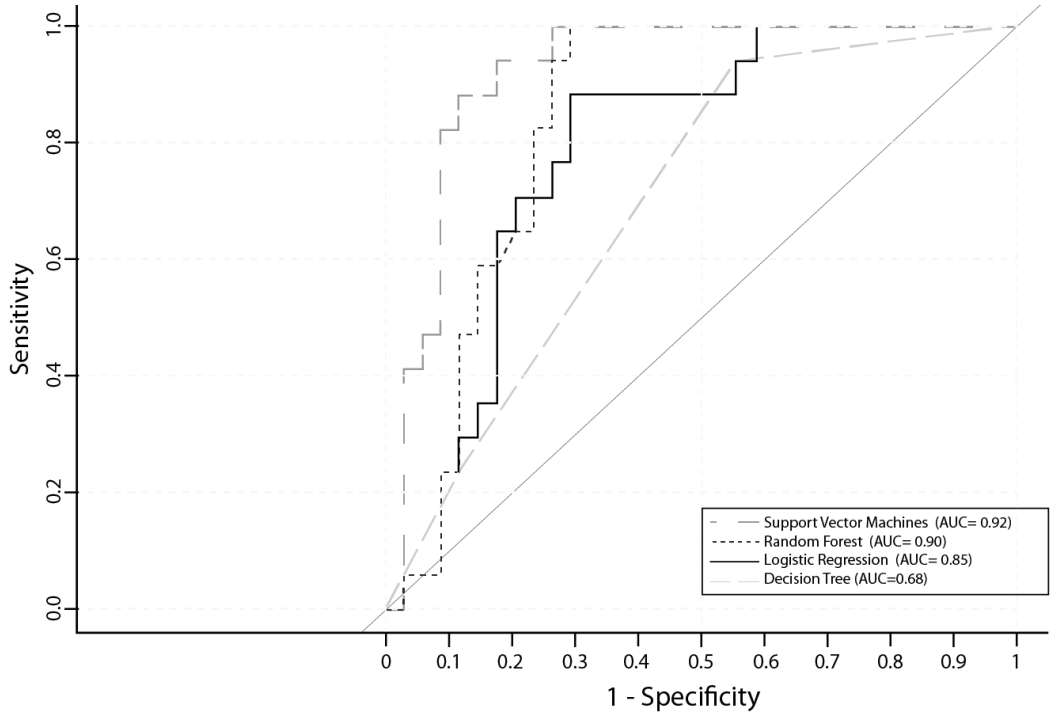

Figure 1. Comparison of the performance of machine learning models using the ROC curve

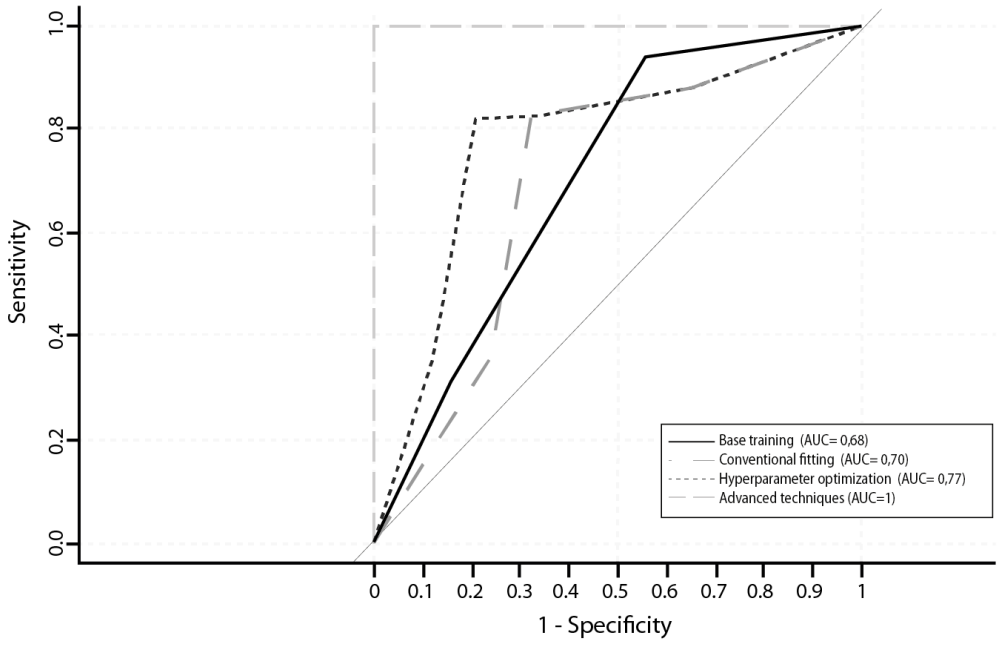

Figure 2. Comparison of the performance of the decision tree models according to their optimization using the ROC curve

Table 2. Comparative chart of decision tree models' performance according to their adjustment

Metric	Model with baseline training	Model with conventional adjustment	Model with hyperparameter optimization	Model with advanced techniques
Overall performance				
Accuracy	0.559	0.725	0.804	1
95 % confidence interval*	0.379 - 0.728	0.583 - 0.841	0.669 - 0.902	0.930 - 1
NIR	0.667	0.667	0.667	0.667
P-value accuracy > NIR**	0.948	0.023	0.023	< 0.001
Kappa	0.230	0.447	0.583	1
McNemar test p-value	0.002	0.062	0.343	1
Sensitivity and specificity				
Sensitivity	0.391	0.676	0.794	1
Specificity	0.909	0.823	0.823	1
Predictive values				
Positive predictive value	0.9	0.885	0.9	1
Negative predictive value	0.417	0.56	0.667	1
Prevalence and detection				
Prevalence	0.676	0.667	0.667	0.667
Detection rate	0.265	0.451	0.529	0.667
Detection prevalence	0.294	0.510	0.588	0.667
Balanced metrics				
Balanced accuracy	0.650	0.750	0.809	1
Area under the curve (AUC)	0.682	0.702	0.778	1
F1 score	0.541	0.767	0.843	1

*Confidence interval of accuracy.
*p-value [accuracy > NIR]: p-value associated with the hypothesis that the model's accuracy is greater than the No Information Rate (NIR).

However, SVM and RF showed the best overall performance, with the highest AUC values, surpassing logistic regression. SVM seeks the best hyperplane to separate classes, enhancing classification accuracy in binary problems.[xiv,xv] RF, on the other hand, combines multiple DTs to improve predictive accuracy and reduce overfitting, thereby improving generalizability and predictive performance.[xvi-xviii]

Selecting the most appropriate model is crucial for accurate data analysis. Each model offers specific advantages depending on the prioritized metric. Accuracy is essential when the goal is to maximize the proportion of correct predictions, which is critical for reliable and safe clinical decision-making. Sensitivity becomes more important when the objective is to detect all positive cases, particularly in prioritized disease diagnoses, where missing a case requiring medical care is unacceptable.[xix]

Specificity is also critical to minimize false positives and avoid unnecessary treatments that may cause harm or discomfort to patients.[xx]

In general, an optimal balance between sensitivity and specificity is required, which can be assessed using composite metrics such as balanced accuracy, F1 score, and AUC. These allow for a more comprehensive evaluation of model performance, considering both detection and discrimination capacity between positive and negative cases.[xx]

The initial DT model showed relatively low performance in both in accuracy and AUC. However, applying hyperparameter tuning and cross-validation significantly improved its performance. These findings support the importance of optimizing ML models through hyperparameter adjustment and cross-validation, as highlighted in previous studies.[xii,xxi,xxii] Hyperparameter tuning involves testing different configurations to find the optimal combination that maximizes performance from training and validation data. [xii,xxi,xxii]

One objective of this research was to demonstrate optimization techniques applied to different models using the R programming language. However, multiple languages, platforms, and environments allow for this type of analysis. The choice of strategy depends on factors such as researcher expertise, problem complexity, and available resources, without limiting the applicability of the presented methods. The flexibility of these approaches facilitates adaptation to different contexts, programming languages, and data, favoring reproducibility in future studies.

Moreover, the ML results were consistent with the literature and existing theories on CKD, highlighting its multifactorial nature. [xxiii] The models identified variables such as sex,[xxiv] hypertension,[xxv] and agrochemical exposure, identified as independent predictors in the DT model for CKD,[xxvi] as documented in similar studies.[xxvii,xxviii]

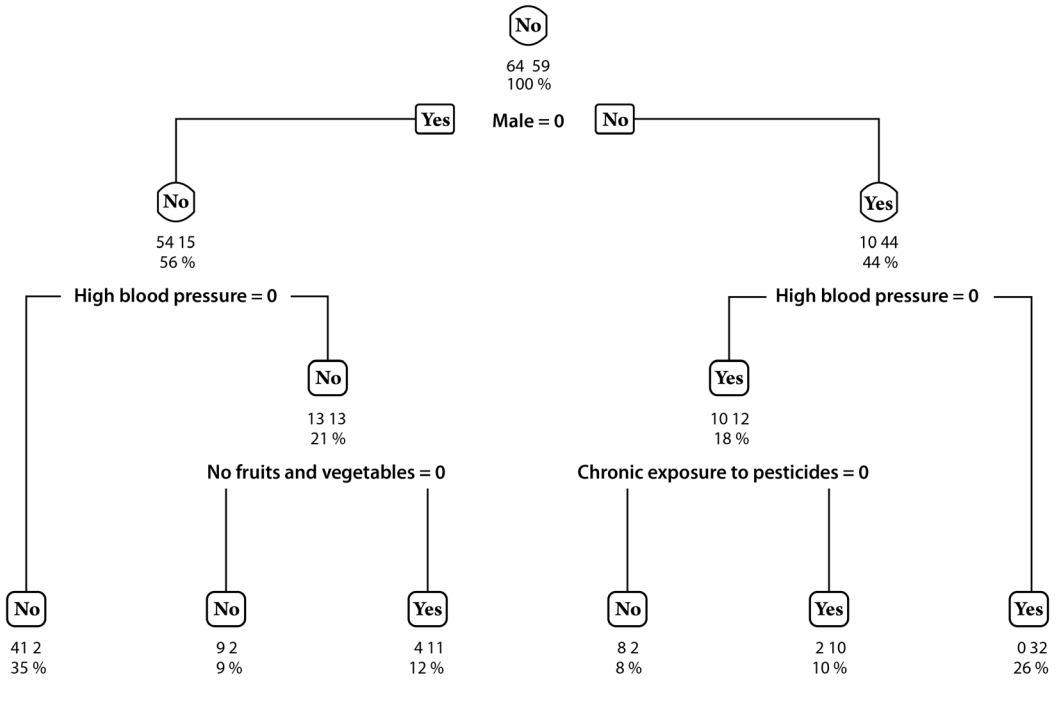

Figure 3. Comparison of the performance of the decision tree models according to their optimization using the ROC curve

However, further research is required to more precisely estimate the risk associated with such exposures and to incorporate additional variables for developing more complex ML models, as in other studies where disease progression has been determined from individuals' using longitudinal or historical patient data.[xxix]

This study highlighted the protective effects of fruit and vegetable consumption[xxx,xxxi] and differences in CKD prevalence by sex. However, its main contribution lies in applying ML models to national data analysis, allowing a more precise and comprehensive evaluation of factors associated with CKD. This approach adapts to different sensitivity, specificity, and accuracy criteria, enabling a deeper understanding of disease-related contributors.[xxvii,xxviii]

Another important aspect of this study is model interpretability. While DTs are relatively easy to interpret due to their visual structure and decision rules, LR, RF, and SVM are more complex and may be more difficult for non-specialist professionals to interpret. Nevertheless, these models can achieve higher predictive performance.[xvi] In this sense, Hendrik Blockeel *et al.* suggest that DTs may be preferable in clinical environments due to their visualization and interpretability.[xxxiii] Nevertheless, interpretability should not be the only criterion when choosing a model. In some cases, higher predictive performance is necessary, as in this study, where SVM and RF may be preferable when greater accuracy is required, particularly for identifying associated factors and supporting decision making.[xxxiv]

Including additional variables, such as laboratory test results or medical imaging, could enrich analysis and improve prediction accuracy. ML models can leverage multiple data sources, both structured and unstructured.[xxxi,xxxii] Incorporating these could provide additional information for CKD diagnosis and prediction, further improving model performance and relevance.[xxix]

This study highlighted important findings in CKD research using ML. A model was developed and compared with another that achieved an AUC of 0.80 using RF and neural networks to predict CKD.[xii] This research achieved similar or even superior results, demonstrating strong predictive power and performance in CKD analysis.[xxxv]

For future research, exploring more advanced methods, such as neural networks, is essential, as they may substantially improve performance. However, these techniques require higher computational power and pose additional interpretation challenges.[xii]

Although the sample size of 178 observations may be considered small in some contexts, literature supports the use of similar or even smaller sizes when applying techniques such as cross-validation, oversampling, and regularization. These maximize data use, improve generalization, and produce reliable results.[xxxvi,xxxvii] Sample size adequacy depends on analysis complexity and variable count: for complex models, a small size may be insufficient, while in simpler analyses with proper techniques, it may suffice for solid results.[xxxviii]

Various studies have validated and applied models with small datasets through simulation, using methods such as Monte Carlo,[xxxix] bootstrap, and generative adversarial networks, which enable controlled expansion of synthetic data to improve training and validation.[xl,xli] These methodologies are applied in genetics, economics, computer vision, and health, proving especially useful when data collection is limited or costly. Their adoption is crucial to strengthening research quality and scope in health.

Studies show that when data exhibit good discriminatory power, effect size and model accuracy increase with sample size, while variance decreases.[xxxviii] Nevertheless, in datasets with low-discrimination power, neither effect size nor accuracy improves significantly with larger samples. An adequate sample size is one that achieves an effect size ≥ 0.5 and ML accuracy ≥ 80 %, since further increases do not yield significant improvements, optimizing the cost-benefit balance.[xxxviii]

Literature also indicates that small sizes may overestimate accuracy due to overfitting, while very large sizes increase costs and time without proportional gain in reliability. Thus, defining an adequate sample size is essential to obtaining reliable and efficient results. In this study, the chosen sample size is justified as it met the effect size and accuracy criteria, ensuringa balance between reliability and available resources.[xxxviii]

The main limitation of this study is the use of secondary data, which carries certain risks. However, since the original research aimed to identify factors associated with CKD, this analysis allowed further exploration of aspects not addressed by traditional methods. Although proper validation and analysis techniques were applied to mitigate potential bias, these limitations should be considered when interpreting results.

Finally, the findings highlight the potential of ML models to improve the analysis of patient data on CKD and other chronic diseases. However, it is essential to continue researching and optimizing these models

to ensure their accuracy, interpretability, and applicability in diverse clinical settings, to generate optimized prognostic models tailored to each region's reality.

Conclusion

Machine learning models are effective tools for identifying factors associated with chronic kidney disease, enabling recognition of complex patterns linked to its occurrence. Logistic regression stood out for overall performance, while SVM and RF achieved the best results across performance metrics.

Optimization through hyperparameter tuning and cross-validation significantly improved the performance of initially less efficient models, such as DTs. Key variables such as male sex, chronic pesticide exposure, high-temperature activities, and hypertension were identified as determinants, along with the protective effect of fruit and vegetable consumption, especially among women.

The developed models showed statistically significant and reliable performance, even with a small sample, thanks to techniques such as oversampling and cross-validation. This research establishes a reproducible methodology for applying ML models to CKD analysis, validating its relevance in public health and its potential to improve prevention and clinical management using national data.

Funding

No funding sources.

References

i. Ahmad Z, Rahim S, Zubair M, Abdul-Ghafar J. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. Diagn. Pathol. 2021;16(1):24. DOI: 10.1186/s13000-021-01085-4

ii. Vargas M, Biggs D. Inteligencia artificial en medicina: Métodos de modelamiento. Rev. Chil. Anest. 2022;51(5). DOI: 10.25237/revchilanestv5129061230

iii. Ruiz RB, Velásquez JD. Inteligencia artificial al servicio de la salud del futuro. Rev. Médica Clínica Las Condes. 2023;34(1):84-91. DOI: 10.1016/j.rmclc.2022.12.001

iv. Pinton P. Impact of artificial intelligence on prognosis, shared decision-making, and precision medicine for patients with inflammatory bowel disease: a perspective and expert opinion. Ann. Med. 2023;55(2):2300670. DOI: 10.1080/07853890.2023.2300670

v. Ahmed M, Stone ML, Stidham RW. Artificial Intelligence and IBD: Where are We Now and Where Will We Be in the Future? Curr. Gastroenterol. Rep. 2024;26(5):137-144. DOI: 10.1007/s11894-024-00918-8

vi. Briceño J, Calleja R, Hervás C. Artificial intelligence and liver transplantation: Looking for the best donor-recipient pairing. Hepatobiliary Pancreat. Dis. Int. 2022;21(4):347-353. DOI: 10.1016/j.hbpd.2022.03.001

vii. Triantafyllidis A, Polychronidou E, Alexiadis A, Rocha CL, Oliveira DN, da Silva AS, *et al.* Computerized decision support and *machine learning* applications for the prevention and treatment of childhood obesity: A systematic review of the literature. Artif. Intell. Med. 2020;104:101844. DOI: 10.1016/j.artmed.2020.101844

viii. Gottlieb ER, Samuel M, Bonventre JV, Celi L, Mattie H. *Machine Learning* for Acute Kidney Injury Prediction in the Intensive Care Unit. Adv. Chronic Kidney Dis. 2022;29(5):431-438. DOI: 10.1053/j.ackd.2022.06.005

ix. Schena FP, Anelli VW, Abbrescia DI, Di Noia T. Prediction of chronic kidney disease and its progression by artificial intelligence algorithms. J. Nephrol. 2022;35(8):1953-1971. DOI: 10.1007/s40620-022-01302-3

x. Toth-Manikowski SM, Yang W, Appel L, Chen J, Deo R, Frydrych A, *et al.* Sex Differences in Cardiovascular Outcomes in CKD: Findings From the CRIC Study. Am. J. Kidney Dis. Off. J. Natl. Kidney Found. 2021;78(2):200-209.e1. DOI: 10.1053/j.ajkd.2021.01.020

xi. Shatte ABR, Hutchinson DM, Teague SJ. *Machine learning* in mental health: a scoping review of methods and applications. Psychol. Med. 2019;49(9):1426-1448. DOI: 10.1017/S0033291719000151

xii. Tejada Peña DA, Velásquez Arteaga CA. Factores de riesgo asociados a la enfermedad renal crónica en Chalatenango, El Salvador. 2025;8(2):160-7. DOI: 10.5377/alerta.v8i2.18760

xiii. Pedrero V, Reynaldos-Grandón K, Ureta-Achurra J, Cortez-Pinto E. Generalidades del *Machine Learning* y su aplicación en la gestión sanitaria en Servicios de Urgencia. Rev. Médica Chile. 2021;149(2):248-254. DOI: 10.4067/s0034-98872021000200248

xiv. Abdel-Fattah MA, Othman NA, Goher N. Predicting Chronic Kidney Disease Using Hybrid *Machine Learning* Based on Apache Spark. Comput. Intell. Neurosci. 2022;9898831. DOI: 10.1155/2022/9898831

xv. Takkavatakarn K, Oh W, Cheng E, Nadkarni GN, Chan L. *Machine learning* models to predict end-stage kidney disease in chronic kidney disease stage 4. BMC Nephrol. 2023;24(1):376. DOI: 10.1186/s12882-023-03424-7

xvi. Bai Q, Su C, Tang W, Li Y. *Machine learning* to predict end stage kidney disease in chronic kidney disease. Sci. Rep. 2022;12(1):8377. DOI: 10.1038/s41598-022-12316-z

xvii. Song Z, Yang Z, Hou M, Shi X. *Machine learning* in predicting cardiac surgery-associated acute kidney injury: A systemic review and meta-analysis. Front. Cardiovasc. Med. 2022;9:951881. DOI: 10.3389/fcvm.2022.951881

xviii. Sattari M, Mohammadi M. Using Data Mining Techniques to Predict Chronic Kidney Disease: A Review Study. Int. J. Prev. Med. 2023;14:110. DOI: 10.4103/ijpvm.ijpvm_482_21

xix. Ghavidel A, Pazos P. *Machine learning* (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review. J. Cancer Surviv. Res. Pract. 2023. DOI: 10.1007/s11764-023-01465-3

xx. Cresta Morgado P, Carusso M, Alonso Alemany L, Acion L. Practical foundations of *machine learning* for addiction research. Part II. Workflow and use cases. Am. J. Drug Alcohol Abuse. 2022;48(3):272-283. DOI: 10.1080/00952990.2021.1966435

xxi. Kernbach JM, Staartjes VE. Foundations of *Machine Learning*-Based Clinical Prediction Modeling: Part II-Generalization and Overfitting. Acta Neurochir. Suppl. 2022;134:15-21. DOI: 10.1007/978-3-030-85292-4_3

xxii. Awaysheh A, Wilcke J, Elvinger F, Rees L, Fan W, Zimmerman KL. Review of Medical Decision Support and *Machine-Learning* Methods. Vet. Pathol. 2019;56(4):512-525. DOI: 10.1177/0300985819829524

xxiii. Pineda JM. Modelos predictivos en salud basados en aprendizaje de maquina (*machine learning*). Rev. Médica Clínica Las Condes. 2022;33(6):583-590. DOI: 10.1016/j.rmclc.2022.11.002

xxiv. Gulamali FF, Sawant AS, Nadkarni GN. *Machine Learning* for Risk Stratification in Kidney Disease. Curr. Opin. Nephrol. Hypertens. 2022;31(6):548-552. DOI: 10.1097/MNH.0000000000000832

xxv. Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, *et al.* Unsupervised *machine learning* for the discovery of latent disease clusters and patient subgroups using electronic health records. J. Biomed. Inform. 2020;102:103364. DOI: 10.1016/j.jbi.2019.103364

xxvi. Molina P, Ojeda R, Blanco A, Alcalde G, Prieto-Velasco M, Aresté N, *et al.* Etiopathogenesis of chronic kidney disease-associated pruritus: putting the pieces of the puzzle together. Nefrologia. 2023;43(1):48-62. DOI: 10.1016/j.nefroe.2023.03.015

xxvii. Ajayi SO, Ekrikpo UE, Ekanem AM, Raji YR, Ogah OS, Ojji DB, *et al.* Prevalence of Chronic Kidney Disease as a Marker of Hypertension Target Organ Damage in Africa: A Systematic Review and Meta-Analysis. Int. J. Hypertens. 2021;2021:7243523. DOI: 10.1155/2021/7243523

xxviii. Zheng J-X, Li X, Zhu J, Guan S-Y, Zhang S-X, Wang W-M. Interpretable *machine learning* for predicting chronic kidney disease progression risk. Digit. Health. 2024;10:20552076231224225. DOI: 10.1177/20552076231224225

xxix. Valdivia Mazeyra MF, Muñoz Ramos P, Serrano R, Alonso Riaño M, Gil Giraldo Y, Quiroga B. Nefropatía endémica mesoamericana: una enfermedad renal crónica de origen no tan desconocido. Nefrología. 2021;41(6):612-619. DOI: 10.1016/j.nefro.2021.03.005

xxx. Wang F-L, Wang W-Z, Zhang F-F, Peng S-Y, Wang H-Y, Chen R, *et al.* Heat exposure and hospitalizations for chronic kidney disease in China: a nationwide time series study in 261 major Chinese cities. Mil. Med. Res. 2023;10(1):41. DOI: 10.1186/s40779-023-00478-4

xxxi. Bach KE, Kelly JT, Palmer SC, Khalesi S, Strippoli GFM, Campbell KL. Healthy Dietary Patterns and Incidence of CKD: A Meta-Analysis of Cohort Studies. Clin. J. Am. Soc. Nephrol. CJASN. 2019;14(10):1441-1449. DOI: 10.2215/CJN.00530119

xxxii. Seidu S, Abdool M, Almaqhawi A, Wilkinson TJ, Kunutsor SK, Khunti K, *et al.* Physical activity and risk of chronic kidney disease: systematic review and meta-analysis of 12 cohort studies involving 1,281,727 participants. Eur. J. Epidemiol. 2023;38(3):267-280. DOI: 10.1007/s10654-022-00961-7

xxxiii. Blockeel H, Devos L, Frénay B, Nanfack G, Nijssen S. Decision trees: from efficient prediction to responsible AI. Front. Artif. Intell. 2023;6:1124553. DOI: 10.3389/frai.2023.1124553

xxxiv. MacEachern SJ, Forkert ND. *Machine learning* for precision medicine. Genome. 2021;64(4):416-425. DOI: 10.1139/gen-2020-0131

xxxv. Sanmarchi F, Fanconi C, Golinelli D, Gori D, Hernandez-Boussard T, Capodici A. Predict, diagnose, and treat chronic kidney

disease with *machine learning*: a systematic literature review. J. Nephrol. 2023;36(4):1101-1117. DOI: 10.1007/s40620-023-01573-4

xxxvi. Kernbach, JM, Staartjes, VE. Foundations of *Machine Learning*-Based Clinical Prediction Modeling: Part II-Generalization and Overfitting. In: Staartjes, V.E., Regli, L., Serra, C. (eds) *Machine Learning* in Clinical Neuroscience. Acta Neurochirurgica Supplement. 2021;134. DOI: 10.1007/978-3-030-85292-4_3

xxxvii. Okita J, Nakata T, Uchida H, Kudo A, Ueno T, Tanigawa M, *et al.* Development and validation of a *machine learning* model to predict time to renal replacement therapy in patients with chronic kidney disease. BMC Nephrol. 2024;25:101. DOI: 10.1186/s12882-024-03527-9

xxxviii. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in *machine learning* applications. BMC Bioinformatics. 2023;24(1):48. DOI: 10.1186/s12859-023-05156-9

xxxix. Lam K, Mondick JT, Peltz G, Wu M, Kraft WK. Bayesian Population Pharmacokinetic Modeling of Ondansetron for Neonatal Opioid Withdrawal Syndrome. Clin Transl Sci. 2025;18(2):e70147. DOI: 10.1111/cts.70147Van Booven

xl. Van Booven DJ, Chen CB, Malpani S, Mirzabeigi Y, Mohammadi M, Wang Y, *et al.* Synthetic Genitourinary Image Synthesis via Generative Adversarial Networks: Enhancing Artificial Intelligence Diagnostic Precision. J Pers Med. 2024;14(7):703. DOI: 10.3390/jpm14070703

xli. Gerstoft P, Groll H, Mecklenbräuker CF. Parametric Bootstrapping of Array Data with A Generative Adversarial Network. 2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop, Hangzhou, China. 2020;1-5p. DOI: 10.1109/SAM48682.2020.9104371.