



Artículo original

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

DOI: 10.5377/alerta.v8i4.21163

David A. Tejada

Observatorio Nacional de Salud, Instituto Nacional de Salud, San Salvador, El Salvador.

Correspondencia

✉ david.tejada@salud.gob.sv

☎ 0000-0003-2502-1433

ACCESO ABIERTO

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

Citación recomendada:

Tejada DA. Desarrollo y optimización de modelos predictivos de *machine learning* para pacientes con enfermedad renal crónica. Alerta. 2025;8(4):354-365.

DOI: 10.5377/alerta.v8i4.21163

Editor:

David Rivera.

Recibido:

7 de junio de 2024.

Aceptado:

15 de agosto de 2025.

Publicado:

31 de octubre de 2025.

Contribución de autoría:

DAT: concepción del estudio, diseño del manuscrito, búsqueda bibliográfica, recolección de datos, manejo de datos o software, análisis de los datos, redacción, revisión y edición.

Conflictos de intereses:

El autor declara no tener conflicto de intereses.

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.

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.

Introducción

A lo largo de los años, la inteligencia artificial (IA) ha evolucionado significativamente, transformándose en una herramienta esencial para mejorar el análisis de la información en diversas disciplinasⁱ. Desde sus inicios en la década de los 50, la IA ha avanzado de

sistemas básicos de reglas a complejos algoritmos de aprendizaje automático o *machine learning* (ML) y aprendizaje profundo o *deep learning* (DL)ⁱⁱ. Esto permite analizar grandes volúmenes de datos con una mayor precisión y rapidez, facilita la realización de investigaciones y optimiza procesos en campos de la medicinaⁱⁱⁱ.

En las ciencias de la salud, el uso de modelos de ML y DL, han transformado la forma en la que se realiza el diagnóstico, tratamiento y pronóstico de enfermedades^{iv}. Estos modelos permiten una exploración más exhaustiva de datos médicos, mejoran la detección temprana de patologías, personalizan tratamientos y aumentan la eficiencia de las intervenciones médicas^v.

Los algoritmos de ML, como la regresión y los árboles de decisión, proporcionan herramientas robustas para la clasificación y predicción de datos médicos^{vi}. Por otro lado, los modelos DL, como las redes neuronales convolucionales y recurrentes, han demostrado ser excelentes en el procesamiento de imágenes y secuencias, como texto, audio o series temporales, avances significativos en la interpretación y el análisis de registros clínicos^{vii}.

En el ámbito de las enfermedades crónicas, la aplicación de estos modelos se vuelve cada vez más indispensable para mejorar la investigación y el manejo médico. Además, permiten a los investigadores identificar con mayor precisión patrones complejos y predecir la evolución de enfermedades prioritarias como la enfermedad renal crónica (ERC), y proporcionan una base más sólida para futuras investigaciones y avances en el campo clínico^{viii-x}.

En ese contexto, el uso de los modelos de ML en el análisis de datos secundarios, como bases de datos, registros electrónicos y estudios previos, permite a los algoritmos de IA extraer información valiosa y patrones que no son posibles de analizar con los modelos convencionales. Esto ha facilitado no solo la identificación de factores de riesgo, sino también el seguimiento de la evolución de la enfermedad y la personalización del tratamiento^{vxi}.

Por lo tanto, en el presente estudio se aplicaron modelos predictivos de ML para procesar datos de pacientes con factores asociados de ERC. Se analizaron las ventajas y la precisión de cada modelo, así como distintas técnicas de optimización, lo que permitió establecer una metodología complementaria para el análisis e investigación de esta enfermedad. Los objetivos de esta investigación fueron desarrollar, optimizar y comparar el desempeño de los modelos de ML en pacientes con ERC.

Metodología

Se desarrollaron modelos de ML para predecir la enfermedad renal crónica, a partir de los registros del estudio, «Factores de riesgo asociados a la enfermedad renal crónica en Chalatenango, El Salvador»^{xii}. La base estaba

compuesta por 174 registros, que incluían 58 casos de personas con ERC y 116 personas sin la enfermedad. La base comprendía 26 variables independientes y la variable dependiente, «enfermedad renal crónica». Para la realización del estudio se incluyeron modelos de regresión logística (RL), máquinas de soporte vectorial (MSV), bosque aleatorio (BA) y árbol de decisiones (AD).

En el desarrollo de los modelos de ML se utilizó una matriz de correlación con un umbral de $\pm 0,7$ para identificar y eliminar predictores altamente correlacionados, con el fin de evitar la multicolinealidad en los modelos lineales.

Las variables incluidas fueron área, edad, sexo, educación formal, nivel educativo, estado civil, actividad laboral, condición de agricultor, consumo nocivo de tabaco y alcohol, ingesta perjudicial de sal, agua, frutas y verduras, sedentarismo, exposición a altas temperaturas, diabetes *mellitus*, hipertensión arterial, obesidad, infecciones urinarias recurrentes, exposición crónica a plaguicidas y consumo perjudicial de antiinflamatorios no esteroideos (AINES).

Las variables predictoras se normalizaron para garantizar una escala homogénea, para facilitar la comparación entre las características y mejorar la convergencia de los algoritmos de aprendizaje automático. Asimismo, en los modelos se incluyeron las mismas variables con el fin de garantizar condiciones similares y comparables entre los algoritmos.

Los datos se dividieron en 80 % para entrenamiento y 20 % para prueba. Luego se evaluó el balance de las clases mediante el análisis de la distribución de la variable de resultado, para ello se compararon las proporciones de casos de ERC y de casos sin ERC con gráficos de distribución y con la prueba de Chi Cuadrado para identificar diferencias significativas entre los grupos.

Para corregir el desbalance de clases, se aplicó al conjunto de entrenamiento la técnica de sobremuestreo, utilizando el paquete «ROSE» (Random OverSampling Examples) en RStudio, mediante la función «ovun.sample» con el método «over», con el fin de aumentar la cantidad de muestras en la clase minoritaria y equilibrar el conjunto de datos de entrenamiento.

Para entrenar los modelos de ML, se ajustaron los parámetros específicos de cada algoritmo. La regresión logística se configuró con método estándar familia «binomial», de esa manera se abordó el problema de clasificación binaria para determinar la presencia de ERC. Las máquinas de soporte vectorial con «kernel radial» mediante el método «svmRadial». Asimismo, se entrenó un modelo de AD mediante el método «rpart».

Para el BA, se utilizó el método «rf», que integra múltiples árboles de decisión para mejorar la precisión predictiva. En todos los casos, se aplicó validación cruzada con diez pliegues «folds» para garantizar una evaluación precisa y estandarizada, con el fin de prevenir el sobreajuste de los modelos.

El rendimiento de los modelos se evaluó mediante la exactitud de los modelos, que mide la proporción de predicciones correctas realizadas por cada modelo, así como los intervalos de confianza al 95 % para la exactitud. También se analizó la tasa de no información (TNI), definida como la proporción de observaciones pertenecientes a la clase mayoritaria, la cual representa la precisión que se obtendría al clasificar todos los casos como pertenecientes a dicha clase. Esta métrica se utilizó como referencia para evaluar la ganancia real del modelo. La exactitud obtenida por el modelo se comparó estadísticamente con la TNI mediante una prueba de hipótesis, considerando un valor de $p < 0,05$ como criterio para determinar si la exactitud era significativamente superior a la de la TNI.

El coeficiente *Kappa* se utilizó para medir la concordancia entre las predicciones del modelo y los valores reales, considerando la posibilidad que una parte de la coincidencia ocurra por azar, lo que permite evaluar con mayor precisión el desempeño del modelo. Se utilizó la prueba de McNemar para evaluar los patrones de errores en la clasificación de cada modelo comparado con la TNI. Esta prueba contrasta los casos clasificados correctamente por el modelo e incorrectamente por la TNI, y viceversa. Un valor de $p < 0,05$ se interpretó como evidencia de diferencias sistemáticas en el desempeño del modelo respecto a la clasificación por el modelo de clase mayoritaria, indicando sesgos sistemáticos o peor desempeño respecto al modelo de referencia.

Además, se estimaron la sensibilidad y la especificidad, junto con los valores predictivos positivos y negativos. Finalmente, se determinó el rendimiento general de los modelos mediante métricas balanceadas que evaluaron diferentes aspectos del desempeño. La exactitud balanceada promedia la sensibilidad y la especificidad para considerar equitativamente ambas clases en situaciones de desequilibrio. El puntaje F1 combina precisión y sensibilidad mediante la media armónica para reflejar un equilibrio entre falsos positivos y falsos negativos. El área bajo la curva (AUC) mide la capacidad global del modelo para discriminar entre clases a través de todos los posibles umbrales de decisión, donde un valor más cercano a la unidad indica un mejor desempeño discriminativo.

Se seleccionó el modelo con el menor desempeño inicial con el objetivo de evidenciar las distintas posibilidades metodológicas para la optimización de los modelos y mejorar la calidad del análisis. Se eligió un AD desarrollado con el paquete «rpart» y que fue configurado con los hiperparámetros que el paquete de R Studio trae por defecto y presenta un entrenamiento base, el cual presentó el rendimiento más bajo entre los modelos de ML evaluados. A este modelo se le aplicaron distintas técnicas y metodologías de optimización para mejorar su desempeño. Se desarrollaron tres modelos adicionales de AD. El primer modelo utilizó una técnica convencional (modelo con ajuste convencional) con un enfoque en «Train-Test Split», dividiendo los datos en un 70 % para entrenamiento y un 30 % para prueba. Los hiperparámetros se dejaron en sus valores predeterminados, lo que significa que no se hicieron ajustes específicos en la estructura del árbol.

Para el segundo AD, se empleó un modelo con optimización de hiperparámetros, utilizando la técnica de «Grid Search» para ajustar los hiperparámetros enfocándose en el parámetro de complejidad (CP), que automáticamente evaluó una gama de valores de CP desde 0,001 hasta 0,1 en incrementos de 0,005, seleccionando el valor que encontró el mejor equilibrio entre las métricas de rendimiento. Además, se aplicó validación cruzada (Cross-Validation) con cinco pliegues (Folds) para asegurar la robustez del modelo en diferentes configuraciones.

En este modelo, el ajuste del parámetro CP y la validación cruzada son técnicas para prevenir el sobreajuste, ya que el CP controla la poda del árbol, limitando su complejidad y evitando que el modelo se adapte en exceso a los datos de entrenamiento, mientras que, la validación cruzada permite evaluar el desempeño en múltiples subconjuntos, favoreciendo una mayor capacidad de generalización.

En el tercer modelo, se implementó un árbol con técnicas avanzadas y combinadas (modelo con técnicas avanzadas) que usó «Train-Test Split» con una división del 70 % para entrenamiento y el 30 % para prueba. Luego, se aplicó validación cruzada con diez pliegues, dividiendo los datos en diez partes para validar el modelo. También se habilitó la «Probability Estimation» para generar probabilidades de clase y se usaron métricas de clasificación binaria (Binary Classification Metrics) para evaluar el rendimiento del modelo.

Finalmente, se seleccionó el modelo optimizado mediante ajuste de hiperparámetros mostró el mejor balance entre desempeño y menor riesgo de sobreajuste, y con este se realizaron las nuevas predicciones sobre la ERC.

En esta investigación no hubo datos faltantes, por lo que no se aplicaron técnicas de imputación. El desempeño se evaluó utilizando las métricas previamente descritas. El procesamiento y análisis de la información se realizó mediante R Studio versión 4.3.2.

La investigación se desarrolló cumpliendo las Buenas Prácticas Clínicas. La base de datos fue codificada para mantener la confidencialidad de los participantes y el protocolo del estudio fue aprobado por el Comité de Ética del Instituto Nacional de Salud bajo el registro CEINS/2024/006.

Resultados

Desempeño de los modelos de machine learning

Se realizó un análisis comparativo del desempeño de cuatro modelos de ML (Tabla 1). En relación a su rendimiento general, la RL presenta la mayor exactitud con un valor de 0,853 (IC95 % = 0,689 - 0,950), logrando clasificar correctamente al 85,29 % de los ca-

sos de ERC. Los BA y las MSV también muestran un buen desempeño, con exactitudes de 0,823 (IC95 % = 0,655-0,932) y 0,794 (IC95 % = 0,621-0,913), respectivamente.

Por otro lado, el AD presentó la exactitud más baja de los cuatro modelos, con un valor de 0,559 (IC95 % = 0,379 - 0,728), se logró clasificar correctamente al 55,88 % de los casos, el cual es inferior a la TNI (67,65 %), (IC95 % = 0,379 - 0,728).

Al analizar la significancia estadística de los modelos y contrastar la hipótesis de igualdad con la TNI, la RL y el BA resultaron ser superiores y estadísticamente significativas ($p < 0,05$).

En cuanto a las métricas de sensibilidad y especificidad, la RL y el BA presentaron la mayor sensibilidad, 0,956 y 0,91, respectivamente, esto indica que tienen la mayor capacidad para identificar correctamente a los pacientes que presentan ERC. Por el contrario, el AD tiene la menor sensibilidad (0,391), pero la mayor especificidad (0,909), por lo tanto, es el modelo más efectivo para identificar a los pacientes que no tienen ERC.

Tabla 1. Cuadro comparativo del desempeño de los distintos modelos de *machine learning*

Métrica	Regresión logística	Máquinas de soporte vectorial	Bosque aleatorio	Árbol de decisiones
Rendimiento general				
Exactitud	0,853	0,794	0,823	0,559
Intervalo de confianza al 95 %*	0,689 - 0,950	0,621 - 0,913	0,6547 - 0,932	0,379 - 0,728
TNI	0,676	0,676	0,676	0,676
P-valor exactitud > TNexactitud > TNI**	0,017	0,097	0,044	0,949
Kappa	0,638	0,518	0,577	0,230
P-valor de la prueba de McNemar	0,371	1	0,683	0,002
Sensibilidad y especificidad				
Sensibilidad	0,956	0,870	0,913	0,391
Especificidad	0,636	0,636	0,636	0,909
Valores predictivos				
Valor predictivo positivo	0,846	0,833	0,840	0,900
Valor predictivo negativo	0,875	0,700	0,778	0,417
Prevalencia y detección				
Prevalencia	0,676	0,676	0,676	0,676
Tasa de detección	0,647	0,588	0,618	0,265
Prevalencia de detección	0,765	0,706	0,735	0,294
Métricas balanceadas				
Exactitud balanceada	0,796	0,753	0,775	0,650
Área bajo la curva (AUC)	0,852	0,921	0,903	0,682
Puntaje F1	0,897	0,851	0,876	0,545

Fuente: Elaboración a partir del procesamiento de los datos en los distintos modelos.

*Intervalo de confianza de la exactitud.

**p valor [exactitud > TNI]: valor de p asociado a la hipótesis que la exactitud del modelo es mayor que la tasa de no información (TNI).

En la Figura 1 se muestra la curva ROC para cada modelo, junto con el valor del AUC correspondiente. Además, se incluye una comparación entre los modelos para evaluar su rendimiento relativo. Cuando se analizan las métricas balanceadas, las MSV y los BA obtuvieron los mejores resultados, con valores de 0,75 y 0,775 para la exactitud balanceada, asimismo 0,921 y 0,903 para el AUC, respectivamente, sugiriendo que estos modelos logran un buen equilibrio entre sensibilidad y especificidad. Aunque el AD tiene la peor exactitud general, presenta una exactitud balanceada y un AUC que superan a la RL, lo que indica que este modelo podría ser el más adecuado en ciertos contextos en los que se requiera un mejor balance entre la identificación de casos positivos y negativos.

Los resultados de la prueba de McNemar muestran que la RL, las MSV y los BA no presentan diferencias significativas en sus patrones de errores de clasificación en comparación con la TNI. Esto indica que estos modelos no tienen sesgos sistemáticos en sus predicciones ($p > 0,05$).

Optimización de los modelos de árboles de decisiones

En la Tabla 2 se muestra un análisis comparativo del desempeño de cuatro modelos de AD con diferentes niveles de ajuste, desde un modelo con entrenamiento base hasta un modelo construido a partir de técnicas avanzadas. En términos de rendimiento general,

se observa que a medida que aumenta el nivel de ajuste del modelo, la exactitud mejora de forma progresiva (Figura 2).

El modelo con técnicas avanzadas de ajuste presentó el mejor desempeño, con una exactitud de uno (IC95 % = 0,930 - 1), logró clasificar correctamente al 100 % de los casos. Los modelos con ajuste convencional y con ajuste de hiperparámetros mostraron una exactitud de 0,725 (IC95 % = 0,583 - 0,841) y 0,804 (IC95 % = 0,669 - 0,902), respectivamente. Estos modelos son estadísticamente significativos al contrastar la hipótesis de igualdad con la TNI, p valor < 0,05.

Con respecto a las métricas de sensibilidad y especificidad, el modelo con técnicas avanzadas presentó valores de uno para ambas métricas. Los modelos con ajuste convencional y con ajuste de hiperparámetros presentaron una sensibilidad de 0,676 y 0,794, y una especificidad de 0,823, respectivamente. En contraste, el modelo base tiene una sensibilidad de 0,391, pero posee la mayor especificidad con 0,909.

Cuando se analizan las métricas balanceadas, el modelo con técnicas avanzadas de ajuste obtuvo un valor de 1 en ambas. Los modelos con ajuste convencional y con optimización de hiperparámetros presentaron una exactitud balanceada de 0,750 y 0,809, y un AUC de 0,702 y 0,777, respectivamente. El modelo con entrenamiento base, en cambio, muestra valores más bajos, con 0,650 de exactitud balanceada y 0,682 de AUC.

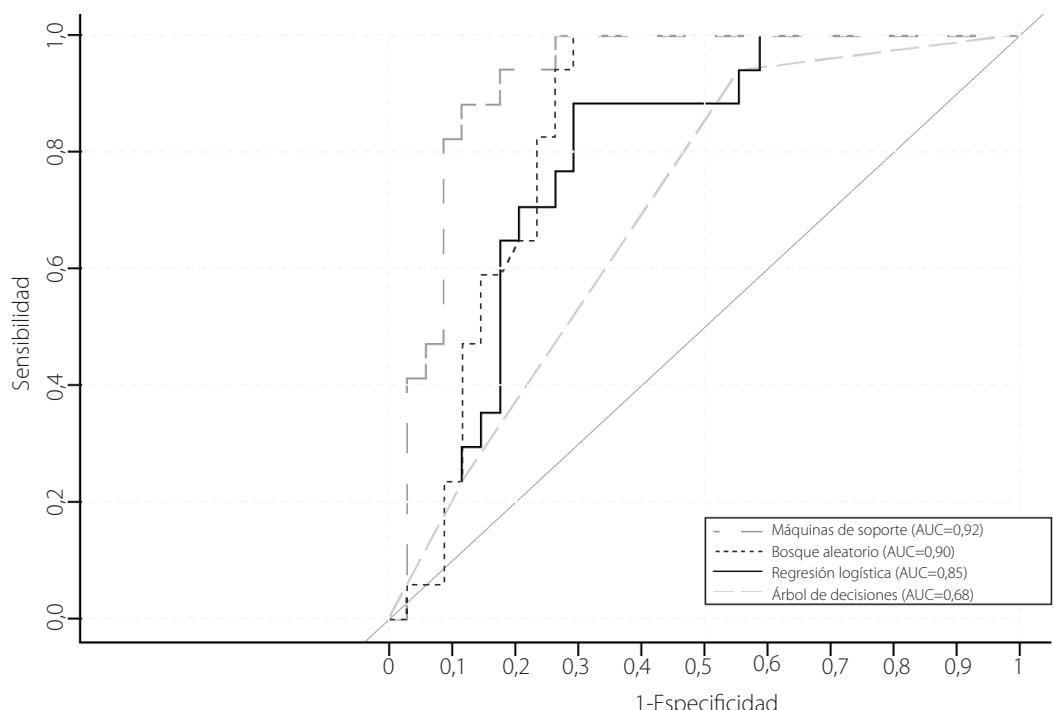


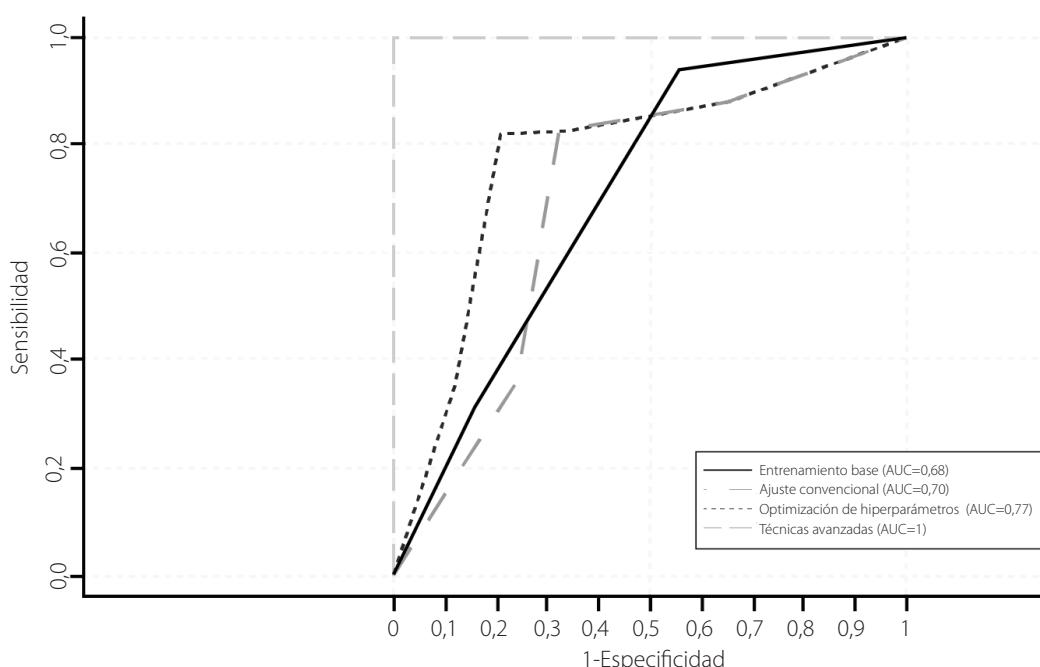
Figura 1. Comparación del rendimiento de los modelos de *machine learning* mediante la curva de ROC

Tabla 2. Cuadro comparativo del desempeño de los modelos de árbol de decisiones según su ajuste

Métrica	Modelo con entrenamiento base	Modelo con ajuste convencional	Modelo con optimización de hiperparámetros	Modelo con técnicas avanzadas
Rendimiento general				
Exactitud	0,559	0,725	0,804	1
Intervalo de confianza al 95 %*	0,379 - 0,728	0,583 - 0,841	0,669 - 0,902	0,930 - 1
TNI	0,667	0,667	0,667	0,667
P-valor exactitud > TNI exactitud > TNI**	0,948	0,023	0,023	< 0,001
Kappa	0,230	0,447	0,583	1
P-valor de la prueba de McNemar	0,002	0,062	0,343	1
Sensibilidad y especificidad				
Sensibilidad	0,391	0,676	0,794	1
Especificidad	0,909	0,823	0,823	1
Valores predictivos				
Valor predictivo positivo	0,9	0,885	0,9	1
Valor predictivo negativo	0,417	0,56	0,667	1
Prevalencia y detección				
Prevalencia	0,676	0,667	0,667	0,667
Tasa de detección	0,265	0,451	0,529	0,667
Prevalencia de detección	0,294	0,510	0,588	0,667
Métricas balanceadas				
Exactitud balanceada	0,650	0,750	0,809	1
Área bajo la curva (AUC)	0,682	0,702	0,778	1
Puntaje F1	0,541	0,767	0,843	1

*Intervalo de confianza de la exactitud.

**p valor [exactitud > TNI]: valor de p asociado a la hipótesis de que la exactitud del modelo es mayor que la tasa de no información (TNI).

**Figura 2.** Comparación del rendimiento de los modelos de árbol de decisiones según su optimización mediante la curva de ROC

La prueba de McNemar muestra que el modelo base presenta diferencias estadísticamente significativas en el patrón de errores de clasificación en comparación con la TNI, $p = 0,002$. Por el contrario, los modelos con ajuste convencional, optimización de hiperparámetros y el de técnicas avanzadas de ajuste, no presentaron diferencias significativas ($p > 0,05$). Lo anterior sugiere que estos modelos no tienen sesgos sistemáticos en sus predicciones ni en los errores de clasificación.

Desempeño del modelo de árbol de decisiones

El AD con optimización de hiperparámetros mostró una exactitud del 80,8 % y una sensibilidad del 79,4 %, así como un área bajo la curva (AUC) de 0,777. Este modelo utilizó principalmente el género y la presencia de hipertensión como factores clave para clasificar a los individuos (Figura 3), comenzó con la identificación de los pacientes sin ERC como la categoría predominante; posteriormente, clasificó a los individuos considerando el género, la hipertensión arterial, el consumo de frutas y verduras y la exposición crónica a plaguicidas.

Para las mujeres, el AD diferenció entre quienes no tenían hipertensión ni otros factores de riesgo y quienes tenían hipertensión, teniendo en cuenta también su consumo de frutas y verduras. Para los hombres, la clasificación se basó en la presencia de hipertensión y la exposición

crónica a plaguicidas. El análisis determinó que la hipertensión arterial y, en el caso de los hombres, la exposición crónica a plaguicidas fueron factores importantes para la predicción de la ERC.

Importancia de las variables predictoras en la decisión del árbol

Los porcentajes de importancia de las variables en un AD indican qué tan relevantes son las variables para la predicción del modelo, estos reflejan su contribución relativa al rendimiento general del árbol. Estos porcentajes se calculan según la frecuencia y calidad de las divisiones realizadas por cada variable en los nodos del árbol; cuanto más frecuente y efectiva sea una variable al mejorar las predicciones, mayor será su importancia.

En el análisis de predicciones de la ERC, los porcentajes de importancia indican qué variables son las más relevantes para identificar la enfermedad, cuáles factores ofrecen la información más determinante para tomar decisiones precisas y segmentar los datos de manera efectiva. En este caso, el sexo masculino (21,62 %) fue la variable más importante, seguido por la exposición crónica a plaguicidas (17,99 %) y las actividades a altas temperaturas (17,62 %), siendo los factores que mejor contribuyen en el rendimiento del modelo al identificar casos de ERC y para entender los elementos que influyen de manera significativa en la enfermedad.

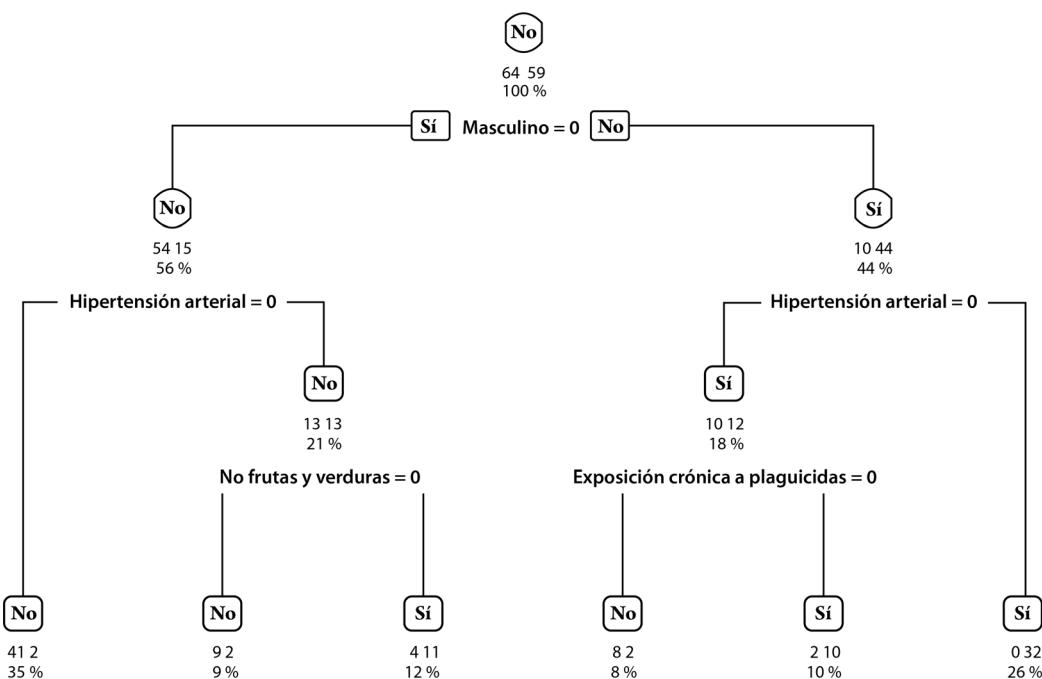


Figura 3. Árbol de decisiones del modelo con optimización de los hiperparámetros

Discusión

El presente estudio utilizó distintos modelos de aprendizaje supervisado para analizar datos de pacientes con factores asociados a la ERC en el departamento de Chalatenango, El Salvador. Los resultados obtenidos muestran el potencial de estos modelos para mejorar la precisión del análisis de datos y optimizar la gestión en sistemas de salud, así como se ha reportado en estudios previos^{xxi}.

En términos generales, el modelo de regresión logística mostró el mejor desempeño, con una exactitud y un puntaje F1 superiores a los otros modelos. Estos resultados concuerdan con investigaciones anteriores que destacan la utilidad de la regresión logística en la predicción de enfermedades crónicas, como la ERC^{xii,xiii}. Esto se debe a su capacidad para modelar la relación entre una variable de respuesta binaria y una o más variables predictoras, demostrando su utilidad y desempeño en datos biométricos con modelos con un similar ajuste^x.

Sin embargo, las MSV y los BA, mostraron el mejor rendimiento global, con los AUC más altos, que superaron al modelo de regresión logística. El enfoque de las MSV busca el mejor hiperplano posible para separar las clases, de esa forma aseguraron una clasificación precisa en problemas de clasificación binaria^{xiv,xv}. En cambio, los BA combinan múltiples AD para mejorar la precisión predictiva^{xvi,xvii} y reducir el sobreajuste que pueda limitar la extrapolación y predicciones posteriores^{xvi-xviii}.

La selección del modelo más adecuado es fundamental para un apropiado análisis de información; cada modelo ofrece ventajas específicas de acuerdo a la métrica que se haya priorizado. La precisión es fundamental cuando se pretende lograr una alta tasa de predicciones correctas, lo cual es esencial para la toma de decisiones clínicas confiables y seguras. Por otra parte, la sensibilidad será más necesaria si se desea detectar a todos los casos positivos, especialmente en el diagnóstico de enfermedades priorizadas, en donde es crucial no pasar por alto ningún caso que requiera atención médica^{xix}.

Asimismo, la especificidad es importante para minimizar los falsos positivos y evitar tratamientos innecesarios que puedan causar daños o malestar a los pacientes^{xx}. En general, es necesario encontrar un equilibrio óptimo entre sensibilidad y especificidad, lo cual puede evaluarse mediante métricas integradoras como la exactitud balanceada, puntaje F1 y el AUC, que permitía analizar el desempeño del modelo de manera

más integral, tomando en consideración tanto la capacidad de detección como la capacidad de discriminación entre casos positivos y negativos^{xx}.

El modelo de AD inicial presentó un desempeño relativamente bajo, tanto en su exactitud como en su AUC. Al aplicar técnicas de ajuste de hiperparámetros y validación cruzada, se logró mejorar significativamente el rendimiento. Estos hallazgos respaldan la importancia de optimizar los modelos de aprendizaje automático mediante el ajuste de hiperparámetros y la validación cruzada, como se ha destacado en estudios previos^{xii,xxi,xxii}. El ajuste de hiperparámetros implica probar diferentes configuraciones para encontrar la combinación óptima que maximice el rendimiento del modelo a partir de los datos de entrenamiento y validación^{xii,xxi,xxii}.

Uno de los objetivos de esta investigación fue mostrar las técnicas de optimización aplicadas a distintos modelos, empleando el lenguaje R. Existen múltiples lenguajes, plataformas y entornos que permiten realizar este tipo de análisis. La elección de la estrategia de análisis dependerá de factores como la experiencia del investigador, la complejidad del problema y la disponibilidad de recursos, sin que ello limite la aplicabilidad de los métodos presentados. La flexibilidad de estos enfoques facilita su adaptación a diversos contextos, lenguajes de programación y datos, para favorecer la replicabilidad en futuros estudios.

Asimismo, los resultados obtenidos mediante técnicas de ML fueron consistentes con la literatura y las teorías existentes sobre la ERC, y destacó su naturaleza multifactorial^{xxiii}. Los modelos predijeron con precisión variables como el sexo^{xxiv}, la hipertensión^{xxv} y la exposición a agroquímicos, que se identificaron como factores de riesgo independientes en el modelo de AD para la ERC^{xxvi}, así como ha sido documentado en estudios similares^{xxvii,xxviii}.

Sin embargo, aún se requiere de más investigaciones para cuantificar de manera precisa el riesgo asociado a estas exposiciones, así como la inclusión de una mayor cantidad de variables para la construcción de modelos de ML mucho más complejos, similar a otros estudios, en donde se ha logrado determinar la progresión de la enfermedad a partir de datos históricos de las personas^{xxix}.

En el estudio se destacan los efectos protectores del consumo de frutas y verduras^{xxx,xxxi} y las diferencias en la prevalencia de la ERC según el sexo^{xxii}. La principal importancia de esta investigación

radica en la aplicación de modelos de ML para el análisis de datos nacionales, lo que permitió una evaluación más precisa y exhaustiva de los factores asociados a la ERC, adaptándose a diversos criterios de sensibilidad, especificidad y precisión, para facilitar una comprensión más profunda de los factores que contribuyen a la enfermedad^{xxvii,xxviii}.

Un aspecto importante a considerar en este estudio es la interpretabilidad de los modelos. Aunque los AD son relativamente fáciles de interpretar debido a su estructura visual y a las reglas de decisión que utilizan para predecir resultados, los modelos de RL, BA y MSV son más complejos y pueden resultar más difíciles de interpretar para los profesionales que no están familiarizados con estos modelos. Sin embargo, estos modelos pueden lograr un mejor rendimiento y precisión en sus predicciones^{xvi}. En este sentido, el estudio de Hendrik Blockeel *et al.*, sugieren que los AD pueden ser preferibles en entornos clínicos debido a su visualización y facilidad de interpretación^{xxxiii}. No obstante, la afirmación anterior va a estar determinada por lo que se espera de cada uno de los modelos. En este sentido, la interpretabilidad no debe ser el único criterio para elegir uno sobre otro. En algunos casos, será necesario tener un mayor rendimiento predictivo, así como en este contexto, en donde las MSV y BA pueden ser ideales cuando se requiere una mayor precisión, especialmente en la identificación de factores asociados y la toma de decisiones^{xxxiv}.

La incorporación de variables adicionales, como datos de pruebas de laboratorio o imágenes médicas, podría enriquecer el análisis y aumentar la precisión de las predicciones. Los modelos de ML son capaces de aprovechar múltiples fuentes de datos, incluyendo datos estructurados y no estructurados^{xxxv,xxxvi}. La incorporación de estas fuentes podría proporcionar información adicional para el diagnóstico y la predicción de la ERC, y mejorar aún más el rendimiento e importancia de los modelos^{xxxvii}.

En este estudio, se destacaron hallazgos importantes en la investigación de la ERC utilizando ML. Se desarrolló un modelo que se puede comparar con otro que obtuvo un AUC de 0,80 utilizando modelos de BA y redes neuronales para predecir la ERC^{xii}. Esta investigación permitió obtener resultados similares e incluso superiores, demostró una alta capacidad predictiva y rendimiento para el estudio de la ERC^{xxxviii}.

Además, para investigaciones futuras, es importante explorar métodos más avanzados, como las redes neuronales,

que pueden mejorar significativamente el rendimiento del modelo. Sin embargo, estas técnicas requieren una mayor capacidad de procesamiento y presentan desafíos adicionales en la interpretación de los resultados^{xix}.

Aunque la muestra de 178 observaciones puede considerarse pequeña en ciertos contextos, la literatura respalda el uso de tamaños similares o incluso menores cuando se aplican técnicas específicas como validación cruzada, sobremuestreo y regularización, que permiten maximizar el aprovechamiento de los datos, mejorar la generalización de los modelos y obtener resultados confiables^{xxxvi,xxxvii}. La idoneidad del tamaño de muestra depende de la complejidad del análisis y del número de variables; en modelos complejos un tamaño reducido puede ser insuficiente, mientras que, en análisis más simples con la utilización de técnicas adecuadas, puede ser suficiente para obtener resultados sólidos^{xxxviii}.

Diversos estudios han validado y utilizado modelos con conjuntos de datos reducidos mediante simulación, empleando técnicas como Monte Carlo^{xxxix} y técnicas como bootstrap y las redes adversariales generativas, que permiten ampliar con precisión los datos sintéticos para mejorar el entrenamiento y la validación de los modelos^{xl,xli}. Estas metodologías se aplican en campos como la genética, la economía, la visión por computadora y la salud, siendo especialmente útiles cuando la obtención de datos es limitada o costosa y cuya adopción es fundamental para fortalecer la calidad y el alcance de la investigación en salud.

Los estudios muestran que, cuando los datos tienen buen poder discriminatorio, el tamaño del efecto y la precisión del modelo aumentan con el tamaño de la muestra, mientras que la varianza disminuye^{xxxviii}. No obstante, en conjuntos con baja discriminación, ni el tamaño del efecto ni la precisión mejoran significativamente al aumentar la muestra. Se considera adecuado un tamaño muestral que logre un tamaño de efecto $\geq 0,5$ y una precisión en aprendizaje automático $\geq 80\%$, ya que aumentos posteriores no generan mejoras significativas, lo que permite optimizar la relación costo-beneficio^{xxxviii}.

La literatura también indica que tamaños pequeños pueden sobreestimar la precisión por sobreajuste, mientras que, muestras muy grandes implican mayores costos y tiempos sin mejoras proporcionales en la confiabilidad. Por ello, definir un tamaño adecuado es clave para obtener resultados confiables y eficientes. En este contexto, el tamaño de muestra empleado en esta

investigación está justificado por cumplir criterios de efecto y precisión que equilibran fiabilidad y recursos disponibles^{xxxviii}.

La principal limitación de este estudio es el uso de datos secundarios, lo cual conlleva ciertos riesgos; no obstante, dado que la investigación original tuvo como propósito identificar factores asociados para ERC, este análisis permitió profundizar en aspectos que métodos tradicionales no abordan. Aunque se aplicaron técnicas adecuadas de validación y análisis para mitigar posibles sesgos, es importante considerar estas limitaciones al interpretar los resultados.

Finalmente, los hallazgos de este estudio resaltan el potencial de los modelos de ML para mejorar el análisis de datos de pacientes con ERC y otros trastornos crónicos. No obstante, es fundamental continuar la investigación y optimización estos modelos para garantizar su precisión, interpretabilidad y aplicabilidad en diversos escenarios y entornos clínicos, con el propósito de generar modelos pronósticos optimizados para la realidad de cada territorio.

Conclusión

Los modelos de ML son herramientas efectivas para identificar factores de riesgo asociados a la enfermedad renal crónica, que permiten reconocer patrones complejos que contribuyen a su desarrollo. La regresión logística destacó por su desempeño general, mientras que, las máquinas de soporte vectorial y los bosques aleatorios obtuvieron los mejores resultados en las métricas de desempeño.

La optimización mediante ajuste de hiperparámetros y validación cruzada mejoró significativamente el rendimiento de modelos inicialmente menos eficientes, como los AD. Variables clave como sexo masculino, exposición crónica a plaguicidas, actividades a altas temperaturas e hipertensión arterial fueron identificadas como determinantes, además del efecto protector del consumo de frutas y verduras, especialmente en mujeres.

Los modelos desarrollados demostraron ser estadísticamente significativos y confiables, incluso con una muestra reducida, gracias a la aplicación de técnicas como el sobremuestreo y la validación cruzada. Esta investigación establece una metodología reproducible para la aplicación de modelos de ML en el análisis de la enfermedad renal crónica, lo que valida la relevancia de este en el ámbito de la salud pública y su potencial para mejorar la prevención y el manejo clínico mediante el uso de datos nacionales.

Financiamiento

No hubo fuentes de financiamiento.

Referencias bibliográficas

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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/ijpm.ijpm_482_21](https://doi.org/10.4103/ijpm.ijpm_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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1007/978-3-030-85292-4_3)
 - xxii. Awasyeh 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](https://doi.org/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](https://doi.org/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.00000000000000832](https://doi.org/10.1097/MNH.00000000000000832)
 - 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.2215/CJN.00530119)
- xxxii. Seidu S, Abdool M, Almaghawi 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1109/SAM48682.2020.9104371).