



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

**PREDICCIÓN DE LA RESPUESTA AL TRATAMIENTO EN
PACIENTES CON ARTRITIS REUMATOIDE UTILIZANDO UN
MODELO BASADO EN MACHINE LEARNING**

**PREDICTING RESPONSE TO TREATMENT IN PATIENTS WITH
RHEUMATOID ARTHRITIS USING MACHINE LEARNING**

Claudia Irene Ibáñez Antequera, ciibaneza@libertadores.edu.co, Fundación Universitaria Los Libertadores

RESUMEN

La artritis reumatoide (AR) es una enfermedad inflamatoria crónica, autoinmune e incapacitante. La respuesta al tratamiento o la remisión sostenida son las principales metas en el manejo de estos pacientes; sin embargo, un número significativo de pacientes no responden al tratamiento. Por lo tanto, predecir el logro de la respuesta o remisión es fundamental para el manejo de los pacientes, ya que permite considerar la disminución gradual del uso de terapias, ayudando a evitar potenciales desenlaces negativos a nivel articular y efectos secundarios no deseados; además de, contribuir en la reducción de los elevados costos del tipo de terapias utilizadas en AR. **Objetivo.** Identificar las variables predictoras de la respuesta al tratamiento o remisión en pacientes con artritis reumatoide atendidos entre 2018 y 2022 en un Centro de Atención Integral en Artritis Reumatoide. **Métodos.** Se desarrollaron modelos *machine learning* utilizando una cohorte retrospectiva de 1226 pacientes con diagnóstico confirmado de AR. Se analizaron descriptivamente las

variables relacionadas con características demográficas y clínicas; los pacientes se clasificaron como respondedores y no respondedores. Se entrenaron, validaron y pusieron a prueba diferentes algoritmos de ML para predecir la respuesta al tratamiento y se comparó su desempeño. **Resultados.** Del total de pacientes, 895 fueron respondedores y 331 no respondedores. El modelo que ofreció mejores resultados fue el modelo Random Forest con una exactitud y AUC de 71 y 62%, respectivamente. Los factores que contribuyeron en mayor medida a la predicción de la respuesta al tratamiento fueron el uso de DMARDs, el DAS28 inicial y la puntuación en el HAQ. **Conclusiones.** El modelo RF clasificó adecuadamente a los pacientes que responden al tratamiento e identificó características clínicas relevantes para predecir la respuesta. Todos los modelos evaluados mostraron pobre desempeño para clasificar correctamente a los pacientes que no responden al tratamiento.

Palabras clave: artritis reumatoide, predicción, tratamiento, machine learning, inteligencia artificial

ABSTRACT

Rheumatoid arthritis (RA) is a chronic inflammatory, autoimmune and disabling disease. Response to treatment or sustained remission are the main goals in the management of these patients; however, a significant number of patients do not respond to treatment. Therefore, predicting the achievement of response or remission is essential for patient management, since it allows considering the gradual decrease in the use of therapies, helping to avoid potential negative outcomes at the joint level and unwanted side effects; in addition to contributing to the reduction of the high costs of the type of therapies used in RA. **Objective.** To identify the predictive variables of the response to treatment or remission in patients with rheumatoid arthritis treated between 2018 and 2022 in a Care Center for Rheumatoid Arthritis. **Methods.** Machine learning models were developed using a retrospective cohort of 1,226 patients with a confirmed diagnosis of RA. Variables related to demographic and clinical characteristics were analyzed descriptively; patients were classified as responders and non-responders. Different ML algorithms to predict treatment response were trained, validated, and tested, and their performance was compared. **Results.** Of the total number of patients, 895 were responders and 331 were non-responders. The model that offered the best

results was the Random Forest model with an accuracy and AUC of 71% and 62%, respectively. The factors that contributed the most to the prediction of treatment response were the use of DMARDs, the initial DAS28 and the HAQ score. **Conclusions.** The RF model adequately classified responders and identified relevant clinical features to predict response. All the models evaluated showed poor performance in correctly classifying patients who do not respond to treatment.

Keywords: R heumatoid arthritis, prediction, treatment, machine learning, artificial intelligence

INTRODUCCIÓN

La artritis reumatoide (AR) es una enfermedad inflamatoria crónica, degenerativa e incapacitante que afecta a los tejidos sinoviales de grandes y pequeñas articulaciones. Se presenta en el 1% de la población y tras padecer la enfermedad durante un tiempo alrededor de 12 años, el 80% de los pacientes se encuentra parcialmente limitado y el 16% totalmente (1). Se ha descrito que el riesgo de mortalidad es entre un 60% y un 70% más alto en pacientes con AR que en aquellos sin la enfermedad (2). Aunque la etiología exacta de la AR aún se encuentra bajo investigación, varios estudios han sugerido un papel importante de numerosos factores genéticos, hormonales y ambientales (3,4).

La respuesta al tratamiento y el logro de la remisión sostenida son las principales metas en el manejo de los pacientes con AR. En la actualidad, aproximadamente el 50% de los pacientes con AR alcanzan una remisión sostenida (5). Esta mejora en los resultados se debe a un diagnóstico más temprano de la enfermedad y un tratamiento más preciso basado en objetivos. Instaurar el tratamiento precozmente es deseable, teniendo en cuenta que en los dos primeros años de inicio de la enfermedad pueden ocurrir daños articulares irreversibles (1).

En el manejo de la AR se ha establecido como terapia estándar el uso de medicamentos antirreumáticos modificadores de la enfermedad (DMARDs, por sus siglas en inglés), que pueden inducir remisión y mantienen el control de la enfermedad articular inflamatoria.

Recientemente, los productos biológicos se consideran una línea prometedora para el tratamiento de estos pacientes. Una expansión del tratamiento de la AR con el uso de medicamentos más efectivos, como los medicamentos biológicos antirreumáticos modificadores de la enfermedad (bDMARDs) o los DMARDs sintéticos dirigidos (tsDMARDs), ha evidenciado una mejoría en la respuesta sostenida de pacientes particularmente resistentes al manejo convencional (6).

Sin embargo, el tratamiento con biológicos puede no ser satisfactorio en algunos casos. Diversos ensayos clínicos han demostrado con frecuencia que aproximadamente el 30-40% de los pacientes no responden al tratamiento con productos biológicos debido a características fisiológicas, clínicas y patológicas disímiles entre los individuos (7, 8). Adicionalmente, dado el creciente número de combinaciones de medicamentos disponibles, la demora en encontrar el mejor tratamiento para un paciente puede ser significativa. Tampoco hay recomendaciones claras y estandarizadas sobre la disminución del tratamiento en caso de enfermedad estable, por lo que el sobretratamiento o el subtratamiento en la AR es usual, lo que puede resultar en brotes destructivos de la enfermedad o efectos secundarios y costos innecesarios (9)

En este momento, predecir el logro de la remisión resulta fundamental para el manejo de estos pacientes, ya que permite considerar la disminución gradual del uso de los DMARDs, ayudando a evitar potenciales desenlaces negativos a nivel articular y efectos secundarios no deseados, además de contribuir en la reducción de los costos elevados de este tipo de terapias. Resultados obtenidos a partir de ensayos clínicos aleatorizados controlados y de estudios observacionales, sobre la reducción escalonada de los DMARDs, han sugerido que hasta el 50% de los pacientes que alcanzan una remisión sostenida pueden reducir con éxito la terapia (6,10). Por lo tanto, los métodos para desarrollar predictores que puedan identificar la respuesta al tratamiento han ganado auge.

Diversos estudios han evaluado la asociación entre factores demográficos y clínicos como edad, sexo, duración y actividad de la enfermedad, y los resultados del tratamiento en pacientes con AR; incluso estudios genéticos y proteómicos han utilizado distintos métodos

estadísticos para identificar biomarcadores que predigan la respuesta a productos biológicos en estos pacientes (11-13). No obstante, la regresión logística fue casi siempre el algoritmo de elección para encontrar predictores independientes en modelos multivariados, cabe señalar que las hipótesis de algunos de estos estudios apoyaban con frecuencia la suposición irreal de que la asociación entre los factores pronósticos y la respuesta o remisión en la AR es directa y aislada (14). Algunos autores sostienen que la regresión logística no es el método apropiado para modelar variables no independientes y que podría ser inadecuado para describir explícitamente la intrincada relación entre los factores pronósticos y la remisión o respuesta al tratamiento, considerando los mecanismos multifactoriales y poco claros de una enfermedad como la AR (14,15). Traducir la información desde algoritmos predictores a la práctica en la vida real, puede ser un tema bastante complejo debido a la heterogeneidad intrínseca de las enfermedades reumatológicas.

Los enfoques de aprendizaje automático o *machine learning*, de uso creciente en investigación clínica, pueden complementar, o incluso superar, los métodos estadísticos habitualmente utilizados para predecir desenlaces clínicos, ya que logran incorporar información de forma tal que es posible obtener predicciones más precisas. La aplicación de modelos *machine learning* ofrece la oportunidad de manejar relaciones no lineales complejas entre los atributos de los individuos que son difíciles de modelar con los métodos estadísticos tradicionales; razón por la cual, el *machine learning* se postula como una herramienta competente para la implementación de algoritmos de decisión multiparamétricos complejos (14).

En el campo de la salud, y particularmente en reumatología, lo anterior representa un paso hacia la llamada ‘medicina de precisión’, conduciendo a la mejora del cuidado médico con una atención más personalizada o individualizada, cuyo propósito es tomar decisiones sobre manejo y tratamiento adaptados a cada paciente.

Por lo anterior, en el presente estudio se buscó identificar las variables predictoras de la respuesta al tratamiento o remisión en pacientes con artritis reumatoide atendidos entre 2018

y 2022 en un Centro de Atención Integral en Artritis Reumatoide mediante un modelo de aprendizaje automático supervisado.

METODOLOGÍA

Datos

Este estudio incluyó una cohorte retrospectiva de pacientes con diagnóstico confirmado de AR, identificados a partir de los registros del Centro de Atención Integral en Artritis Reumatoide Biomab. Se tuvieron en cuenta pacientes atendidos entre 2018 y 2022, mayores de 18 años que cumplieron los criterios internacionales validados para AR (2010 *Rheumatoid Arthritis Classification Criteria American College of Rheumatology/European League Against Rheumatism Collaborative Initiative*). Se excluyeron pacientes con inicio de la artritis antes de los 16 años. Los datos de línea de base fueron obtenidos del momento de la primera consulta del paciente al centro, los datos de seguimiento relacionados con la respuesta al tratamiento se tomaron del control más próximo realizado entre los 6 y 12 meses desde la primera consulta.

Definición de la variable respuesta

La respuesta al tratamiento fue evaluada con base en múltiples criterios de respuesta, y fue representada como una variable Booleana, de forma que los pacientes se clasificaron como respondedores y no respondedores. Tabla 1.

Tabla 1. Definición de la variable objetivo

Valor variable objetivo	Significado	Condiciones
1	Respondedor	Una de las siguientes: <ul style="list-style-type: none">● Lograr baja actividad de la enfermedad (DAS28 \leq3.2) o remisión (DAS28 $<$2.6) durante el seguimiento● Cumplir con los criterios de respuesta de la Liga Europea contra el Reumatismo (EULAR)(16) para respuesta buena o moderada:<ul style="list-style-type: none">○ Buena respuesta: DAS28 \leq3.2 y una reducción $>$1.2 respecto a la puntuación basal;○ Respuesta moderada: DAS28 \leq3.2 y una reducción $>$0.6 y \leq1.2 respecto a la puntuación basal, DAS28 $>$3.2 y \leq5.1 y una reducción $>$0.6 respecto a la puntuación basal,

		DAS28 >5.1 y una reducción >1.2 respecto a la puntuación basal
0	No respondedor	No mostrar respuesta según los criterios EULAR o no cumplir criterios para baja actividad de la enfermedad o remisión

Preparación inicial de los datos y variables de entrada

Los datos recolectados aportaron información relacionada con variables sociodemográficas (edad, sexo), comorbilidades (hipertensión arterial, diabetes mellitus, enfermedad cardiovascular, enfermedad renal crónica, osteoporosis, síndrome de Sjögren), resultados de laboratorio (proteína C reactiva, factor reumatoide, Anticuerpos Antipéptido Citrulinado Cíclico anti-CCP, hemoglobina, creatinina, velocidad de sedimentación globular, leucocitos, alanina aminotransferasa, tasa de filtración glomerular, proteinuria), actividad de la enfermedad (*Disease Activity Score* DAS28 inicial), capacidad funcional del paciente (*Health Assessment Questionnaire* HAQ) y tratamiento recibido (DMARDs biológicos, metotrexate, otros sintéticos convencionales, sintéticos dirigidos o pequeñas moléculas, antimaláricos, corticoides, AINES, analgésicos opioides y no opioides).

Inicialmente se llevó a cabo un análisis descriptivo de los datos, con el fin de evaluar la distribución de estos e identificar valores faltantes. Como se reportaron comorbilidades con muy baja frecuencia se decidió agruparlas todas en una sola variable, de igual manera se procedió con los DMARDs. Durante la selección de las variables de entrada, no se tuvieron en cuenta aquellas con una proporción de valores perdidos superior a 40% ni variables categóricas con poca variabilidad (categoría < 10%). Como resultado 18 variables fueron preseleccionadas para realizar la modelación. Las variables de entrada seleccionadas se describen en la tabla 2.

Tabla 2. Variables explicativas consideradas en la modelación

Variable	Definición	Valores
Edad	Edad en años	Años
IMC	Índice de masa corporal	Kg/m ²
PCR	Proteína C reactiva	Reactiva; No reactiva
VSG	Velocidad de sedimentación globular	mm/hr
HB	Hemoglobina	gr/dL
Leucocitos	Recuento de leucocitos en sangre	Valor entre 0 y 20000
Creatinina	Nivel de creatinina sérica	mg/dL
Duración de la enfermedad	Tiempo de duración de la enfermedad desde el diagnóstico en años	Años
DAS-28 inicial	<i>Disease Activity Score</i> en la línea de base	Valor entre 0 y 10
HAQ	<i>Health Assessment Questionnaire</i> en la línea de base	Valor entre 0 y 3
Anti-CCP	Anticuerpos Antipéptido Citrulinado Cíclico	Positivo; Negativo
FR	Factor Reumatoide	Positivo; Negativo

Comorbilidades	Diagnóstico concomitant de alguna de las siguientes enfermedades: hipertensión arterial, diabetes mellitus, enfermedad cardiovascular, enfermedad renal crónica, osteoporosis, síndrome de Sjögren	Si; No
DMARDs	Uso de DMARDs: biológicos (Adalimumab, Certolizumab, Etanercept, Golimumab, Infliximab, Abatacept, Rituximab, Tocilizumab) sintéticos convencionales (Leflunomida, Sulfazalazina, Azatioprina, Ciclosporina, Ciclofosfamida*), sintéticos dirigidos o pequeñas moléculas (Tofacitinib)	Si; No
Metotrexate	Uso de metotrexate	Si; No
Corticoides	Uso de corticoides	Si; No
Analgésicos	Uso de analgésicos opioide (tramadol, codeína) y no opioides (acetaminofén, dipirona)	Si; No
Antimaláricos	Uso de antimaláricos (cloroquina, hidroxicloroquina)	Si; No

*Es un inmunosupresor no DMARDs, por haber muy pocos casos, se analizó en este grupo

Preprocesamiento y modelación

Con el propósito de obtener métricas para ser utilizadas como valores de referencia, se construyó un modelo basado en reglas, utilizando la totalidad de las observaciones. Se planteó como regla un valor de DAS28 inicial $<3,2$ y una duración de la enfermedad <10 años. Posteriormente los datos se dividieron en conjuntos independientes de datos de entrenamiento (80%, $n = 980$) y de prueba (20%, $n = 246$). Los modelos predictivos se generaron mediante *machine learning* utilizando solo el conjunto de datos de entrenamiento. El conjunto de datos de prueba no participó en ningún entrenamiento o validación interna de los modelos predictivos. Solo se utilizó para la validación externa final de cada modelo entrenado. Los datos faltantes en el conjunto de entrenamiento fueron imputados recurriendo a una imputación simple, la media para las variables numéricas y la categoría valor constante (“no disponible”) para las variables categóricas. Para medir el comportamiento de los modelos y el sobreajuste de estos, sobre el conjunto de entrenamiento se llevó a cabo una validación cruzada con 10 segmentos.

Para corregir el desbalance de clases se utilizó la técnica SMOTE (*Synthetic Minority Oversampling Technique*), la cual permite aumentar la clase minoritaria a partir del aumento sintético de los datos. En cuanto a la modelación, se realizó un análisis global de los posibles modelos de clasificación para la predicción de la respuesta al tratamiento, de los cuales se seleccionaron los cinco modelos con los puntajes más altos respecto a las métricas de desempeño: *accuracy* o exactitud, área bajo la curva (AUC por sus siglas en inglés), *recall* o

sensibilidad, precisión y puntaje F1; se examinaron las respectivas matrices de confusión y las características más importantes en cada modelo. Se evaluó el pronóstico en los datos de prueba y posteriormente se realizó un ajuste de hiperparámetros para optimizar la exactitud y el área bajo la curva con el fin de observar si mejoraba el desempeño de los modelos. Se compararon todas las métricas y se escogió el mejor modelo. Finalmente, se estudiaron las características más importantes que contribuyeron a la predicción en el modelo elegido, para lo cual se utilizó el método de explicaciones aditivas SHAP.

Todos los modelos de aprendizaje automático se generaron y evaluaron utilizando Python 3.10, mediante las librerías Pycaret y Scikit-learn.

RESULTADOS

Características de los pacientes

Se incluyeron un total de 1226 pacientes con diagnóstico de AR, con una edad promedio de 59.3 ± 14 años, la mayoría (77.4%) eran de sexo femenino. Del total de pacientes, 331 (27%) fueron clasificados como no respondedores y 895 (73%) como respondedores.

La duración promedio de la enfermedad fue 4.5 ± 7.9 años, con un 81% (991/1226) de los pacientes con una duración de la enfermedad menor a 10 años. El factor reumatoide y el anti-CCP no se encontraron disponibles para todos los pacientes, reportando una positividad del 79.2% (929/1162) y 74.3% (832/1120), respectivamente. El valor promedio del puntaje DAS28 inicial fue de 3.5 ± 1.3 , con un 46% de los pacientes con un DAS28 inicial ≤ 3.2 considerado como baja actividad de la enfermedad.

El antecedente más frecuente fue hipertensión arterial, presente en un 28.8% de los casos, seguido de osteoporosis en un 16.6% y sólo un 6% de la población incluida reportó la presencia de Síndrome de Sjögren. El tratamiento con DMARDs más comúnmente recibido fue metrotexate (74.6%), seguido de otros DMARDs sintéticos convencionales. La mediana de tiempo de seguimiento fue de 7 meses. Tabla 3.

Tabla 3. Características generales de los pacientes

Variable	No respondedores n = 331	Respondedores n = 895	Todos N = 1226
Edad en años, media (DE)	5,4 (13,5)	60 (14,1)	59,3 (14,0)
Sexo femenino, n (%)	256 (77,3)	693 (77,4)	949 (77,4)
Duración de la enfermedad en años, media (DE)	5,3 (8,4)	2,2 (7,7)	4,5 (7,9)
IMC kg/m, media (DE)	26 (4,9)	25,6 (4,4)	25,7 (4,5)
Factor reumatoide positivo, n (%)	265/316 (83,9)	655/846 (77,4)	929/1162 (79,2)
Anti-CCP positivo, n (%)	234/297 (78,8)	598/823 (72,7)	832/1120 (74,3)
Puntaje DAS-28 inicial, media (DE)	4 (1,4)	3,3 (1,2)	3,5 (1,3)
Puntuación HAQ, mediana (RIC)	0,7 (0,1 - 1,1)	0,2 (0 - 1)	0,4 (0 - 1)
PCR reactiva, n (%)	73/134 (54,5)	132/359 (36,8)	205/493 (41,6)
VSG en mm/hr, media (DE)	29 (23,0)	20,3 (19,3)	22,6 (20,7)
Hipertensión arterial, n (%)	77 (23,3)	276 (30,8)	353 (28,8)
Diabetes mellitus, n (%)	26 (7,9)	62 (6,9)	88 (7,2)
Enfermedad cardiovascular, n (%)	15 (4,5)	38 (4,2)	53 (4,3)
Enfermedad renal crónica, n (%)	1 (0,3)	6 (0,7)	7 (0,6)
Osteoporosis, n (%)	48 (14,5)	155 (17,3)	203 (16,6)
Síndrome de Sjögren, n (%)	20 (6,0)	47 (5,2)	67 (6,0)
Analgésicos No Opioides (Acetaminofén, Diproina), n (%)	144 (43,5)	384 (42,9)	528 (43,1)
Analgésicos Opioides (Codeína, Tramadol), n (%)	61 (18,4)	92 (10,3)	153 (12,5)
AINES, n (%)	7 (2,1)	14 (1,6)	21 (1,7)
Corticoides, n (%)	264 (79,8)	574 (64,1)	838 (68,4)
Metotrexate, n (%)	252 (76,1)	663 (74,1)	915 (74,6)
DMARDs biológicos, n (%)	40 (12,1)	89 (9,9)	129 (10,5)
Otros DMARDs sintéticos convencionales No Metotrexate, n (%)	185 (55,9)	351 (39,2)	536 (43,7)
DMARDs sintéticos dirigidos o moléculas pequeñas, n (%)	8 (2,4)	5 (0,6)	13 (1,1)
Antimaláricos, n (%)	65 (19,6)	209 (23,4)	274 (22,4)

DE desviación estándar; IMC índice de masa corporal; Anti-CCP Antipéptido Citrulinado Cíclico; DMARDs medicamentos antirreumáticos modificadores de la enfermedad; PCR proteína C reactiva; HAQ Health Assessment Questionnaire; AINES antiinflamatorios no esteroideos

Predicción de la respuesta al tratamiento

En general, el desempeño de todos los modelos machine learning evaluados fue muy similar de acuerdo con todas las métricas utilizadas y mostraron mejores resultados que el modelo base. La exactitud y AUC se encontraron en los rangos entre 55 – 72% y 55 – 68%, respectivamente. La tabla 4 presenta las métricas de los 5 modelos considerados como los de

mejor desempeño durante el entrenamiento con validación cruzada. El modelo con mejores resultados fue el modelo basado en árboles de decisión: Random Forest, el cual obtuvo una exactitud de 72%, AUC 66%, sensibilidad de 88%, precisión 77% y puntaje F1 de 82%, (ver Tabla 4).

Tabla 4. Desempeño promedio de los modelos entrenados en la validación cruzada

Modelo		Exactitud (Accuracy)	AUC	Sensibilidad (Recall)	Precisión	Puntaje F1
RF	Random Forest Classifier	0.72	0.66	0.88	0.77	0.82
ET	Extra Trees Classifier	0.72	0.64	0.89	0.77	0.82
LGBM	Light Gradient Boosting Machine	0.72	0.67	0.87	0.78	0.82
GBC	Gradient Boosting Classifier	0.71	0.67	0.85	0.78	0.81
ADA	Ada Boost Classifier	0.70	0.63	0.82	0.78	0.80
Modelo basado en reglas		0.49	-	0.39	0.81	0.53

Para todos los modelos se observó un pobre rendimiento para clasificar correctamente la clase “No respondedor”. La sensibilidad más alta alcanzada para esta clase fue registrada por el modelo ADA con un 34.5%, mientras que la más baja fue registrada en el modelo ET con un 20%. Por su parte, para la clase “Respondedor” se logró una sensibilidad superior al 85% en todos los (ver Anexos: Figura S1).

Al realizar la evaluación del pronóstico sobre el conjunto de datos de prueba, todos los modelos mostraron un ligero deterioro en su comportamiento, con una disminución en al menos una de las métricas de desempeño. La figura 1 muestra la matriz de confusión para los modelos RF y ADA en los datos de prueba, obteniendo un AUC 62% y 60%, sensibilidad 84% y 82%, precisión 78% y 77% y puntaje F1 81% y 79%, respectivamente. Con igual proporción de falsos positivos (34% vs. 34%), pero una mayor proporción de falsos negativos en el modelo ADA (18% vs. 15%).

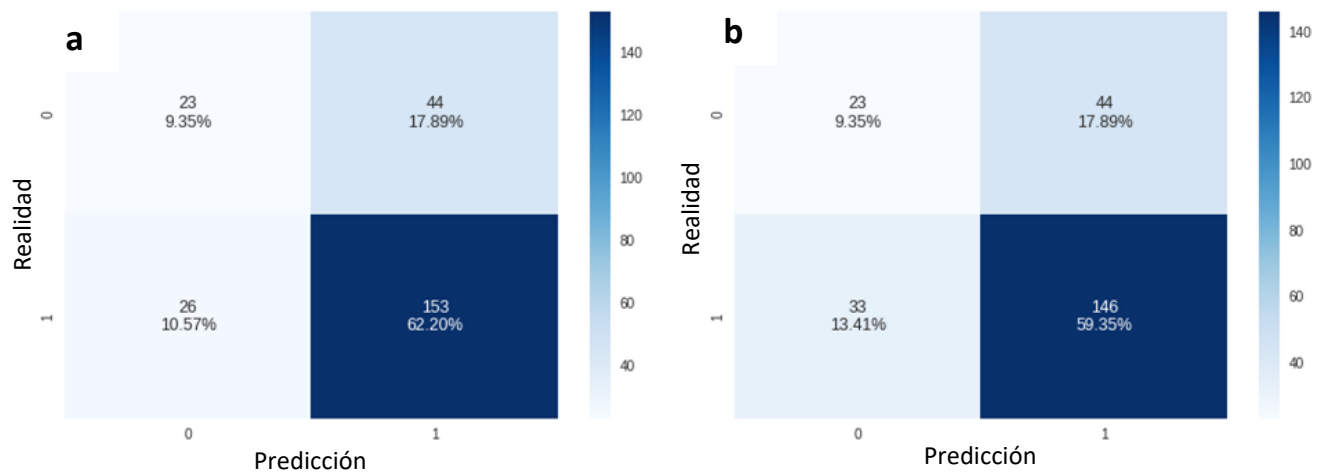


Figura 1. Matrices de confusión de los modelos RF (a) y ADA (b) en los datos de prueba.

Luego de realizar el ajuste de los hiperparámetros para la optimización de la exactitud y el AUC, explorando una potencial mejoría del rendimiento de los modelos, no se observó una recuperación significativa en las métricas de desempeño. Figuras 3 y Anexos: Figura S2.

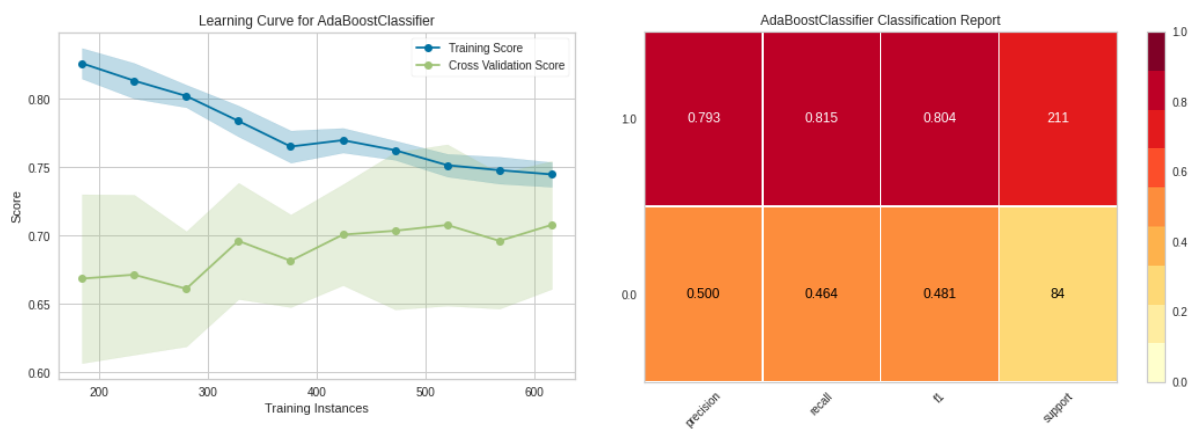


Figura 3. Curva de aprendizaje y reporte de clases del modelo ADA luego de la optimización de hiperparámetros.

Características importantes para la predicción de la respuesta al tratamiento

Se utilizó el método SHAP para determinar la influencia de las variables que contribuyeron a la respuesta en el modelo de predicción Random forest. El uso de DMARDs fue la característica más importante para predecir la respuesta, revelando que su uso se asocia a una contribución negativa, es decir, a la no respuesta al tratamiento. El DAS28 inicial fue la segunda característica en importancia, con bajos valores de DAS28 asociados a una contribución positiva (respuesta) y valores altos asociados a una contribución negativa (no respuesta). Esto mismo se observó sobre la puntuación en el cuestionario HAQ y la VSG que ocuparon el tercer y quinto lugar en importancia, respectivamente. Por otro lado, se observó que el no uso de corticoides y la presencia de comorbilidades aumentan la probabilidad de respuesta al tratamiento. La edad, la duración de la enfermedad, el FR y el IMC también se encontraron entre los principales factores que contribuyen a la respuesta. Las interpretaciones de la importancia de cada característica se muestran en la figura 5.



Figura 5. Importancia de las características que contribuyeron a la predicción de la respuesta al tratamiento.

DISCUSIÓN

Una adecuada respuesta al tratamiento en pacientes con AR se asocia con una mejoría a largo plazo de resultados relacionados con la funcionalidad, desenlaces reportados por el paciente y su supervivencia (14). El logro de la respuesta al tratamiento influye en gran manera en el manejo que se ofrece al paciente ya que permite una reducción gradual del mismo. Por otro lado, poder detectar tempranamente a aquellos individuos que son menos propensos a responder al tratamiento resulta decisivo para instaurar una terapia y un seguimiento más rigurosos. En el presente estudio se buscó desarrollar un modelo para predecir la respuesta al tratamiento en pacientes con AR e identificar los factores que más contribuyen a la misma, mediante modelos *machine learning*.

En general, los resultados mostraron una buena capacidad de los modelos para identificar y clasificar a los pacientes que potencialmente responderán al manejo instaurado, el método RF fue el que mostró ser más adecuado para predecir la respuesta al tratamiento con mayor exactitud; sin embargo, se evidenció una pobre capacidad para detectar a los no respondedores, con proporciones significativas de falsos positivos. Por un lado, esto puede ser explicado por la menor representación, en este estudio, de los pacientes que no responden al tratamiento, lo que pudo llevar a una reducción de la capacidad de generalización de los modelos; por otro lado, los mecanismos multifactoriales y poco claros de la AR, hacen pensar en la posible inclusión de un mayor número de individuos con cuadros clínicos ambiguos, poco definidos y muy disímiles en el grupo de no respondedores, esto sumado a: 1) la falta de información en variables relevantes para el pronóstico en estos pacientes, como la PCR, que en otros estudios ha sido descrita como una variable predictora de la respuesta (17) y 2) la no inclusión de otros factores que pueden actuar como mediadores de la “no respuesta”, como factores ambientales, genéticos y hormonales (18).

En todo caso, un resultado falso positivo en el escenario bajo estudio, debe ser estudiado con cuidado por los efectos que podría tener sobre el paciente, ya que juzgar que este último mostrará una adecuada respuesta al manejo que se le ofrece cuando la realidad será diferente,

puede traducirse en el deterioro significativo e irreversible de su condición, con un alto impacto en su salud y su calidad de vida. Mientras que un resultado falso negativo, si bien puede tener efectos sobre el paciente, relacionados, quizás, con eventos adversos por el uso de terapias, se podría pensar que el impacto más notorio en este caso se da a nivel del sistema de salud, por los costos innecesarios, y no tanto a nivel individual. Aquí hay que considerar que los eventos adversos mencionados pueden ser detectados y controlados a tiempo cuando los pacientes se encuentran bajo una estrategia de seguimiento estrecha.

El principal factor identificado como más relevante para la respuesta al tratamiento fue el uso de DMARDs; sin embargo, la contribución expuesta fue contraria a la esperada, ya que el uso de estos medicamentos modificadores de la enfermedad se asoció a una menor respuesta. La interpretación de este hallazgo se debe realizar con precaución, ya que es posible que se esté presentando un sesgo de confusión por indicación (19), de forma que a los pacientes con mayor actividad de la enfermedad sea a quienes se les prescribe con mayor frecuencia el uso de DMARDs, y a su vez estos muestren una menor respuesta al tratamiento relacionada con la severidad de su condición. Adicionalmente, es importante tener en cuenta la relativa baja frecuencia de uso de estas terapias por sus altos costos en Colombia, lo que puede limitar su uso en casos con menor actividad de la enfermedad (20). De hecho, los pacientes no respondedores tuvieron un DAS28 inicial y una puntuación en el HAQ mayores que aquellos clasificados como respondedores, esto indica que es probable que en el primer grupo existiera un mayor número de pacientes con actividad moderada o elevada de la enfermedad.

Lo anterior también está relacionado con la contribución del DAS28 inicial como la segunda característica más importante para la predicción, poniendo de manifiesto que valores altos de DAS28 se asocian a menor respuesta. Un resultado similar fue descrito por Koo y cols. (17), quienes reportaron el DAS28 como la segunda característica con mayor contribución a su modelo de predicción, con un valor SHAP de -0,091, mostrando una asociación negativa con la variable respuesta.

Los resultados del presente trabajo deben ser interpretados teniendo en cuenta sus limitaciones, entre las cuales se pueden mencionar las siguientes: a) el desbalance de clases,

con el grupo de no respondedores como clase minoritaria, puede afectar la capacidad de generalización de los modelos; b) es posible que las características de entrada hayan sido insuficientes para clasificar adecuadamente a pacientes como respondedores y no respondedores; c) para disminuir la pérdida de información se incluyeron variables con un porcentaje importante de valores nulos (>20%), aunque fueron imputados. Además, no se debe descartar la posibilidad de que otro tipo de modelamiento pueda explicar mejor la mencionada intrincada relación entre las características basales de los pacientes con AR y la respuesta al tratamiento, con menores proporciones de falsos positivos y negativos.

CONCLUSIONES

En el presente estudio se desarrollaron y evaluaron diferentes modelos de aprendizaje automático supervisado, se propone un modelo basado en árboles de decisión como el más adecuado para predecir la respuesta al tratamiento en pacientes con AR, teniendo reserva por su pobre capacidad para clasificar correctamente a los pacientes que no responden al manejo y su significativa proporción de falsos positivos. Las características que se identificaron como las más relevantes para la predicción de la respuesta son el uso de DMARDs, el DAS28 inicial y la puntuación en el cuestionario HAQ.

REFERENCIAS BIBLIOGRÁFICAS

1. Guzmán Moreno R, Restrepo Suárez JF. Artritis reumatoide temprana. *Revista Colombiana de Reumatología*. 2002;171–5.
2. Dadoun, S. et al. Mortality in rheumatoid arthritis over the last fifty years: Systematic review and meta-analysis. *Joint Bone Spine*. 2013;80:29–33.
3. Coenen, M. J. H. & Gregersen, P. K. Rheumatoid arthritis: A view of the current genetic landscape. *Genes Immun*. 2009;10: 101–111.
4. Raychaudhuri, S. et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. *Nat. Genet*. 2009;41:1313–1318.
5. Combe B, Rincheval N, Benessiano J, Berenbaum F, Cantagrel A, Daures JP, Dougados M, Fardellone P, Fautrel B, Flipo RM, et al. Five-year favorable outcome of patients with early rheumatoid arthritis in the 2000s: data from the ESPOIR cohort. *J Rheumatol*. 2013;40(10):1650–7.
6. Vodencarevic A, Tascilar K, Hartmann F, Reiser M, Hueber AJ, Haschka J, et al. Advanced machine learning for predicting individual risk of flares in rheumatoid arthritis patients tapering biologic drugs. *Arthritis Res Ther*. 2021;23(1):67.
7. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum*. 2004;50(5):1400–11.
8. Weinblatt ME, Kremer JM, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Fox RI, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor: Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *N Engl J Med*. 1999;340(4):253–
9. Kalweit M, Walker UA, Finckh A, Müller R, Kalweit G, Scherer A, et al. Personalized prediction of disease activity in patients with rheumatoid arthritis using an adaptive deep neural network. *PLoS One* [Internet]. 2021;16(6):e0252289. Disponible en: <http://dx.doi.org/10.1371/journal.pone.0252289>
10. Schett G, Emery P, Tanaka Y, Burmester G, Pisetsky DS, Naredo E, Fautrel B, van Vollenhoven R. Tapering biologic and conventional DMARD therapy in rheumatoid arthritis: current evidence and future directions. *Ann Rheum Dis*. 2016;75(8):1428–37.

11. Park YJ, Chung MK, Hwang D, Kim WU. Proteomics in rheumatoid arthritis research. *Immune Netw.* 2015;15(4):177–85.
12. Plant D, Barton A. Adding value to real-world data: the role of biomarkers. *Rheumatology (Oxford).* 2020;59(1):31–8.
13. Sutcliffe M, Radley G, Barton A. Personalized medicine in rheumatic diseases: how close are we to being able to use genetic biomarkers to predict response to TNF inhibitors? *Expert Rev Clin Immunol.* 2020;16(4):389–96.
14. Venerito V, Angelini O, Fornaro M, Cacciapaglia F, Lopalco G, Iannone F. A machine learning approach for predicting sustained remission in rheumatoid arthritis patients on biologic agents. *J Clin Rheumatol.* 2022;28(2):e334–9.
15. Mo X, Chen X, Li H, et al. Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning. *Front Pharmacol.* 2019;10:1155.
16. EULAR response criteria. (s/f). Das-score.nl. Recuperado el 11 de septiembre de 2022, de <https://www.das-score.nl/en/das-and-das28/das28-why/eular-response-criteria>
17. Koo, B. S., Eun, S., Shin, K., Yoon, H., Hong, C., Kim, D.-H., Hong, S., Kim, Y.-G., Lee, C.-K., Yoo, B., & Oh, J. S. (2021). Machine learning model for identifying important clinical features for predicting remission in patients with rheumatoid arthritis treated with biologics. *Arthritis Research & Therapy*, 23(1), 178.
18. Luque-Tévar M, Perez-Sanchez C, Patiño-Trives AM, Barbarroja N, Arias de la Rosa I, Abalos-Aguilera MC, et al. Integrative clinical, molecular, and computational analysis identify novel biomarkers and differential profiles of anti-TNF response in Rheumatoid Arthritis. *Front Immunol* [Internet]. 2021;12:631662. Disponible en: <http://dx.doi.org/10.3389/fimmu.2021.631662>
19. Kyriacou DN, Lewis RJ. Confounding by Indication in Clinical Research. *JAMA.* 2016;316(17):1818–1819. doi:10.1001/jama.2016.16435
20. Jáuregui, E., Bautista, W., Beltrán, A., Felipe, O., Fernández, A., Fernández, D., Jaller, J. J., Saldarriaga, L., & Toro, C. (2019). Asociación Colombiana de Reumatología. Consenso sobre recomendaciones para disminución y discontinuación de la terapia biológica en pacientes con artritis reumatoide, espondilitis anquilosante y artritis psoriásica. *Revista Colombiana de Reumatología*, 26(1), 11–23.

ANEXOS

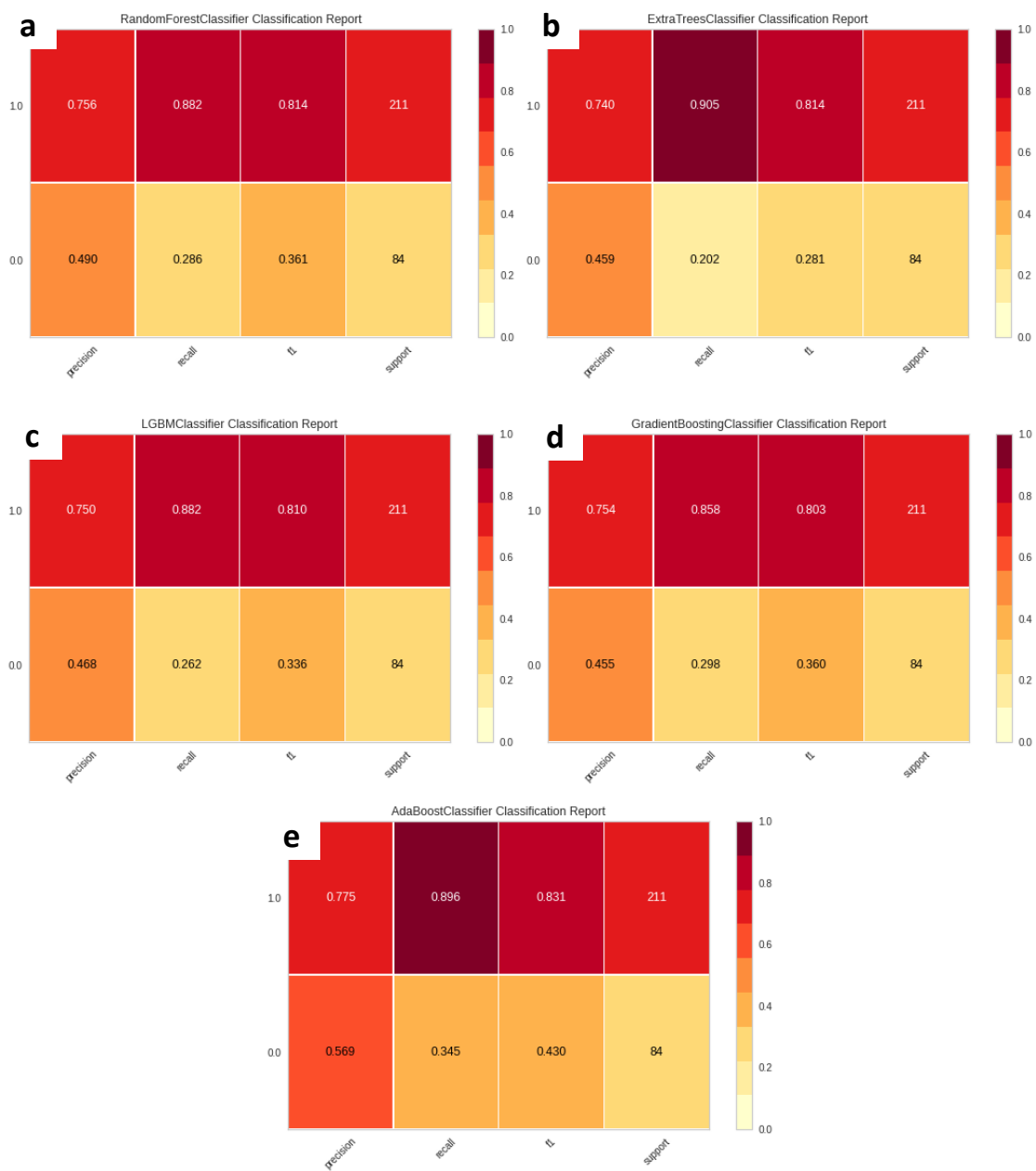


Figura S1. Reporte de clasificación de los modelos. a, modelo RF; b, modelo ET; c, modelo LGBM; d, modelo GBC; e, modelo ADA.

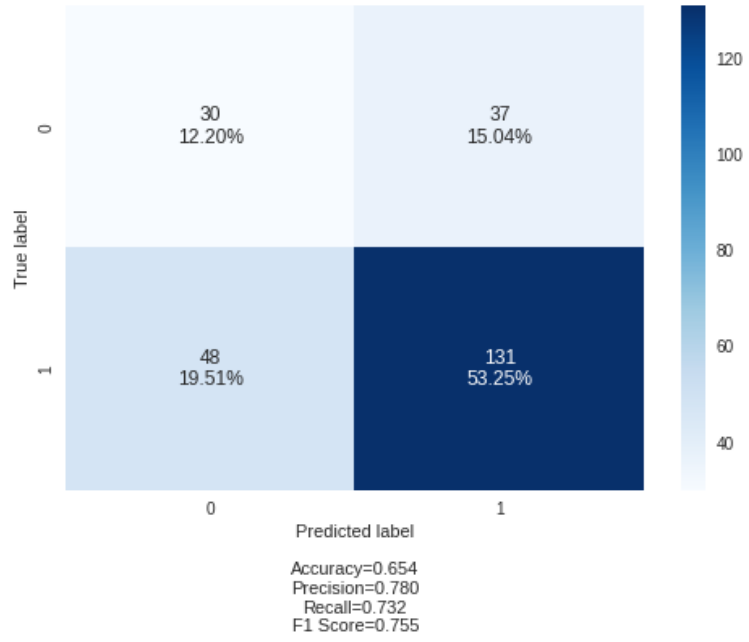


Figura S2. Matriz de confusión del modelo ADA optimizado frente a los datos de prueba