



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Predicción de tiempo en rango terapéutico para pacientes
con fibrilación auricular no valvular (FANV) anticoagulados
con antivitamina k (AVK)

Trabajo Fin de Grado

Grado en Ciencia de Datos

AUTOR/A: Perez Orero, Leandro

Tutor/a: Fuster i Garcia, Elies

Cotutor/a externo: LINARES GARCIA, MARIANO

CURSO ACADÉMICO: 2022/2023

Dedicatoria

A mis padres y hermana por su apoyo, paciencia y sacrificio, que ha permitido conseguir mis objetivos, tanto para este trabajo como también durante mi formación universitaria.

A mis amigos por acompañarme, y en especial durante estos meses.

Y al resto de mi familia por estar siempre a mi lado.

Agradecimientos

Especial y sincero agradecimiento a mis tutores, el Dr Mariano Linares y Elies Fuster, por su dedicación, constante apoyo y estímulo en la dirección de mi Trabajo Fin de Grado.

Al Servicio de Hematología y Hemoterapia del Consorcio Hospital General Universitario de Valencia por su colaboración en la aportación del material necesario e imprescindible para la realización de este trabajo.

Al Consorcio Hospital General Universitario y concretamente a su Directora Gerente Inmaculada Renart, por darme la oportunidad de realizar este trabajo. Pues sin su autorización hubiera sido imposible su desarrollo.



Resumen

Introducción: La fibrilación auricular (FA) es la arritmia cardiaca más frecuente y está relacionada con eventos embólicos. La eficacia y seguridad de la anticoagulación con antivitaminas K (AVK), entre ellos el acenocumarol, en la prevención del ictus depende del tiempo que el paciente se mantiene en rango terapéutico (TRT). Se han diseñado algunas escalas para predecir el comportamiento del TRT al inicio de la terapia. El objetivo de este trabajo fue evaluar la capacidad predictiva y discriminativa del *score* SAME-TT₂R₂ en una población caucásica, así como investigar y comparar diferentes modelos predictivos de TRT al inicio de la anticoagulación como herramienta para decidir si acenocumarol es el tratamiento adecuado o hay que recurrir a anticoagulantes orales de acción directa.

Material y métodos: Se ha realizado un estudio analítico de cohorte retrospectiva en 1000 pacientes con fibrilación auricular sin patología valvular (FANV) en tratamiento activo con acenocumarol entre 2007 y 2021 en el Consorcio Hospital General Universitario de Valencia (CHGUV), tras la previa recopilación y digitalización de 57 probables predictores de TRT. Se calculó el TRT durante los 6 meses siguientes al mes inicial de ajuste de dosis y se compararon distintos modelos predictivos de TRT, usando aproximaciones de regresión y clasificación (TRT: <60%, <65%, <70%), con la puntuación del *score* SAME-TT₂R₂. Tras el tratamiento de los datos (imputación de valores nulos, anómalos, sesgo, estandarización), se eliminaron 12 variables categóricas por su baja frecuencia en una de sus categorías, así como 13 variables por su alta correlación con otras. Se adoptaron 2 estrategias por lo que respecta a la selección de variables: Etapa 1 con las 32 restantes y el *score* SAME-TT₂R₂, y Etapa 2 con las 13 variables cuyas pruebas de significancia con el TRT mostraron un valor $p \leq 0.2$. Ambas estrategias se utilizaron en el entrenamiento y test de los siguientes modelos de *Machine Learning* (ML): regresión lineal/logística, árboles de decisión, redes neuronales (*Multilayer Perceptron*: MLP), *ensembles* (*Adaboost* y *Voting*), Máquinas de Vectores de Soporte (SVM) y algoritmos de los K Vecinos más Cercanos (KNN). Finalmente, se realizó una validación independiente en 50 pacientes recogidos de forma posterior a la selección de los modelos óptimos en la fase de test.

Resultados: Los resultados de los modelos de ML utilizando únicamente el *score* SAME-TT₂R₂ fueron mejorados al introducir otras variables adicionales. En cuanto al problema de regresión, los mejores resultados se obtuvieron con los modelos MLP y regresión lineal al incluir las 6 variables con un valor $p < 0.05$ en las pruebas de significancia con el TRT, explicando un 3.5% y 4% respectivamente de la variabilidad de los datos y los únicos modelos con un R² ajustado positivo. Para el problema de clasificación, los modelos más equilibrados en referencia al *f1 score*, precisión y *recall*/sensibilidad (métricas entre 0.5 y 0.6) fueron KNN y regresión logística, incluyendo las variables de la Etapa 2 en la predicción de TRT<60%; además de tener una *balanced accuracy* y área bajo la curva de aproximadamente 0.6.

Conclusiones: Este estudio ha evidenciado las limitaciones del *score* SAME-TT₂R₂ como factor predictor del TRT y que los modelos de ML pueden ayudar a mejorar la calidad de la predicción si se les alimenta con las variables más relevantes. A pesar de ello, ninguno de los modelos de ML desarrollados en este trabajo han conseguido suficiente capacidad predictiva. Estos resultados pueden haber estado influenciados por el tipo de cohorte (retrospectiva), el número de pacientes, el tipo de variables así como la falta de homogeneidad en los diagnósticos clínicos a constar en historial electrónico. Por lo que sería necesario futuros estudios prospectivos incluyendo mayor número de pacientes y otro tipo de variables.



Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

Palabras clave: *score* SAME-TT₂R₂, fibrilación auricular, tiempo en rango terapéutico, regresión, árboles de decisión, redes neuronales, *ensembles*, SVM, KNN, acenocumarol.

Abstract

Introduction: Atrial fibrillation (AF) is the most common cardiac arrhythmia and is associated with embolic events. The efficacy and safety of anticoagulation with antivitamin K (VKA), including acenocumarol, in the prevention of stroke depends on the time the patient remains in therapeutic range (TRT). Some scales have been designed to predict the behavior of TRT at the start of therapy. The aim of this study was to evaluate the predictive and discriminative capacity of the SAME-TT₂R₂ score in a Caucasian population, as well as to investigate and compare different predictive models of TRT at the start of anticoagulation as a tool to decide whether acenocumarol is the appropriate treatment or whether direct-acting oral anticoagulants should be used.

Material and methods: An analytical retrospective cohort study was performed in 1000 patients with atrial fibrillation without valvular pathology (NVAF) under active treatment with acenocumarol between 2007 and 2021 at the Consorcio Hospital General Universitario de Valencia (CHGUV), after prior collection and digitization of 57 probable predictors of TRT. TRT was calculated during the 6 months following the initial month of dose adjustment and different predictive models of TRT were compared, using regression and classification approaches (TRT: <60%, <65%, <70%), with the SAME-TT₂R₂ score. After data processing (imputation of null values, outliers, skew, standardization), 12 categorical variables were eliminated due to their low frequency in one of their categories, as well as 13 variables due to their high correlation with others. Two strategies were adopted with regard to variable selection: Stage 1 with the remaining 32 variables and the SAME-TT₂R₂ score, and Stage 2 with the 13 variables whose significance tests with the TRT showed a p value ≤ 0.2 . Both strategies were used in the training and testing of the following Machine Learning (ML) models: linear/logistic regression, decision trees, neural networks (Multilayer Perceptron: MLP), ensembles (Adaboost and Voting), Support Vector Machines (SVM) and K Nearest Neighbors (KNN) algorithms. Finally, an independent validation was performed on 50 patients collected after the selection of the optimal models in the test phase.

Results: The results of the ML models using only the SAME-TT₂R₂ score were improved by introducing additional variables. Regarding the regression problem, the best results were obtained with the MLP and linear regression models when including the 6 variables with a p-value < 0.05 in the significance tests with the TRT, explaining 3.5% and 4% respectively of the variability of the data and the only models with a positive adjusted R². For the classification problem, the most balanced models in reference to f1 score, precision and recall/sensitivity (metrics between 0.5 and 0.6) were KNN and logistic regression, including the Stage 2 variables in the prediction of TRT<60%; in addition to having a balanced accuracy and area under the curve of approximately 0.6.

Conclusions: This study has evidenced the limitations of the SAME-TT₂R₂ score as a predictor of TRT and that ML models can help improve the quality of prediction if fed with the most relevant variables. Despite this, none of the ML models developed in this work have achieved sufficient predictive ability. These results may have been influenced by the type of cohort (retrospective), the number of patients, the type of variables as well as the lack of homogeneity in the clinical diagnoses to be recorded in the electronic history. Future



prospective studies including a larger number of patients and other types of variables would be necessary.

Key words: SAME-TT₂R₂ score, atrial fibrillation, time in therapeutic range, regression, decision trees, neural networks, ensembles, SVM, KNN, acenocoumarol.



Índice de contenido

1. Introducción.....	1
1.1. Motivación.....	1
1.2. Objetivos.....	2
1.3. Impacto esperado.....	2
1.4. Metodología.....	3
1.5. Estructura.....	4
1.6. Colaboraciones.....	4
1.7. Convenciones.....	4
2. Estado del arte.....	5
2.1. Fibrilación auricular.....	5
2.1.1. Qué es la fibrilación auricular y su tratamiento anticoagulante.....	5
2.1.2. Control del tratamiento anticoagulante.....	6
2.1.3. Estado actual de las investigaciones sobre la predicción del tiempo en rango terapéutico (TRT).....	7
2.2. Machine Learning o aprendizaje automatizado.....	8
2.2.1. Introducción.....	8
2.2.2. Máquinas de Vectores de Soporte.....	9
2.2.3. Métodos de ensembles: Boosting (AdaBoost) y Voting.....	10
2.2.4. Redes neuronales: Multilayer Perceptron.....	11
2.2.5. Árboles de decisión.....	12
2.2.6. Algoritmo de los K Vecinos más Cercanos.....	13
2.2.7. Regresión: lineal y logística.....	14
2.3. Crítica al estado del arte.....	15
2.3.1. Crítica a la metodología utilizada en el score SAME-TT2R2.....	15
2.3.2. Crítica a los modelos predictivos en salud.....	16
2.4. Propuesta.....	17
3. Análisis del problema.....	17
3.1. Introducción del problema.....	17
3.2. Análisis de seguridad.....	18
3.3. Análisis energético o de eficiencia algorítmica.....	18



3.4. Análisis del marco legal y ético.....	18
3.4.1. Análisis de la protección de datos.....	18
3.4.2. Propiedad intelectual.....	19
3.4.3. Ética.....	19
3.5. Análisis de riesgos.....	19
3.6. Identificación y análisis de soluciones posibles.....	20
3.6.1. Detección precoz de un TRT subóptimo.....	20
3.6.2. Predicción del TRT.....	20
3.7. Solución propuesta.....	21
3.8. Plan de trabajo.....	21
4. Preparación y Comprensión de los Datos.....	23
4.1. Población.....	23
4.2. Variable a predecir: Tiempo en Rango Terapéutico (TRT).....	23
4.3. Variables potencialmente predictoras.....	24
4.3.1. Características identificativas.....	25
4.3.2. Características demográficas.....	25
4.3.3. Comorbilidades y factores de riesgo.....	25
4.3.4. Fármacos.....	27
4.3.5. Hallazgos de laboratorio.....	29
4.3.6. Hallazgos ecocardiográficos.....	29
4.3.7. Variables derivadas.....	30
4.3.8. Descriptiva de las características clínico-biológicas, analíticas y ecocardiográficas de los pacientes.....	31
4.4. Evaluación del score SAME-TT2R2.....	33
4.5. Preparación de datos.....	37
4.5.1. Reorganización de datos.....	37
4.5.2. Imputación de valores nulos o faltantes.....	38
4.5.3. Evaluación de valores anómalos.....	38
4.5.4. Análisis de distribución de variables numéricas.....	42
4.5.5. Tratamiento de valores anómalos.....	44
4.5.6. Estandarización de variables numéricas.....	44
4.5.7. Análisis bivariente entre TRT y variables numéricas.....	44



Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

4.5.8. Análisis univariante de variables categóricas.....	49
4.5.9. Análisis bivariante entre TRT y variables categóricas.....	50
4.5.10. Selección de variables.....	53
5. Entrenamiento y Test de los modelos predictivos.....	53
5.1. Modelos de clasificación binaria.....	54
5.1.1. Balanceo de datos.....	54
5.1.2. Métricas de rendimiento.....	54
5.1.3. Resultados con TRT < 60%.....	55
5.1.4. Resultados con TRT < 65%.....	56
5.1.5. Resultados con TRT < 70%.....	57
5.1.6. Curvas ROC (Receiver Operating Characteristic Curve).....	58
5.1.7. Selección de modelos de clasificación.....	59
5.1.8. Valor del score SAME-TT2R2 ≥ 3 para predecir TRT < 60%.....	61
5.2. Modelos de regresión.....	61
5.2.1. Métricas de rendimiento.....	61
5.2.2. Valor predictivo del score SAME-TT2R2 ≥ 3	62
5.2.3. Resultados incluyendo las 33 variables.....	63
5.2.4. Resultados con las 13 variables seleccionadas.....	64
5.2.5. Resultados con las 6 variables significativas.....	66
5.2.6. Selección de modelos de regresión.....	67
5.3. Tiempos de ejecución.....	67
6. Validación de los modelos predictivos óptimos.....	69
6.1. Modelos de clasificación.....	69
6.2. Modelos de regresión.....	70
7. Conclusiones.....	73
7.1. Legado.....	74
7.2. Relación del trabajo desarrollado con los estudios cursados.....	74
8. Trabajos futuros.....	75
9. Referencias.....	77
10. Anexos: ODS.....	80
11. Glosario de abreviaciones.....	81



Índice de tablas

TABLA 1: Fármacos considerados como leves/moderados/graves en la interacción con acenocumarol...	28
TABLA 2: Variables incluidas en el score SAME-TT ₂ R ₂ y su aportación.....	30
TABLA 3: Descripción de las características de las variables categóricas en el conjunto de datos (total, entrenamiento y prueba)	32
TABLA 4: Descripción de las características de las variables numéricas en el conjunto de datos (total, entrenamiento y prueba).....	33
TABLA 5: Datos estadísticos del TRT por puntuación del <i>score</i> SAME-TT ₂ R ₂	34
TABLA 6: Pruebas de significancia aplicadas según el tipo de variable.....	35
TABLA 7: N° pacientes (%) en rango subóptimo y mediana del TRT por puntuación del SAME-TT ₂ R ₂	35
TABLA 8: N° pacientes (%) en rango subóptimo y mediana del TRT según SAME-TT ₂ R ₂ < 2 o ≥ 2.....	36
TABLA 9: N° pacientes (%) en rango subóptimo y mediana del TRT según SAME-TT ₂ R ₂ < 3 o ≥ 3.....	36
TABLA 10: N° pacientes (%) en rango subóptimo y mediana del TRT según SAME-TT ₂ R ₂ < 4 o ≥ 4.....	37
TABLA 11: Rangos de valores no considerados anómalos para cada variable según RIQ y <i>Z-score</i>	41
TABLA 12: Sesgo de las variables numéricas.....	43
TABLA 13: Valor p de la prueba de <i>Mann-Whitney U</i> de las variables categóricas (vs TRT) con p≤0.2....	51
TABLA 14: Correlación de Pearson de las variables categóricas con r > 0.7.....	53
TABLA 15: Resultados de clasificación para los modelos con 33 y 13 variables con TRT < 60%.....	56
TABLA 16: Resultados de clasificación para los modelos con 33 y 13 variables con TRT < 65%.....	57
TABLA 17: Resultados de clasificación para los modelos con 33 y 13 variables con TRT < 70%.....	58
TABLA 18: Resultados de clasificación para los modelos con el <i>score</i> SAME-TT ₂ R ₂ ≥3 y TRT<60%....	61
TABLA 19: Resultados de regresión para los modelos con SAME-TT ₂ R ₂ ≥ 3.....	63
TABLA 20: Resultados de regresión para los modelos con 33 variables.....	64
TABLA 21: Resultados de regresión para los modelos con 13 variables.....	64
TABLA 22: Resultados de regresión para los modelos con 6 variables.....	66
TABLA 23: Comparación del n° de parámetros libres y tiempos de ejecución (s) para ambas etapas y todos los modelos (clasificación).....	68
TABLA 24: Comparación del n° de parámetros libres y tiempos de ejecución (s) para los 3 conjuntos de variables y todos los modelos (regresión).....	69



Índice de figuras

FIGURA 1: Efecto de la FA en el ritmo sinusal.....	6
FIGURA 2: Riesgos de tener un control INR alto/bajo.....	7
FIGURA 3: Ramas del <i>Machine Learning</i>	9
FIGURA 4: Ejemplos de SVC con diferentes kernels.....	10
FIGURA 5: Ejemplo de SVR.....	10
FIGURA 6: Ejemplo de <i>AdaBoost</i> en un problema de clasificación.....	11
FIGURA 7: Ejemplo de arquitectura <i>Voting</i> en un problema de clasificación/regresión.....	11
FIGURA 8: Ejemplo de arquitectura <i>Multilayer Perceptron</i>	12
FIGURA 9: Ejemplo de arquitectura de un árbol de decisión.....	13
FIGURA 10: Ejemplo de KNN en un problema de clasificación.....	14
FIGURA 11: Ejemplo de modelo de regresión lineal y logística.....	15
FIGURA 12: Diagrama de Gantt sobre el plan de trabajo (en días).....	22
FIGURA 13: Distribución del Tiempo en Rango Terapéutico (TRT) en el conjunto de datos.....	24
FIGURA 14: Comparación de la distribución de TRT según cada puntuación del SAME-TT ₂ R ₂	34
FIGURA 15: Comparación de la distribución de TRT según SAME-TT ₂ R ₂ ≥ 3 y < 3.....	36
FIGURA 16: Número de valores considerados anómalos por variable según RIQ.....	39
FIGURA 17: Número de valores considerados anómalos por variable según <i>Z-score</i>	40
FIGURA 18: Distribución original de las variables numéricas.....	42
FIGURA 19: Distribución de las variables numéricas tras las transformaciones.....	43
FIGURA 20: Varianza explicada acumulada del PCA con las variables numéricas.....	45
FIGURA 21: Importancia de las variables numéricas en las primeras 2 componentes principales y TRT < 60%.....	46
FIGURA 22: Mapa de calor de la correlación de <i>Spearman</i> con las variables numéricas y la variable objetivo TRT.....	47
FIGURA 23: Mapa de calor de la correlación de <i>Pearson</i> con las variables numéricas y la variable objetivo TRT.....	48
FIGURA 24: Grado de asociación entre las variables numéricas con alta correlación ($r > 0.7$ o < -0.7).....	49
FIGURA 25: Frecuencia de las variables categóricas con baja representación en alguna de sus clases (<5%).....	50
FIGURA 26: Comparación de medianas entre clases de las variables binarias con valor $p \leq 0.2$	52



FIGURA 27: Curvas ROC para todos los modelos con puntos de corte de TRT: 60%, 65% y 70%.....	59
FIGURA 28: Curvas de calibración para KNN y regresión logística con TRT<60%.....	60
FIGURA 29: Curvas de <i>precision-recall</i> para KNN y regresión logística con TRT<60%.....	60
FIGURA 30: Comparación de residuos de los modelos de regresión con 13 variables.....	65
FIGURA 31: Comparación de residuos de los modelos de regresión con 6 variables.....	67
FIGURA 32: Matriz de confusión y curva ROC para regresión logística y TRT<60%.....	70
FIGURA 33: Matriz de confusión y curva ROC para KNN y TRT<60%.....	70
FIGURA 34: Análisis de residuos para MLP con 6 variables.....	72
FIGURA 35: Análisis de residuos para regresión lineal con 6 variables.....	72



1. Introducción

Durante los últimos veinte años, la fibrilación auricular (FA) parece haber incrementado y seguirá aumentando en los próximos 30 años [1], y a pesar de los avances en su tratamiento, sigue siendo una de las principales causas de accidente cerebrovascular y morbimortalidad cardiovascular en el mundo, convirtiéndose en uno de los mayores desafíos de salud pública.

Al tratarse de una patología trombotogénica, en la mayoría de los casos es necesaria la anticoagulación [2]. Los agentes más usados son los antagonistas de la vitamina K o antivitamina K (AVK) y su actividad anticoagulante se determina mediante el Índice Normalizado Internacional (INR). Los AVK han demostrado ser altamente efectivos reduciendo el ictus y la mortalidad en la FA. Sin embargo, su seguridad y efectividad están condicionadas por el adecuado control de los niveles de anticoagulación, reflejado en el Tiempo en Rango Terapéutico (TRT) del INR. Un INR por debajo del rango terapéutico disminuye el efecto antitrombótico, mientras que niveles superiores al mismo incrementan el riesgo hemorrágico.

Así pues, en los últimos años han empezado a surgir intentos de predicción de TRT basados en *Machine Learning* (ML), destacando el *score* SAME-TT₂R₂ [3, 4]. Aunque no hay ningún método consolidado en la práctica clínica.

En este Trabajo Fin de Grado (TFG) se estudia y se analiza la anticoagulación oral en pacientes con fibrilación auricular sin patología valvular (FANV) con el objetivo de intentar predecir el TRT que tendría un paciente con unas características determinadas. Para ello, se ha investigado ML con diferentes técnicas y algoritmos que pueden ayudar a solucionar este problema.

1.1. Motivación

La FA es una patología trombotogénica en la que los AVK han demostrado ser altamente efectivos. Sin embargo, no están exentos de riesgos que hacen que su seguridad y efectividad estén determinadas por un adecuado TRT. Se estima que hasta un 50% de los pacientes con FANV anticoagulados con AVK presentan un TRT subóptimo (<60-70%, según diferentes publicaciones previas), por lo que en estos casos estaría indicado el cambio a un anticoagulante oral de acción directa (ACOD) como alternativa terapéutica a los AVK [5]. Ello ha generado interés por predecir si un paciente recién diagnosticado de una FANV va a presentar o no, en los próximos meses, un adecuado TRT de anticoagulación, que hasta la fecha los estudios previos no han logrado precisar. Por tanto, existe un nicho de conocimiento sobre el que es interesante investigar y en cuyo estudio la ciencia de datos y las herramientas de inteligencia artificial (IA) pueden aportar luz en este campo.

Además, la realización de un trabajo como el presente contiene características muy relevantes para la formación de un científico de datos y adquisición de conocimiento en análisis de datos biomédicos masivos (monitorización, predicción, etc.). Conocimiento de ámbito clínico que solamente se puede acceder a través de la colaboración con un centro sanitario como el Consorcio Hospital General Universitario de Valencia (CHGUV), que ha proporcionado los medios y datos necesarios para el desarrollo de este trabajo de carácter multidisciplinar e intersectorial (clínico y técnico).

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

El sistema sanitario público requiere soporte en ciencia de datos / estadística dada la enorme cantidad y complejidad de los datos que constan en las historias clínicas de los pacientes así como en temas de logística e informatización de equipos y sistemas. Debido al alto grado de informatización existente en todo el equipamiento médico como por ejemplo historias clínicas, es necesario el empleo de herramientas y metodología especialmente diseñadas para el tratamiento de cada vez mayores cantidades de datos. Por ello, la Ciencia de Datos también es clave para el desarrollo de proyectos de investigación relacionados con temas de salud, en muchos casos proyectos multidisciplinares, multicéntricos, nacionales o internacionales.

1.2. Objetivos

El objetivo principal que se desea alcanzar en este estudio es obtener un modelo predictivo que sea capaz de determinar el TRT (tiempo en rango terapéutico) en pacientes con FANV anticoagulados con AVK (acenocumarol).

Para ello, se plantearon los siguientes objetivos:

- Evaluar la capacidad predictora del *score* SAME-TT₂R₂ para determinar el TRT subóptimo en una muestra propia recopilada en el presente TFG.
- Estimar el valor añadido del uso de modelos más complejos basados en técnicas de *Machine Learning* (ML), así como la inclusión de datos adicionales al *score* SAME-TT₂R₂.

1.3. Impacto esperado

En FANV, los ACOD muestran en general un mejor perfil de seguridad (reducción de complicaciones hemorrágicas) y también de efectividad (reducción de complicaciones tromboembólicas) que los AVK, así como resultan ser coste-efectivos en más del 70% de los casos según publicaciones previas [6]. Sin embargo, en España, no está financiada la sustitución generalizada de acenocumarol por los ACOD. En los pacientes con FANV anticoagulados con AVK que presentan un TRT subóptimo durante los últimos 6 meses previos, sí que está recomendado, aprobado y financiado el cambio a un ACOD. Pero dado que la mayoría de ensayos controlados aleatorizados han comparado ACOD vs warfarina y pocos han comparado ACOD vs acenocumarol [7], debe tenerse en cuenta el impacto presupuestario de la sustitución de este último (aproximadamente entre 10.35-13.8 €/mes) por los ACOD (aproximadamente entre 80-100 €/mes, dependiendo del ACOD administrado y su dosis ajustada). En este sentido, la evaluación económica más reciente en España [8], financiada por el fabricante de dabigatrán, determinó que el coste total por paciente con FANV (durante toda la vida del paciente) fue más bajo con acenocumarol (22230 €) que con los ACOD (24564 € con dabigatrán, 24655 € con apixabán y 25900 con rivaroxabán), pero los ACOD aportaron más años de vida ajustados por calidad (AVAC) que acenocumarol. Teniendo en cuenta el umbral para España de 20000 €/AVAC, la mejor ratio coste-efectividad incremental frente a acenocumarol fue para dabigatrán (6397 €/AVAC), seguido por apixabán (8039 euros/AVAC).



Por todo ello, el disponer de un modelo predictivo de TRT rápido y robusto que identifique los grupos de pacientes candidatos para recibir desde su diagnóstico estos nuevos tratamientos (ACOD) permitiría establecer una estrategia racional eficaz, segura y coste-efectiva para su uso. De esta manera, se evitaría que un grupo importante de pacientes con FANV anticoagulados con AVK permanecieran durante una parte sustancial del periodo inicial de tratamiento fuera de rango terapéutico con el riesgo de eventos trombóticos y/o hemorrágicos que ello conlleva.

Además, el presente proyecto de investigación en el campo de la predicción del TRT de los AVK tiene una clara relación con el Objetivo de Desarrollo Sostenible (ODS) número 3: Salud y bienestar. Este objetivo se centra en garantizar una vida saludable y promover el bienestar para todas las personas en todas las edades. El proyecto contribuye a este objetivo al abordar un aspecto crucial de la salud, que es la eficacia del tratamiento anticoagulante en pacientes con FANV. Al mejorar la precisión en la predicción del TRT, se puede lograr una mejor gestión del tratamiento y, en última instancia, mejorar la salud y el bienestar de los pacientes de una forma racional.

A ello hay que añadir que al utilizar técnicas avanzadas de análisis de datos y modelos de predicción, el proyecto también se alinea con el ODS 9: Industria, innovación e infraestructura. La aplicación de algoritmos de aprendizaje automático y el desarrollo de modelos predictivos más sofisticados implican un enfoque innovador que puede mejorar significativamente la calidad de la atención médica y la toma de decisiones clínicas.

1.4. Metodología

Se realizó un estudio analítico de cohorte retrospectiva en 1000 pacientes con FANV en tratamiento activo con acenocumarol entre 2007 y 2021, registrados en el programa informático GOTA del tratamiento anticoagulante en el CHGUV. Se calculó el TRT como el porcentaje de INRs en rango (INR= 2-3) durante los 6 meses siguientes al mes inicial de ajuste de dosis.

Se recogieron las características de los pacientes:

- Las incluidas en el *score* SAME-TT₂R₂: sexo, edad < 60 años, historial médico con más de dos comorbilidades, tratamiento con amiodarona, tabaquismo y etnia.
- Otra serie de variables con posible potencial predictivo en base a lo referido en estudios previos, incluyendo variables sociodemográficas, clínicas, ecocardiográficas y analíticas al diagnóstico

Se abordaron diferentes enfoques para analizar la relación entre el TRT (mediana, <60%, <65% y <70%) y el *score* SAME-TT₂R₂ (sistema de puntuación original entre 0-8 puntos, ≥ 2 , ≥ 3 y ≥ 4), mediante la prueba chi cuadrado / Fisher en el caso del TRT clasificado según puntos de corte y Kruskal-Wallis / Mann-Whitney U para la comparación entre medianas del TRT.

Tras el pretratamiento de los datos (imputación de valores nulos y anómalos, análisis del sesgo y estandarización) se adoptaron dos estrategias por lo que respecta a la selección de variables.

- Etapa 1: 33 / 58 variables (incluyendo SAME-TT₂R₂) fueron seleccionadas tras la eliminación de 25 de ellas (12 por baja frecuencia y 13 por multicolinealidad).

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

- Etapa 2: 13 / 33 variables cuyas pruebas de significancia con el valor de TRT (*Kruskal-Wallis* o *Mann-Whitney U* para variables categóricas y correlación de *Spearman* y *Pearson* para variables continuas) mostraron un valor $p \leq 0.2$.

Ambas estrategias se utilizaron en el entrenamiento (800 pacientes) y prueba (200 pacientes) de varios modelos de aprendizaje automático o ML para su posterior comparación entre ellos y con la capacidad predictora del *score* SAME-TT₂R₂. Entre estos modelos se incluyeron: regresión lineal/logística, árboles de decisión, redes neuronales (*Multilayer Perceptron* o MLP), *ensembles* (*Voting* y *AdaBoost*), Máquinas de Vectores de Soporte o *Support Vector Machines* (SVM) y algoritmos de los K Vecinos más Cercanos o *K-Nearest Neighbors* (KNN). En cuanto a la validación final de los modelos, se utilizaron otros 50 pacientes distintos a los del conjunto inicial de datos para realizar una validación interna.

Como lenguaje de programación se utilizó *Python* versión 3.9.4 (en *Visual Studio Code*) y *Excel* para el almacenamiento de los datos.

1.5. Estructura

Este documento se estructura de la siguiente manera. En el apartado 1 se introduce el problema. El apartado 2 presenta el estado actual del control del tratamiento anticoagulante en la FANV así como su relación con los métodos de aprendizaje automático. Se describe cómo estos métodos pueden ser aplicados en este contexto para predecir el tiempo en rango terapéutico. Se presenta así mismo un listado de algoritmos y cómo funcionan. Se expone una crítica a lo ya existente, justificando la propuesta para llevar a cabo este TFG. En el apartado 3 se analiza el problema, incluyendo análisis del marco legal y ético, y presenta soluciones posibles exponiendo, por último, la solución propuesta en este trabajo. El apartado 4 muestra el proceso de obtención del conjunto de datos con el que finalmente se va a trabajar. Se describe el proceso completo de obtención, comprensión, limpieza, transformación, integración y preparación de los datos. El apartado 5 presenta los resultados obtenidos tras el entrenamiento y test de los algoritmos seleccionados, incluyendo las herramientas y métricas que se han utilizado para analizar su rendimiento. El apartado 6 presenta la validación de dichos algoritmos predictivos en un subgrupo de pacientes. El apartado 7 recoge las conclusiones, legado así como la relación del trabajo desarrollado con los estudios cursados; y finalmente, en el apartado 8 se analizan posibles trabajos futuros. También se incluye un glosario de abreviaciones (apartado 11), anexos (apartado 10) y referencias (apartado 9) al final del TFG.

1.6. Colaboraciones

El presente TFG se ha realizado en colaboración con la Universitat Politècnica de València (Escola Tècnica Superior d'Enginyeria Informàtica) y la empresa CHGUV (Servicio de Hematología y Hemoterapia), de cuyos ficheros electrónicos se han recogido los datos de los pacientes necesarios para la realización de este trabajo.

1.7. Convenciones

Las palabras extranjeras se remarcaron en cursiva.



2. Estado del arte

En este capítulo se establece el punto de partida del trabajo realizado. Primero, se describe este punto de partida desde el punto de vista del hecho a analizar qué es la fibrilación auricular y su tratamiento anticoagulante, así como el control de dicho tratamiento. Posteriormente se describen los algoritmos de aprendizaje automático apropiados para ser aplicados sobre el conjunto de datos. Por último se analiza el sistema de puntuación más ampliamente utilizado para predecir TRT (el *score* SAME-TT₂R₂) y se propone que diferencias podría aportar este trabajo para intentar mejorar lo ya existente.

2.1. Fibrilación auricular

2.1.1. Qué es la fibrilación auricular y su tratamiento anticoagulante

La FA es la arritmia cardíaca crónica más frecuente [9] (Figura 1), y a pesar de los avances en su tratamiento, sigue siendo una de las principales causas de accidente cerebrovascular, insuficiencia cardíaca, muerte súbita y morbimortalidad cardiovascular en el mundo.

Durante los últimos veinte años, la FA parece haber incrementado [1], probablemente debido al envejecimiento de la población, al incremento de enfermedades crónicas y a la optimización en técnicas para detectarla. Su prevalencia es del 1-2% de la población [10], y se incrementa con la edad estimándose que la padece un 8,5% de la población española mayor de 60 años [11].

La complicación mayor relacionada con la FA es el accidente isquémico cerebrovascular, aumentando 5 veces el riesgo de ictus [12]. En comparación con los ictus causados por otras etiologías, los ictus asociados a la FA tienen mayor mortalidad, dejan más secuelas y tienen un mayor riesgo de recurrencias [13].

Los factores de riesgo habituales se resumen en la escala clínica de riesgo CHA₂DS₂-VASc (insuficiencia cardíaca, hipertensión, edad >65 años, diabetes mellitus, ictus, enfermedad vascular y sexo femenino) [14]. Al tratarse de una patología trombogénica, en la mayoría de los casos (CHA₂DS₂-VASc ≥2) es necesaria la anticoagulación oral [2].

Los AVK [acenocumarol (Sintrom®) y warfarina (Aldocumar®)] han sido durante décadas la única opción disponible para la terapia anticoagulante oral en la prevención de las complicaciones tromboembólicas de la FA, siendo el representante más usado en España el acenocumarol (Sintrom®), al contrario que en los países anglosajones, donde el AVK más utilizado es la warfarina (Aldocumar®). Pero, en los últimos años se han desarrollado los ACOD (dabigatrán, rivaroxabán, apixabán y edoxabán) que han demostrado un beneficio-riesgo favorable en la prevención del ictus y la embolia sistémica en pacientes con FA sin patología valvular (FANV), definiéndose la FANV como aquella que no se asocia a valvulopatía significativa [15]. En cambio, en pacientes con FA valvular se sigue recomendando como tratamiento anticoagulante de elección los AVK.

En este sentido, en septiembre de 2012, la Agencia Española del Medicamento (AEMPS) publicó un primer informe de posicionamiento terapéutico sobre el uso de estos medicamentos (AVK versus ACOD) en FANV, actualizado por última vez en noviembre de 2016 [5]. En este

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

informe se recomienda el uso de AVK en: 1) pacientes ya en tratamiento con AVK y buen control de INR; 2) nuevos pacientes con FANV en los que esté indicada la anticoagulación; 3) fibrilación auricular con afectación valvular. Por el contrario, serán de elección los ACOD en: 1) pacientes con hipersensibilidad conocida o con contraindicación específica al uso de acenocumarol o warfarina; 2) pacientes con antecedentes de hemorragia intracraneal (excepto durante la fase aguda) en los que se valore que los beneficios de la anticoagulación superan el riesgo hemorrágico; 3) pacientes con ictus isquémico que presenten criterios clínicos y de neuroimagen de alto riesgo de hemorragia intracraneal; 4) pacientes en tratamiento con AVK que sufren episodios tromboembólicos arteriales graves a pesar de un buen control del INR; 5) pacientes que han iniciado tratamiento con AVK en los que no es posible mantener un control de INR dentro de rango (2-3) a pesar de un buen cumplimiento terapéutico; 6) imposibilidad de acceso al control de INR convencional.

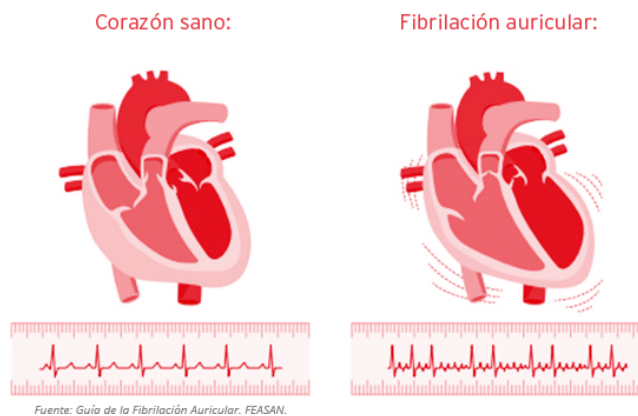


Figura 1. Efecto de la FA en el ritmo sinusal (cardioalianza.org)

2.1.2. Control del tratamiento anticoagulante

Aunque existe amplia experiencia clínica con los AVK, la realidad es que poseen varias desventajas (estrecha ventana terapéutica, variabilidad del efecto anticoagulante, controles periódicos de la anticoagulación, ajustes frecuentes de dosis) que pueden limitar su uso en la práctica clínica.

En este sentido, el acenocumarol tiene abundantes interacciones medicamentosas por su metabolismo y grandes variaciones en su biodisponibilidad. Además, tiene un índice terapéutico muy estrecho, dado que menores dosis pueden provocar disminución del efecto del fármaco, mientras que dosis mayores pueden provocar hemorragias por aumento del efecto del fármaco (Figura 2). Para monitorizar su efecto y evitar un riesgo innecesario de eventos tromboembólicos y/o hemorrágicos, se utiliza el tiempo de protrombina, el cual se prolonga con la acción del fármaco. Debido a que los parámetros de tiempo de protrombina varían de un centro a otro, se creó el *International Normalized Ratio* (INR), el cual nos sirve para monitorizar el efecto anticoagulante del fármaco. En el caso de la FANV se debe mantener el INR entre 2 y 3 para una anticoagulación óptima. Como el INR es un dato de laboratorio que nos muestra el efecto del acenocumarol en un momento en el tiempo, éste puede tener fluctuaciones por lo que se deben hacer controles periódicos para mantener por mayor tiempo el rango terapéutico [16]. Una forma de monitorizar la efectividad de tratamiento es el cálculo del TRT a partir del porcentaje de INRs que se encuentran en el rango terapéutico [17]. Se

recomienda que el TRT esté por encima del 70%, de acuerdo a las Guías Europeas de la FA; aunque otras guías proponen que este rango sea mayor a 60 ó 65% [18].

Respecto a los ACOD, éstos tienen una cinética y un efecto anticoagulante predecible además de una ventana terapéutica amplia, por lo que no es necesario monitorizar la actividad anticoagulante y se pueden prescribir a dosis fijas, de acuerdo a ciertas características clínicas [19]. Todo esto hace que sean cómodos y seguros para los pacientes.

Sin embargo, los ACOD también presentan inconvenientes como la falta de un antídoto específico (salvo en el caso del dabigatrán), la dificultad de asegurar el cumplimiento terapéutico al no requerir monitorización, y su mayor coste.

En resumen, los AVK han demostrado ser altamente efectivos reduciendo el ictus y la mortalidad en la FA. Su seguridad y efectividad están determinadas por el adecuado control de los niveles de anticoagulación, reflejado en el TRT. En este sentido y en el marco del Sistema Nacional de Salud, la elección del anticoagulante en pacientes con FANV en los que el control INR es subóptimo (TRT<60-65%) sería uno de los ACOD aprobados y financiados en estos casos [5]. Por lo que se requiere de un modelo predictivo de TRT para cubrir la necesidad de establecer unas pautas claras de manejo de la anticoagulación en pacientes con FANV.

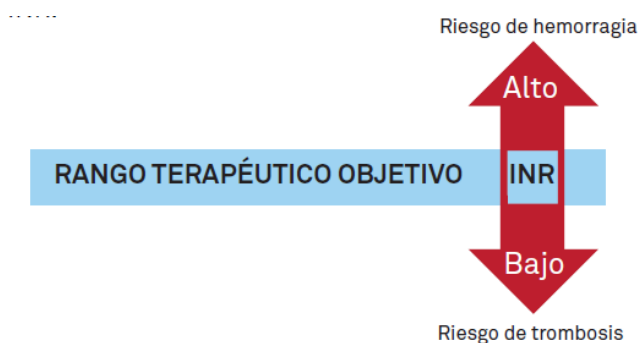


Figura 2. Riesgos de tener un control INR alto/bajo (osakidetza.euskadi.eus)

2.1.3. Estado actual de las investigaciones sobre la predicción del tiempo en rango terapéutico (TRT)

En medicina, hay una tendencia hacia la medicina individualizada pues las enfermedades suelen mostrar diferencias individuales en lo que respecta a su presentación y curso, así como respuesta al tratamiento. Ello justifica un enfoque individualizado para el tratamiento de las enfermedades que sea capaz de predecir respuestas individuales. El tratamiento con AVK ha sido uno de los escenarios tradicionales de individualización del tratamiento debido a la variabilidad interindividual en la respuesta a estos fármacos.

En este sentido, el artículo publicado por Apostolakis et al [3] ha sido una contribución en este campo mostrando la validación de un *score* simple y fácil de usar, el *score* SAME-TT₂R₂ (sexo femenino, edad < 60 años, historial médico con más de dos comorbilidades, tratamiento con amiodarona, tabaquismo con doble puntuación y etnia no caucásica con doble puntuación)

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

para predecir si el paciente va a mantener un buen rango de anticoagulación ($score < 2$) o en caso contrario, va a presentar un TRT subóptimo ($\geq 2 score$).

Otras condiciones no incluidas el $score$ SAME-TT₂R₂ pero que han mostrado asociación independiente con un menor TRT según estudios previos [6, 20 - 24] han sido: analfabetismo funcional, HAS-BLED, sangrado previo, trombocitopenia, tromboembolismo venoso, neoplasia, demencia, aspirina, uso de medicamentos para el dolor, albúmina, porcentaje de neutrófilos, recuento de glóbulos rojos, ancho de distribución eritrocitaria (ADE), índice de masa corporal (IMC), alcoholismo y anemia.

2.2. Machine Learning o aprendizaje automatizado

2.2.1. Introducción

En los últimos años, el campo del *Machine Learning* (ML) ha experimentado un crecimiento exponencial y se ha convertido en una disciplina fundamental en el ámbito de la IA (Figura 3). El ML se basa en la idea de que los sistemas informáticos pueden aprender y mejorar automáticamente a partir de datos sin ser programados explícitamente [25].

El ML se ha convertido en una herramienta poderosa para extraer información y conocimiento a partir de grandes volúmenes de datos. En lugar de seguir enfoques tradicionales basados en reglas o algoritmos explícitos, los algoritmos de ML aprenden patrones y estructuras a partir de los datos, lo que les permite realizar tareas complejas de manera automatizada. Esto ha llevado a avances significativos en áreas como la clasificación de datos, la predicción, la detección de anomalías y la toma de decisiones.

A medida que el ML continúa evolucionando, también se están explorando nuevas técnicas y enfoques. Por ejemplo, el aprendizaje profundo (*Deep Learning*) ha surgido como una rama del ML que se basa en redes neuronales artificiales, con múltiples capas, capaces de aprender representaciones de alto nivel de los datos.

Dentro del aprendizaje automático se pueden diferenciar dos tipos: aprendizaje supervisado y aprendizaje no supervisado. Esta clasificación se realiza en función de la cantidad y el tipo de supervisión que reciben los algoritmos durante el entrenamiento [26].

Entre los enfoques más comunes se encuentran los modelos de aprendizaje supervisado, donde se utilizan datos etiquetados para entrenar un modelo que pueda realizar predicciones precisas sobre nuevos datos. Dentro del aprendizaje supervisado existen dos aproximaciones, clasificación y regresión. En la clasificación, el objetivo es predecir la clase de una etiqueta; y en la regresión, el objetivo es predecir un número continuo. Estos modelos de aprendizaje supervisado incluyen algoritmos como árboles de decisión, regresión logística, SVM, KNN, métodos de *ensembles* y redes neuronales.

Por otro lado, el aprendizaje no supervisado se centra en descubrir patrones y estructuras ocultas en los datos sin la necesidad de etiquetas. Algunos de los algoritmos utilizados en este enfoque son el Clustering, la Reducción de Dimensionalidad y las Redes Generativas Adversariales.



En lo que respecta al campo de la medicina, el ML se ha convertido en una herramienta poderosa permitiendo a los profesionales de la salud analizar grandes cantidades de datos de pacientes y extraer conocimientos valiosos para la toma de decisiones clínicas, siendo los modelos supervisados ampliamente utilizados tanto en problemas de regresión como de clasificación.

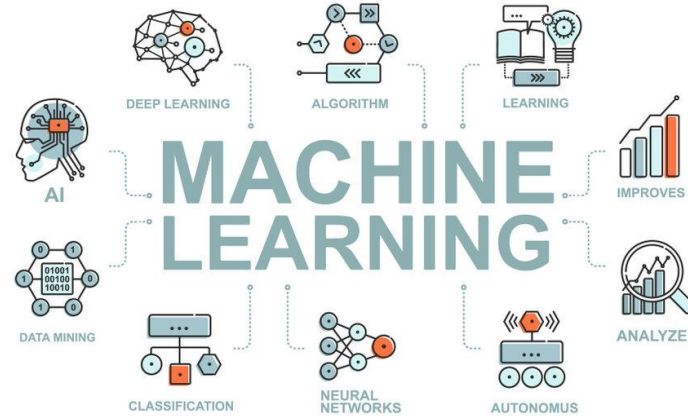


Figura 3. Ramas del *Machine Learning* ([fsm.ac.in](http://sm.ac.in))

Los algoritmos de aprendizaje supervisado que se detallan a continuación son aquellos que se han aplicado sobre el conjunto de datos y que resuelven el problema planteado en este TFG, en aproximaciones tanto de regresión como de clasificación.

2.2.2. Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte (SVM) son algoritmos de aprendizaje supervisado no paramétrico que se utilizan para problemas de clasificación (SVC) lineal y no lineal mediante el uso de funciones de kernel (Figura 4), y regresión (SVR) (Figura 5) o incluso para la detección de valores atípicos [27]. Los SVM pueden aplicarse en diversas áreas médicas, como el diagnóstico de enfermedades, la predicción de resultados clínicos o la identificación de biomarcadores [28].

Durante el entrenamiento, el algoritmo representa los puntos de entrenamiento en el espacio y aprende la importancia de cada uno de ellos para posteriormente representar el límite de decisión (hiperplano). Para definir el hiperplano es necesario un subconjunto de puntos de entrenamiento que se encuentran en la frontera y que se denominan vectores de soporte. Para realizar la predicción de una nueva muestra, se mide la distancia a cada uno de los vectores de soporte [26].

Entre las ventajas de los SVM en medicina se encuentran: buen rendimiento en conjuntos de datos de tamaño pequeño a mediano, capacidad para manejar datos no lineales mediante el uso de kernels y capacidad para generalizar a partir de un conjunto de entrenamiento limitado. Sin embargo, es importante tener en cuenta algunas desventajas. Los SVM pueden ser computacionalmente costosos en conjuntos de datos grandes, ya que el proceso de entrenamiento implica resolver un problema de optimización complejo. Además, los resultados

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

pueden ser sensibles a la selección de hiperparámetros, lo que requiere una cuidadosa configuración y validación cruzada.

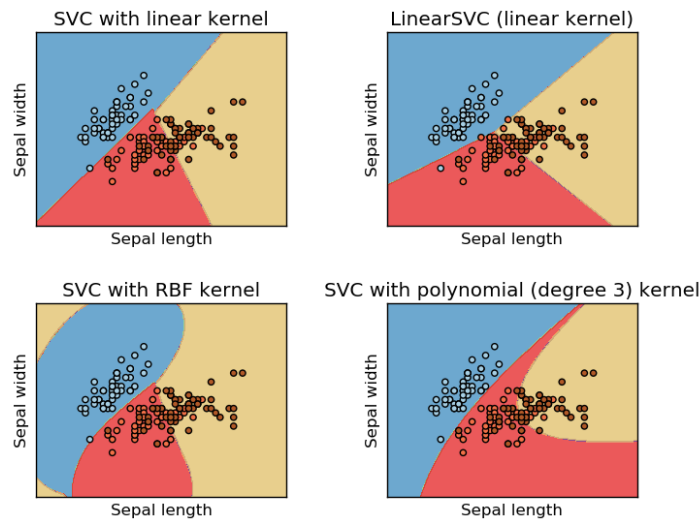


Figura 4. Ejemplos de SVC con diferentes kernels (medium.com)

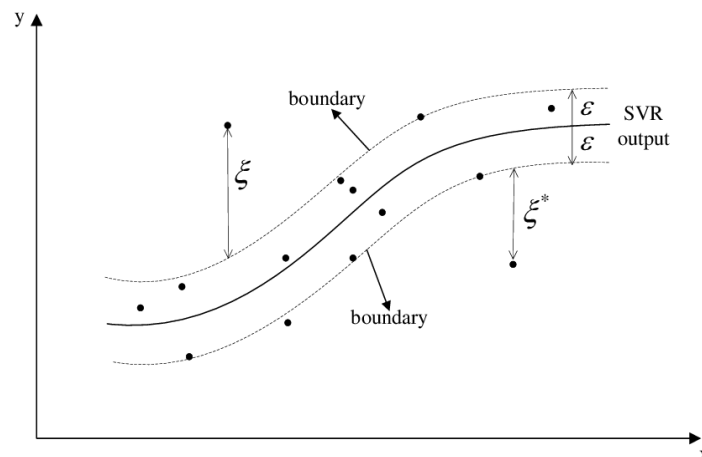


Figura 5. Ejemplo de SVR (researchgate.net)

2.2.3. Métodos de *ensembles*: *Boosting* (*AdaBoost*) y *Voting*

Los métodos de *ensembles*, como el *AdaBoost* y el *Voting*, se basan en la combinación de múltiples modelos base de regresión o clasificación débiles para obtener un solo modelo de predicción más preciso y generalizable [29]. La combinación reduce los errores y mejora significativamente la varianza y el sesgo del modelo de predicción.

Técnicas de *ensembles* como el *Voting* (Figura 7) utilizan la salida de los modelos base para obtener una clasificación, se elige la mayor votación, ya sea votación simple (para problemas de clasificación) o suma/votación ponderada (para problemas de regresión).



Dentro de las técnicas de *ensembles* que utilizan técnicas de fusión de clasificadores destaca el método *Boosting* (el impulso). Con *boosting* se presenta un ajuste secuencial y cada nuevo modelo emplea información del modelo anterior para aprender de sus errores mejorando con cada iteración. *Boosting* tiene variantes o diferentes algoritmos de potenciación o refuerzo, entre ellos la potenciación adaptativa o *AdaBoost* (Figura 6).

Entre las ventajas de los *ensembles* en medicina, se encuentra su capacidad para manejar conjuntos de datos complejos y de alta dimensionalidad, su resistencia al sobreajuste y su capacidad para capturar relaciones no lineales en los datos. Estas técnicas también pueden ser útiles para identificar interacciones entre variables y capturar patrones sutiles en los datos médicos.

Sin embargo, es importante tener en cuenta algunas desventajas. Los *ensembles* pueden ser más difíciles de interpretar en comparación con modelos individuales, lo que puede limitar la comprensión de los factores subyacentes que contribuyen a las predicciones. Además, los *ensembles* requieren ajuste de hiperparámetros adecuado para optimizar su rendimiento, lo que puede ser un desafío.

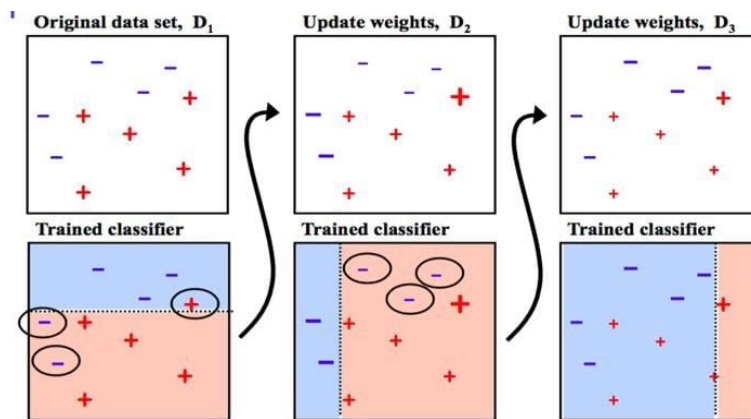


Figura 6. Ejemplo de *AdaBoost* en un problema de clasificación (pub.towardsai.net)

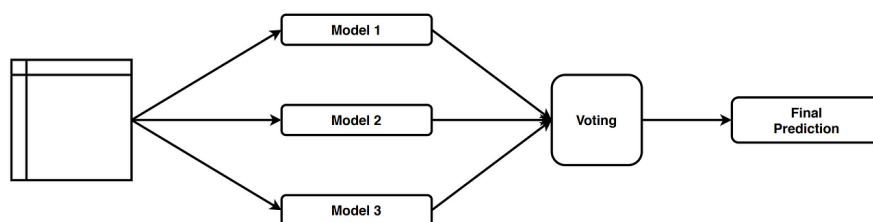


Figura 7. Ejemplo de arquitectura *Voting* en un problema de clasificación/regresión (levelup.gitconnected.com)

2.2.4. Redes neuronales: *Multilayer Perceptron*

Las redes neuronales, como por ejemplo MLP, utilizan formas de procesamiento análogas a las del cerebro humano como base para desarrollar algoritmos capaces de hallar relaciones complejas no lineales entre variables de entrada y salida (Figura 8), siendo ésta una de las

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

principales ventajas de las redes neuronales en medicina [30]. Además, las redes neuronales tienen un buen rendimiento en conjuntos de datos grandes y pueden manejar datos de alta dimensionalidad, como imágenes médicas o datos genómicos.

Otra gran ventaja es la facilidad para escalar su capacidad predictiva en función de la cantidad de información que recibe, permitiendo mayor precisión simplemente añadiendo más datos de entrada. Así mismo, eliminan la necesidad de tratar extensamente las variables de entrada del modelo, no siendo necesario realizar un gran análisis exploratorio para determinar qué variables son más relevantes y cuáles se deben excluir.

Sin embargo, también existen algunas desventajas asociadas con el uso de redes neuronales en medicina. En primer lugar, las redes neuronales requieren grandes cantidades de datos de entrenamiento para obtener resultados óptimos. Esto puede ser un desafío en el campo médico, donde los conjuntos de datos pueden ser limitados debido a la disponibilidad de muestras o a la confidencialidad de los datos. Además, las redes neuronales pueden ser computacionalmente costosas, especialmente cuando se trabaja con conjuntos de datos grandes y modelos complejos. También pueden ser difíciles de interpretar y explicar, lo que puede limitar su adopción en entornos médicos donde la interpretabilidad es esencial.

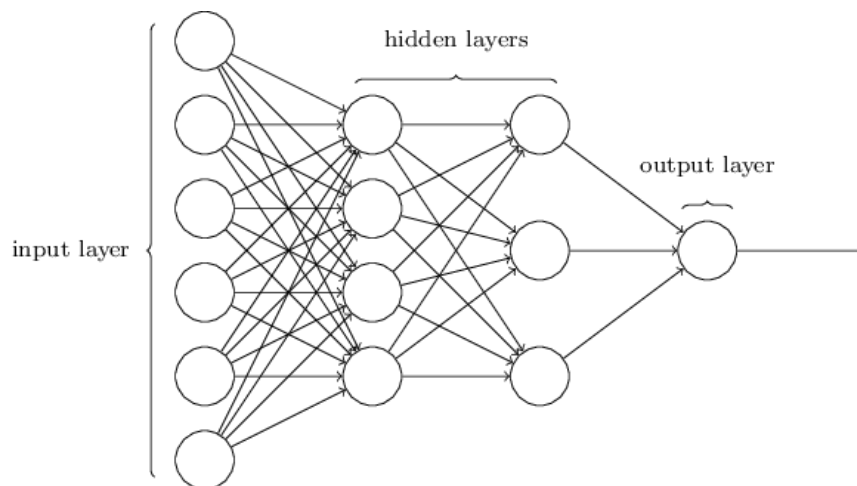


Figura 8. Ejemplo de arquitectura *Multilayer Perceptron* (github.com)

2.2.5. Árboles de decisión

Los árboles de decisión son algoritmos de aprendizaje supervisado no paramétrico que se utiliza para problemas de clasificación y regresión. Estos algoritmos se basan en segmentar poblaciones de forma eficiente y donde cada subgrupo que se genera tiene un significado propio [31]. Los subgrupos formados son bien definidos y mutuamente excluyentes, cuyos miembros presentan características similares. Los árboles de decisión tienen como objetivo crear un modelo capaz de predecir el valor de una variable objetivo mediante el aprendizaje de reglas de decisión simples inferidas de las características de los datos. Estas reglas de decisión son preguntas que conducen a una decisión, “sí” o “no”, pudiendo expresarse en forma de árbol. Cada nodo representa una pregunta o un nodo terminal que contiene la respuesta (Figura 9).

Una de las principales ventajas de los árboles de decisión en medicina es su facilidad de interpretación y comprensión. Estos modelos generan estructuras de árbol que muestran de



manera explícita cómo se toman las decisiones basadas en las características de los datos. Esto permite a los médicos y profesionales de la salud comprender y explicar el proceso de toma de decisiones del modelo, lo que es crucial para la confianza y aceptación en el campo médico.

Además, los árboles de decisión manejan naturalmente variables categóricas y pueden capturar interacciones no lineales entre las características. También son menos sensibles a los valores atípicos en comparación con otros modelos y pueden funcionar bien incluso con conjuntos de datos pequeños.

En cambio, los árboles de decisión también presentan algunas desventajas, ya que no siempre capturan relaciones complejas en los datos, especialmente cuando las relaciones son no lineales o cuando se requiere un alto grado de precisión. Otro aspecto a considerar es la necesidad de ajustar correctamente los parámetros del árbol, como la profundidad máxima o el número mínimo de muestras requeridas en un nodo para realizar una división. Ello puede hacerles propensos al sobreajuste, lo que significa que pueden ajustarse demasiado a los datos de entrenamiento y no generalizar bien a nuevos datos.

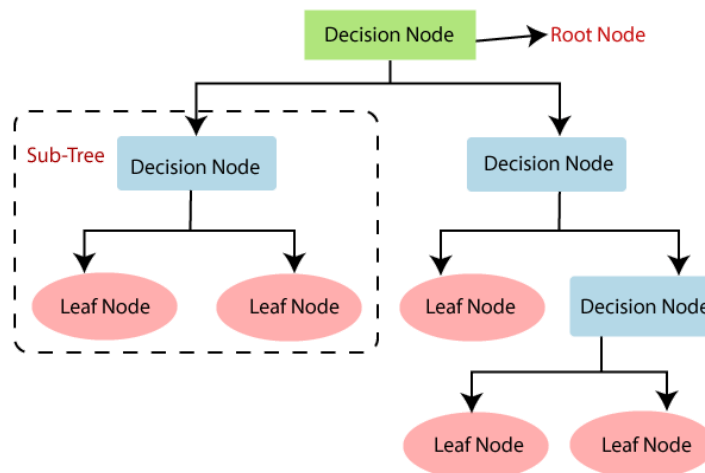


Figura 9. Ejemplo de arquitectura de un árbol de decisión (anderfernandez.com)

2.2.6. Algoritmo de los K Vecinos más Cercanos

El algoritmo de K Vecinos más Cercanos (KNN) es un algoritmo de aprendizaje supervisado no paramétrico ampliamente utilizado en el campo de la medicina debido a su simplicidad y capacidad para manejar problemas de clasificación y regresión. KNN se basa en el principio de que los puntos de datos similares tienden a agruparse en el espacio de características. Estos algoritmos “memorizan” los datos de entrenamiento y, a la hora de realizar una predicción para un nuevo dato, sencillamente devuelve el dato de entrenamiento más próximo (o el resultado de considerar los k datos de entrenamiento más próximos) (Figura 10).

En medicina, KNN se aplica en diversos escenarios. En problemas de regresión, KNN utiliza la información de los K vecinos más cercanos para predecir el valor de una variable continua. Calcula el promedio de los valores de los vecinos más cercanos y lo utiliza como estimación. Por ejemplo, en la predicción de la presión arterial de un paciente, KNN puede utilizar la información de pacientes similares para estimar su presión arterial.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

En problemas de clasificación, KNN asigna una muestra a la clase más frecuente entre sus vecinos más cercanos. Esto es especialmente útil en medicina para el diagnóstico de enfermedades o la predicción de resultados clínicos. Por ejemplo, en la detección de cáncer, KNN puede utilizar información de pacientes previos para clasificar a un nuevo paciente como positivo o negativo para la enfermedad.

Una de las principales ventajas de KNN es su fácil implementación y comprensión. No requiere una fase de entrenamiento explícita y no hace suposiciones sobre la distribución de los datos. Además, KNN tiene la capacidad de manejar relaciones no lineales y es adecuado para conjuntos de datos pequeños o con características complejas.

Sin embargo, KNN también tiene algunas limitaciones. Es sensible a la elección del parámetro K , que representa el número de vecinos considerados. Un valor demasiado pequeño puede llevar a una clasificación inestable y sensible al ruido, mientras que un valor demasiado grande puede generar un sesgo hacia las clases mayoritarias. Aparte, en conjuntos de datos grandes, el cálculo de la distancia entre los puntos puede ser computacionalmente costoso.

Otro inconveniente de KNN es que no maneja bien las variables irrelevantes o el ruido en los datos. La presencia de características irrelevantes puede afectar negativamente la precisión del algoritmo y generar predicciones incorrectas.

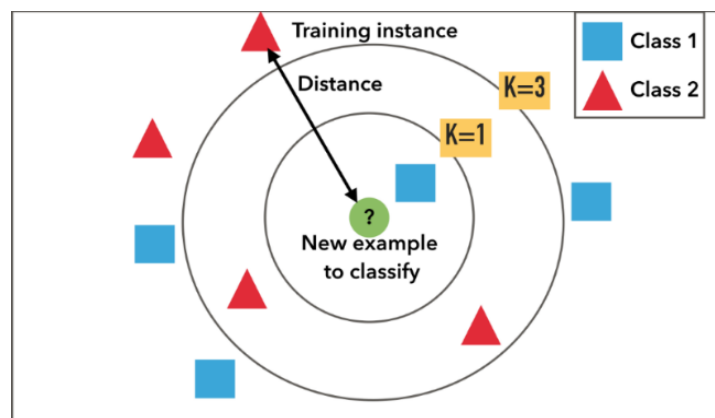


Figura 10. Ejemplo de KNN en un problema de clasificación (deepai.org)

2.2.7. Regresión: lineal y logística

La regresión lineal es el método más sencillo entre los modelos lineales. Este modelo estima la variable objetivo mediante una combinación lineal de las características predictivas, minimizando la suma de los cuadrados de los residuales (método de los mínimos cuadrados).

En problemas de regresión, la regresión lineal es uno de los métodos más comunes utilizados en medicina. Busca establecer una relación lineal entre las variables independientes y la variable dependiente. Esto permite estimar y predecir el valor de la variable dependiente en función de los valores de las variables independientes. En medicina, la regresión se aplica en diversos escenarios, como la predicción de resultados clínicos, la estimación de la progresión de enfermedades o la identificación de factores de riesgo.

Una de las principales ventajas de la regresión es su fácil interpretación. Permite analizar la contribución individual de cada variable independiente y evaluar su importancia en la predicción del resultado. Además, la regresión es computacionalmente eficiente, especialmente en comparación con algoritmos más complejos como las redes neuronales.

Sin embargo, la regresión también tiene algunas limitaciones. Supone una relación lineal entre las variables, lo cual puede ser una simplificación excesiva en situaciones donde existen relaciones no lineales o complejas. Además, asume que los residuos (diferencias entre los valores observados y los valores predichos) siguen una distribución normal. Si esta suposición no se cumple, puede afectar la validez de los resultados y las inferencias.

Otra desventaja de la regresión es su capacidad limitada para manejar variables categóricas o no lineales de forma directa. En estos casos, pueden requerirse transformaciones de variables o el uso de técnicas adicionales (Figura 11), como la regresión logística [32] o modelos no lineales.

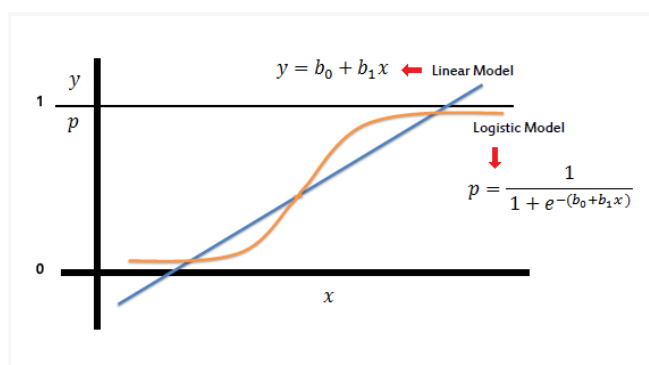


Figura 11. Ejemplo de modelo de regresión lineal y logística (themachinelearners.com)

2.3. Crítica al estado del arte

2.3.1. Crítica a la metodología utilizada en el *score* SAME-TT₂R₂

La metodología utilizada en el desarrollo del *score* SAME-TT₂R₂ para la predicción del control de la anticoagulación oral en pacientes con FA tratados con warfarina presenta algunas limitaciones que deben ser consideradas, entre ellas la utilización de la regresión logística binaria como único modelo predictivo.

Otra crítica importante es la falta de validación externa del *score* en algunos centros. Es necesario evaluar su desempeño en poblaciones y entornos clínicos diferentes para determinar su generalización. La falta de su validación en algunos de los estudios publicados posteriormente al desarrollo del *score* SAME-TT₂R₂ [6, 22 - 24, 33] limita la confiabilidad de los resultados y la aplicabilidad de dicho *score* en la práctica clínica diaria.

Además, la selección de variables incluidas en el *score* puede plantear interrogantes. La etnia parece estar sobre-representada en este modelo, sin embargo estará infra-representada cuando el modelo sea aplicado a población española, como es la población del presente TFG. En el *score* SAME-TT₂R₂ solo se ha incluido tratamiento con antiarrítmicos, antihipertensivos y estrógenos, pero hay multitud de fármacos que interaccionan con el acenocumarol. También existe la

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

posibilidad de que se hayan omitido otros factores importantes que podrían influir en el control de la anticoagulación, tales como distancia al centro médico, nivel de educación, farmacogenética, alcohol, etc. La exclusión de ciertas variables puede afectar la precisión y la capacidad predictiva del *score*.

Otra crítica se relaciona con la elección de los puntos de corte utilizados para clasificar a los pacientes en categorías de riesgo. Según trabajos previos, la determinación de los puntos de corte del sistema de puntuación SAME-TT₂R₂, como riesgo de TRT subóptimo, ha variado entre ≥ 1 [34], ≥ 2 [35] y ≥ 3 en otros estudios [21, 33]. Ello implica un sesgo en la asignación de los pacientes a diferentes categorías de riesgo. Además, la sensibilidad y especificidad del *score* podrían variar según los puntos de corte utilizados, lo que puede afectar su utilidad clínica en la toma de decisiones.

Adicionalmente, es importante considerar que el *score* SAME-TT₂R₂ se desarrolló específicamente para pacientes tratados con warfarina, y su aplicabilidad en pacientes que reciben otros AVK, como acenocumarol, puede ser limitada. En este sentido, la efectividad de este *score* en la predicción del control de la anticoagulación en pacientes anticoagulados con acenocumarol ha sido escasamente evaluada [24].

2.3.2. Crítica a los modelos predictivos en salud

Muchos estudios se centran en cohortes de pacientes específicos, lo que limita la extrapolación de los resultados a otros entornos. Para superar esta limitación, es necesario llevar a cabo investigaciones más amplias y diversificadas que incluyan datos de diferentes grupos de pacientes y entornos clínicos (estudios multicéntricos), así como validaciones externas de los modelos predictivos.

Además, la calidad y disponibilidad de los datos utilizados en la construcción de los modelos también son objeto de críticas. La falta de estandarización en la recopilación y el registro de datos clínicos dificulta la comparación y la combinación de conjuntos de datos entre diferentes estudios. Asimismo, la presencia de datos incompletos o de baja calidad puede afectar la precisión y confiabilidad de los modelos predictivos. Por lo tanto, es crucial establecer estándares claros en la captura y registro de datos clínicos para mejorar la calidad de los modelos y garantizar resultados más confiables.

Otra crítica se centra en la interpretación y aplicabilidad de algunos modelos predictivos en el campo de la medicina. Algunos modelos, como las redes neuronales o los *ensembles*, son considerados cajas negras debido a su complejidad y falta de transparencia en la toma de decisiones. Esto plantea preocupaciones en cuanto a la confianza y aceptación de los resultados por parte de los médicos y profesionales de la salud. Sin embargo, estos modelos son capaces de capturar relaciones más complejas en los datos.

Por último, la investigación en ciencias de la computación y ciencias de datos aplicadas a salud continúa mejorando la precisión de las predicciones clínicas y optimizando la asignación de recursos en la atención médica, pero es posible que incluso un modelo de predicción perfectamente ajustado no se traduzca en una mejor atención clínica ni pueda eliminar la incertidumbre de la toma de decisiones en medicina. También es indiscutible que, en la salud, los expertos del dominio son un eslabón fundamental e irremplazable en el modelo de trabajo de



ciencia de datos, que le dan sentido y modulan tanto la creación como el uso clínico de estos algoritmos.

2.4. Propuesta

En el contexto de la FANV, actualmente hay alternativas terapéuticas a los AVK, concretamente los ACOD, sin necesidad de ser monitorizadas de forma periódica. Sin embargo, por cuestiones socioeconómicas el acenocumarol se sigue usando ampliamente en el tratamiento de los pacientes con FANV a pesar del elevado porcentaje de pacientes con TRT subóptimo. Por ello, el presente TFG propone identificar el subgrupo de pacientes con TRT subóptimo que se beneficiarían de un uso precoz de los ACOD de forma coste-efectiva, además de ser ésta una de sus indicaciones financiada por la AEMPS. Para ello, se han incluido un amplio rango de variables no incluidas en el clásico *score* SAME-TT₂R₂ (todas ellas especificadas en el apartado 4.3), así como se han investigado diferentes modelos de aprendizaje automático como alternativa a las regresiones tan comúnmente utilizadas en medicina y concretamente en dicho *score*.

3. Análisis del problema

3.1. Introducción del problema

La FA implica una carga importante tanto para los pacientes como para la salud y la economía de la sociedad, tal y como se ha especificado previamente (apartado 2.1).

Además, en el departamento de salud del CHGUV, al igual que sucede en el resto de España y en otros países, existe un número elevado de pacientes anticoagulados con AVK por FANV con mal control (alrededor de un 50%), siendo los ACODs una alternativa terapéutica, ya financiada en España, de la se podrían beneficiar este subgrupo de pacientes con TRT subóptimo. Este cambio a un ACOD, habiéndose seleccionado el más adecuado según el perfil del paciente, supondría una reducción de las complicaciones tromboembólicas y hemorrágicas, así como una mejor calidad de vida de los pacientes y su entorno familiar.

Sin embargo en la práctica clínica de los profesionales de la salud, la decisión terapéutica para un paciente en particular no solo tiene en cuenta el resultado final, sino también debe sopesar los riesgos y los beneficios de un procedimiento terapéutico concreto. En este sentido, la utilización de los ACODs se asocia con beneficios (no requieren control rutinario y presentan menor número de complicaciones hemorrágicas y tromboembólicas) pero también inconvenientes (mayor coste) respecto al uso de AVK, por lo que han sido ya materia de amplia discusión. Por ello es necesario establecer unas recomendaciones de uso para los ACODs, identificando aquellos pacientes en los que las ventajas de estos nuevos tratamientos sean mayores y asegurando que el incremento del número de pacientes tratados se produzca de forma prudente y acompañada con el conocimiento científico, tal y como podría ser el caso de los pacientes con un TRT subóptimo.

3.2. Análisis de seguridad

Existe el riesgo de que personas no autorizadas puedan acceder a los datos del proyecto, lo que podría comprometer la privacidad y confidencialidad de la información médica sensible de los pacientes. Para mitigar este riesgo solo las personas con una necesidad relevante (el tutor en el CHGUV, el tutor en la UPV y el alumno del presente trabajo) han podido ver y trabajar los datos recogidos, tras aprobación y firma del acuerdo de colaboración entre la UPV y CHGUV por parte de la Directora Gerente del CHGUV. Así mismo la información se ha almacenado únicamente durante la realización del presente TFG.

3.3. Análisis energético o de eficiencia algorítmica

Es importante evaluar la eficiencia de los algoritmos utilizados para el procesamiento de datos en términos de tiempo de ejecución y consumo de recursos computacionales. Algoritmos más eficientes pueden reducir la carga de trabajo del sistema, disminuir el tiempo de respuesta y optimizar el consumo energético. Comparar diferentes modelos en función de su rendimiento y eficiencia puede ayudar a identificar aquellos que logren un equilibrio óptimo entre precisión y consumo de energía. Para ello, se evaluaron los modelos mediante las métricas consideradas, así como su tiempo de ejecución (datos mostrados en apartado 5).

3.4. Análisis del marco legal y ético

En aras de salvaguardar la privacidad y garantizar la anonimización de los datos utilizados, se llevó a cabo un proceso de eliminación de variables identificativas de los pacientes, como el número de historia clínica (NHC), fecha de nacimiento y los nombres individuales, asegurando así el cumplimiento de las normativas y principios éticos en relación con la privacidad y confidencialidad de la información.

3.4.1. Análisis de la protección de datos

Para la realización de este proyecto se ha seguido la LOPD (Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales) como normativa a seguir en lo relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos, y completar sus disposiciones. También se ha tenido en cuenta el reglamento europeo RGPD (Reglamento General de Protección de Datos) 2016/679, de 27 de abril de 2016.

Esto implica garantizar la privacidad de los pacientes y proteger la confidencialidad de su información personal y médica, así como cumplir con los requisitos de almacenamiento seguro, uso apropiado, anonimización y eliminación adecuada de los datos cuando ya no sean necesarios. En este sentido, el presente TFG es un estudio observacional retrospectivo en el que se han utilizado datos clínicos, analíticos, farmacológicos y ecocardiográficos que habitualmente constan en la historia clínica electrónica (HCE) de los pacientes, tomando las medidas adecuadas para proteger la identidad de los mismos y asegurar que los datos se han manejado de manera segura y se han utilizado únicamente con fines específicos para este proyecto.

La protección de datos, mencionada anteriormente, puede implicar la implementación de acuerdos de confidencialidad. En este sentido Leandro Pérez Orero, como autor del presente



TFG, ha firmado un compromiso de confidencialidad depositado en la dirección de investigación y docencia del CHGUV.

3.4.2. Propiedad intelectual

Destacar que la propiedad intelectual será del autor del TFG, Leandro Pérez Orero, que será quien pondrá las restricciones para su distribución y modificación por parte de terceros, teniendo siempre en cuenta la protección de datos, confidencialidad y la transparencia en la divulgación de los resultados del presente proyecto.

3.4.3. Ética

Aunque la búsqueda de un modelo predictivo para detectar pacientes con un TRT subóptimo conllevaría a una introducción precoz de ACOD de lo que se puede beneficiar la industria farmacéutica que los comercializa, queremos aclarar que el autor de TFG no presenta en este sentido ningún conflicto de interés, pues no ha recibido ninguna ayuda por parte de la industria farmacéutica y solo ha buscado un beneficio potencial para la salud de los pacientes.

Por otro lado, este estudio se ha llevado a cabo de acuerdo con los requerimientos expresados en la Declaración de Helsinki revisada en 2013, así como en la legislación vigente en España de acuerdo a lo dispuesto en el Real Decreto 957/2020 por el que se regulan los estudios observacionales con medicamentos de uso humano.

3.5. Análisis de riesgos

El presente estudio no representó un riesgo adicional para los pacientes, ya que al ser un estudio observacional no se realizó ninguna intervención adicional al proceso diagnóstico/terapéutico habitual, por lo que no aumentó el número de exploraciones complementarias necesarias ni implicó un cambio terapéutico diferente a lo establecido en la práctica habitual. Solo se realizó revisión de la HCE de los pacientes preservando la privacidad de los participantes, tal y como he explicado anteriormente.

Además, el proyecto puede enfrentar varios riesgos relacionados con la calidad de los datos, el sesgo, la generalización limitada, las decisiones clínicas basadas en la predicción y la aceptación/adopción del modelo.

La calidad y precisión de los datos son fundamentales, por lo que se deben realizar controles de calidad y validación para garantizar la confiabilidad de los resultados y, así, evitar errores o datos incompletos.

Además, se debe abordar el sesgo en los datos para evitar conclusiones erróneas o discriminación en la predicción como la falta de diversidad en los pacientes incluidos o la carencia de representatividad de la muestra. Por ello, debe considerarse la generalización del modelo a otras poblaciones y entornos clínicos, realizando validaciones y estudios adicionales.

Es crucial destacar que el modelo debe ser utilizado como una herramienta complementaria a la evaluación clínica, evitando decisiones basadas únicamente en la predicción y considerando otros factores como el conocimiento de los expertos en el área de salud.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

Por último, se debe trabajar en la comunicación y educación para garantizar la aceptación y adopción del modelo por parte de los profesionales de la salud y los pacientes, informándoles sobre los beneficios y limitaciones del mismo.

3.6. Identificación y análisis de soluciones posibles

3.6.1. Detección precoz de un TRT subóptimo

Para enfrentar este desafío desde un enfoque basado en ciencia de datos, se podrían aplicar las siguientes estrategias:

- Utilizar algoritmos de aprendizaje automático para desarrollar modelos de detección de anomalías. Estos modelos podrían analizar los datos de los pacientes e identificar patrones o comportamientos atípicos que indiquen una desviación del TRT.
- Implementación de un sistema de monitoreo continuo en tiempo real que permitiera un estricto y regular seguimiento de los niveles de anticoagulación de los pacientes. Esto podría realizarse mediante la integración de dispositivos de monitoreo como medidores de coagulación portátiles o aplicaciones móviles, utilizando algoritmos de procesamiento de señales para detectar cambios significativos en los niveles de anticoagulación. Ello permitiría a los pacientes y a los profesionales de la salud acceder a los datos de manera rápida y precisa.

Este último enfoque permite una detección temprana de un TRT subóptimo y por tanto una intervención precoz para optimizar el tratamiento anticoagulante de elección en cada caso, minimizando así el riesgo de complicaciones hemorrágicas y tromboembólicas.

Sin embargo, requiere de un sistema de monitoreo continuo y análisis en tiempo real, lo que puede ser costoso y necesita de recursos adicionales. Además, es importante garantizar la precisión y confiabilidad de los dispositivos de monitoreo utilizados.

3.6.2. Predicción del TRT

Para afrontar esta situación desde una perspectiva de análisis de datos, se podrían utilizar algoritmos de aprendizaje automático, como redes neuronales o árboles de decisión, para desarrollar modelos de predicción que permitieran anticipar cuándo un paciente estará fuera del rango terapéutico. Estos modelos podrían basarse en características sociodemográficas, clínicas, analíticas, farmacológicas, ecocardiográficas y otros datos adicionales como características genéticas, biomarcadores específicos o datos ambientales que puedan tener influencia en la respuesta al tratamiento anticoagulante.

Esta solución permite una intervención preventiva para evitar que los pacientes se desvíen del rango terapéutico. También, proporciona una herramienta útil para la toma de decisiones clínicas y la optimización del tratamiento anticoagulante.

No obstante, requiere datos adicionales y el desarrollo de modelos de predicción complejos. La precisión de las predicciones puede depender de la calidad y disponibilidad de los datos utilizados. Además, es necesario garantizar que las predicciones se interpreten y apliquen de manera adecuada en el contexto clínico.



3.7. Solución propuesta

La solución propuesta en el presente TFG se enfoca en desarrollar un modelo de predicción que pueda anticipar la detección de los pacientes antes de que se desvíen del rango terapéutico establecido. A continuación, se detallan las fases clave para implementar esta solución:

1. Recopilación de datos: Recopilar datos relevantes de los pacientes, como características clínicas, datos de monitoreo, resultados de análisis de laboratorio, tratamientos previos y variables adicionales que puedan influir en la respuesta al tratamiento anticoagulante.
2. Transformación/Integración/Preparación de los datos: Transformar e integrar los datos recopilados en un formato adecuado para el análisis. Realizar tareas de limpieza de datos, manejo de valores faltantes y eliminación de variables irrelevantes. Además, aplicar técnicas de ingeniería de características para crear nuevas variables que capturen información relevante.
3. Evaluación de la capacidad discriminativa del *score* SAME-TT₂R₂: Analizar la capacidad discriminativa del *score* SAME-TT₂R₂ en los datos recopilados. Esto implica evaluar cómo el *score* SAME-TT₂R₂ se relaciona con la desviación del rango terapéutico y determinar si es suficientemente discriminativo para predecir qué pacientes van a tener un TRT subóptimo.
4. Desarrollo del modelo de predicción: Utilizar técnicas de aprendizaje automático, como redes neuronales, árboles de decisión o algoritmos de clasificación para desarrollar un modelo de predicción que utilice las variables relevantes para detectar a los pacientes antes de que se desvíen del rango terapéutico. Los modelos se deben entrenar utilizando datos históricos y se optimizarán para obtener la mejor capacidad predictiva posible mediante la división en conjuntos de entrenamiento y prueba.
5. Comparación de modelos: comparar el rendimiento de los modelos desarrollados mediante las métricas oportunas en cuanto a clasificación y para regresión.
6. Validación del modelo: realizar una validación interna del modelo final con un conjunto independiente del usado para entrenar y testear los modelos, con la finalidad de evaluar la capacidad del modelo propuesto para generalizar y predecir con precisión el diagnóstico de nuevos pacientes.
7. Comparación del modelo con el SAME-TT₂R₂: Realizar una comparación entre el modelo de predicción propuesto y el *score* SAME-TT₂R₂ para determinar cuál es más efectivo en la predicción del TRT.

3.8. Plan de trabajo

El estudio se ha realizado durante un periodo de 5 meses, iniciándose tras la autorización de la Directora Gerente del CHGUV, según la distribución representada en el siguiente diagrama de Gantt (Figura 12).

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

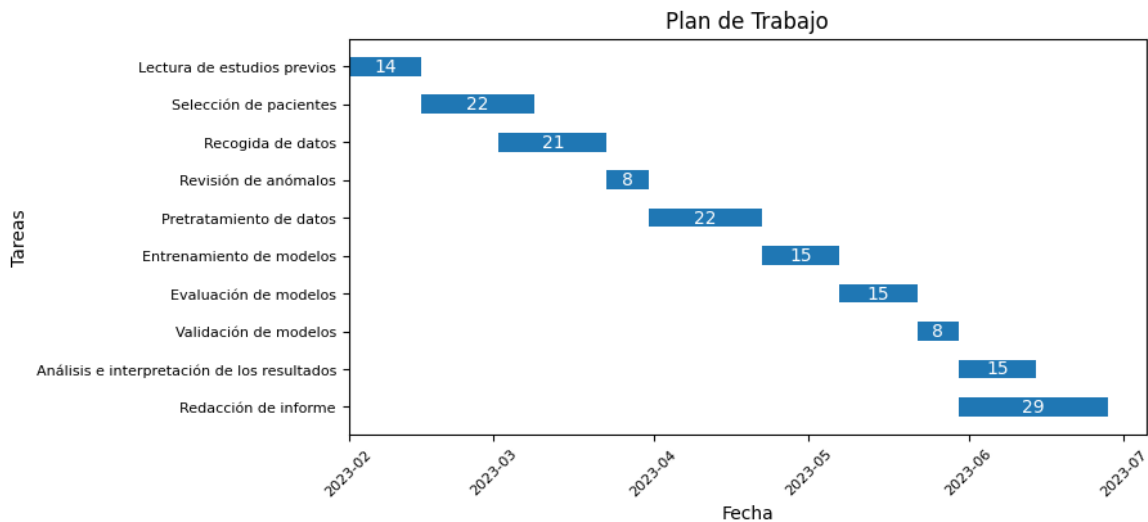


Figura 12. Diagrama de Gantt sobre el plan de trabajo (en días)

1. Lectura de estudios previos (14 días - fecha de inicio: 01/02/2023): revisión y análisis de investigaciones anteriores relacionadas con el tema, para la obtención de conocimientos necesarios y suficiente evidencia científica.
2. Selección de pacientes (22 días - fecha de inicio: 15/02/2023): clasificación de paciente a partir del sistema informático GOTA, y posterior aplicación de los criterios de inclusión y exclusión.
3. Recogida de datos (21 días - fecha de inicio: 02/03/2023): obtención de datos sociodemográficos, clínicos, analíticos, farmacológicos y ecocardiográficos a partir de la revisión de la HCE de los pacientes e introducción en la base de datos *Excel*.
4. Revisión de los datos anómalos (8 días - fecha de inicio: 23/03/2023): corrección de los datos erróneos o incompletos.
5. Pretratamiento de datos (22 días - fecha de inicio: 31/03/2023): preprocesamiento de los datos, incluyendo tratamiento de faltantes, valores anómalos, análisis descriptivo, pruebas de significancia, transformaciones de variables y selección de variables.
6. Entrenamiento de modelos (15 días - fecha de inicio: 22/04/2023): selección y entrenamiento de los modelos predictivos.
7. Evaluación de los modelos (15 días - fecha de inicio: 07/05/2023): prueba de los modelos con los datos de test y optimización de parámetros.
8. Validación de modelos (8 días - fecha de inicio: 22/05/2023): validación de los modelos finales mediante el conjunto de validación interna..
9. Análisis e interpretación de los resultados (15 días - fecha de inicio: 30/05/2023): comparación y comprensión de los hallazgos obtenidos, identificación de patrones y tendencias, evaluación de la precisión y confiabilidad del modelo.
10. Redacción de informe (29 días - fecha de inicio: 30/05/2023): introducción; estado del arte; análisis del problema; documentación de los resultados obtenidos incluyendo el



análisis exploratorio de datos, las metodologías utilizadas, los modelos desarrollados y sus respectivas evaluaciones; conclusiones y recomendaciones futuras basadas en los hallazgos del estudio.

4. Preparación y Comprensión de los Datos

4.1. Población

De los 9209 pacientes con diagnóstico de FA anticoagulados con acenocumarol en tratamiento activo entre 1 de enero de 2007 al 31 de diciembre de 2021 (15 años) registrados en el programa informático GOTA de control del tratamiento anticoagulante en el CHGUV, se planteó estudiar a 1000 pacientes para el desarrollo de los modelos predictivos. Entre los criterios de inclusión se exigió el disponer de un seguimiento con INRs durante al menos 7 meses desde el inicio del tratamiento anticoagulante y que la ecocardiografía no mostrara valvulopatía (ni mitral, ni aórtica, ni tricúspide) o en caso de valvulopatía, ésta fuera ligera. Todos estos casos (sin valvulopatía o con valvulopatía ligera) fueron considerados como FANV en el presente estudio.

Para obtener los 1000 pacientes, se tuvieron que revisar un total de 3823 pacientes seleccionados al azar, pues durante dicha selección quedaron excluidos 2823 pacientes por alguno de los siguientes motivos:

- No disponer de informe clínico electrónico completo al diagnóstico (pacientes que iniciaron la anticoagulación anteriormente al 2005).
- Seguimiento con INRs inferior a 7 meses consecutivos (al tratarse de una FA no permanente, fallecimiento o traslado de domicilio).
- Ecocardiografía con presencia de valvulopatía mitral, aórtica y/o tricúspide moderada/grave.
- No disponer de ecocardiografía.

Posteriormente, se realizó el estudio observacional retrospectivo en los 1000 pacientes con FANV seleccionados.

4.2. Variable a predecir: Tiempo en Rango Terapéutico (TRT)

Se definió el TRT como una variable continua expresada como porcentaje de INRs en rango terapéutico ($INR = 2-3$), obteniendo los valores de INRs a partir del sistema informático GOTA. Para dicho cálculo se tuvo en cuenta INRs de los primeros 7 meses descartando INRs del primer mes de tratamiento (ajuste inicial de dosis) o periodos como intervenciones quirúrgicas o dentales u otros procedimientos invasivos, etc. que conlleven la modificación de la pauta de acenocumarol. Se consideró un resultado más óptimo cuanto mayor fuese el valor del TRT.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

La media del número de INRs utilizados para el cálculo del TRT fue de 8.33 ± 1.63 , rango = 4 - 17. Al igual, la media del TRT calculado mediante porcentaje de INRs en rango fue de $57.03 \pm 19.49\%$, rango = 0 - 100 y su mediana fue de 55.6, con un rango intercuartílico (RIQ) = 44 - 71.4.

La distribución de todos los pacientes según su TRT se muestra en la Figura 13, donde se observa que la mayoría se encuentran entre 40% y 60% de TRT (sesgo = 0.18). El porcentaje de pacientes con TRT subóptimo < 60% fue de 57.3% (573 pacientes en total), con TRT subóptimo < 65% fue de 66% (660 pacientes en total) y con TRT subóptimo < 70% fue de 70.7% (707 pacientes en total).

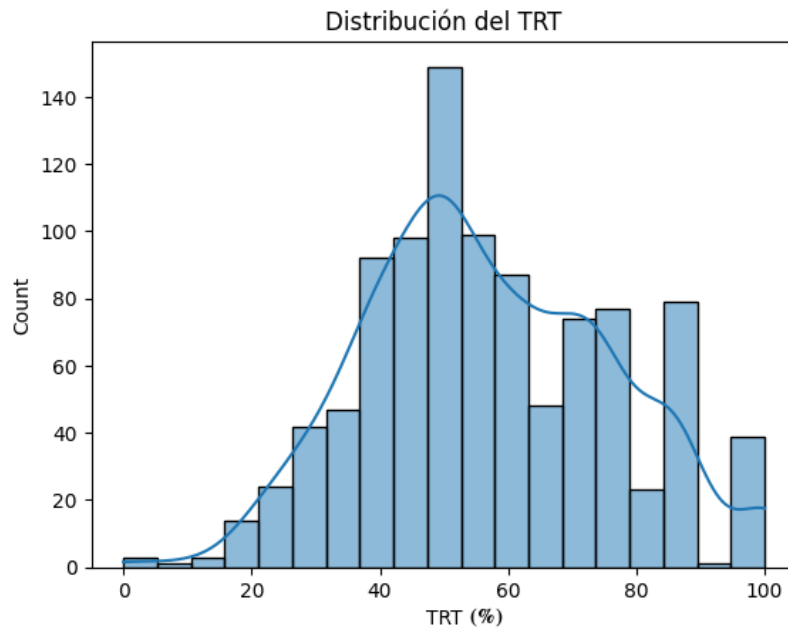


Figura 13. Distribución del Tiempo en Rango Terapéutico (TRT) en el conjunto de datos

4.3. Variables potencialmente predictoras

Se extrajeron los siguientes datos de la HCE: nombre (apellidos, nombre), NHC, fecha de nacimiento, sexo, etnia, IMC, dislipemia, enfermedad renal, enfermedad hepática, enfermedad pulmonar, neoplasia, apnea del sueño, tabaquismo en los dos últimos años, abuso de alcohol, cardiopatía isquémica, hipertensión arterial (HTA), diabetes mellitus (DM), arteriopatía periférica, insuficiencia cardíaca (IC), deterioro cognitivo/demencia, ictus previo, embolia sistémica previa, sangrado mayor previo, fecha de inicio de tratamiento anticoagulante con acenocumarol, tratamiento antiarrítmico con amiodarona, otros antiarrítmicos, verapamilo, diltiazem, digoxina, sotalol, otros β -bloqueantes, aspirina, otros antiagregantes plaquetarios, interacciones del acenocumarol con otros fármacos de uso crónico, número total de pastillas diarias de otros fármacos distintos al acenocumarol, ADE, leucocitos, neutrófilos, linfocitos, monocitos, hemoglobina, plaquetas, disfunción ventricular y tamaño de la aurícula izquierda aumentado por ecocardiografía.

A continuación se presentan todas las variables, potencialmente predictoras, escogidas para realizar el análisis junto con una breve descripción y qué factores se han considerado como positivos para cada una de ellas.



4.3.1. Características identificativas

- **Nombre:** nombre completo del paciente (apellidos, nombre).
- **NHC:** número de historia clínica del paciente.

4.3.2. Características demográficas

- **Fecha de nacimiento:** fecha de nacimiento del paciente (dd/mm/yyyy).
- **Sexo:** variable categórica donde H = hombre y M = mujer.
- **Etnia:** todos los pacientes fueron caucásicos (blancos) incluyendo diferentes nacionalidades, 992 de nacionalidad española, 1 alemana, 2 rusas, 1 italiana, 1 rumana, 2 marroquíes y 1 paciente de origen desconocido. Al tratarse todos de caucásicos, no se tuvo en cuenta esta variable para el desarrollo de los modelos predictivos. Únicamente se utilizó la etnia para el cálculo del score SAME-TT₂R₂ con cero puntos para la variable etnia no caucásica.

4.3.3. Comorbilidades y factores de riesgo

- **IMC (kg/m²):** variable continua que indica la relación entre el peso y la estatura de una persona.
- **Dislipemia:** trastorno metabólico que se caracteriza por niveles anormales de lípidos en la sangre. Es una variable binaria cuyo valor es 1 si el diagnóstico del paciente incluye alguno de los siguientes términos: hipercolesterolemia / hipertrigliceridemia / hiperlipemia, y 0 en caso contrario.
- **Enfermedad renal:** condición en la que los riñones no pueden filtrar adecuadamente los desechos y el exceso de líquidos de la sangre. Es una variable binaria cuyo valor es 1 si el diagnóstico del paciente incluye alguno de los siguientes términos: diálisis / trasplante / creatinina > 2.2 miligramos/decilitros (mg/dL) / aclaramiento creatinina ≤ 30 mililitros/minuto (ml/min) / enfermedad renal, y 0 en caso contrario.
- **Enfermedad hepática:** cualquier cambio en la estructura o función del hígado que puede ser causado por diversas enfermedades o factores de riesgo. Es una variable binaria cuyo valor es 1 si el diagnóstico del paciente incluye alguno de los siguientes términos: cirrosis / hepatopatía crónica / bilirrubina > 2 x normal (0.3-1.2 mg/dL) / transaminasas > 3 x normal (10-34 unidades/litro) / esteatosis hepática, y 0 en caso contrario.
- **Enfermedad pulmonar:** trastornos que afectan la función y la estructura de los pulmones. Es una variable binaria cuyo valor es 1 si el diagnóstico del paciente incluye los siguientes términos: EPOC (enfermedad pulmonar obstructiva crónica) / bronquiectasias / asma / bronquitis crónica / atelectasia / hiperreactividad bronquial, y 0 en caso contrario.
- **Neoplasia:** crecimiento anormal y descontrolado de células que puede ser benigno o maligno. En nuestro caso se han seleccionado las neoplasias malignas activas y casos con antecedentes de neoplasia maligna, sin tener en cuenta las neoplasias cutáneas

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

excepto los casos de melanoma. Para considerar neoplasia se tuvieron en cuenta los siguientes términos: tumor / neoplasia / cáncer / carcinoma / melanoma. La variable se divide en 3 categorías: tener neoplasia activa (± 1 año), haber tenido antecedentes de neoplasia y no tener neoplasia.

- **Apnea del sueño:** trastorno en el que la respiración se interrumpe repetidamente durante el sueño debido a la obstrucción de las vías respiratorias superiores o a la falta de señales del cerebro que controlan la respiración. Es una variable binaria cuyo valor es 1 si el paciente tiene antecedentes de SAOS / SAHS = síndrome de apnea del sueño, y 0 en caso contrario.
- **Tabaquismo:** variable binaria cuyo valor es 1 si el paciente ha fumado en los dos últimos años, y 0 en caso contrario.
- **Abuso de alcohol:** variable binaria cuyo valor es 1 si el paciente tiene antecedentes de alcoholemia, y 0 en caso contrario.
- **Cardiopatía isquémica:** enfermedad cardiovascular que se produce cuando las arterias coronarias que suministran sangre al corazón se estrechan o se bloquean. Es una variable binaria cuyo valor es 1 si el diagnóstico incluye alguno de los siguientes términos: angina / angor / infarto de miocardio (IAM) / enfermedad isquémica aguda/crónica cardíaca / angiografía significativa / estenosis arterias coronarias.
- **Hipertensión arterial (HTA):** condición en la que la presión arterial es elevada de forma sostenida. Es una variable binaria cuyo valor es 1 si el paciente tiene antecedentes de HTA, y 0 en caso contrario.
- **Diabetes mellitus (DM):** enfermedad crónica que se caracteriza por niveles elevados de azúcar en la sangre. Es una variable binaria cuyo valor es 1 si el paciente tiene antecedentes de DM, y 0 en caso contrario.
- **Arteriopatía periférica:** enfermedad en la que las arterias que suministran sangre a las piernas y los pies se estrechan o se bloquean. Es una variable binaria cuyo valor es 1 si el diagnóstico del paciente contiene alguno de los siguientes términos: claudicación intermitente / arteriosclerosis / angiopatía / enfermedad vascular o arterial, y 0 en caso contrario.
- **Insuficiencia cardíaca (IC):** condición en la que el corazón no puede bombear suficiente sangre para satisfacer las necesidades del cuerpo. Es una variable binaria cuyo valor es 1 si el paciente tiene antecedentes de edema agudo de pulmón o insuficiencia cardíaca, y 0 en caso contrario.
- **Deterioro cognitivo/demencia:** conjunto de trastornos neurológicos que afectan la memoria, el pensamiento, la comunicación y otras habilidades mentales. Es una variable binaria cuyo valor es 1 si el paciente tiene antecedentes de demencia senil / alzheimer, y 0 en caso contrario.
- **Ictus previo:** variable binaria cuyo valor es 1 si el diagnóstico del paciente incluye los siguientes términos: ACV (accidente cerebrovascular) / AIT (accidente isquémico transitorio) / infarto cerebral, y 0 en caso contrario.



- **Embolia sistémica previa:** momento en el que se bloquea el flujo de sangre causando daño en el órgano o tejido afectado. Es una variable binaria cuyo valor es 1 si el paciente tiene antecedentes de embolia sistémica, y 0 en caso contrario.
- **Sangrado mayor previo:** variable binaria cuyo valor es 1 si el paciente tiene antecedentes de hemorragia mayor (cerebral, retroperitoneal, digestiva) y 0 en caso contrario.

4.3.4. Fármacos

Desde los historiales farmacológicos electrónicos, se recogieron los fármacos de uso crónico que el paciente ya tomaba o empezó a tomar durante el primer mes de tratamiento anticoagulante.

- **Fecha de inicio de tratamiento** anticoagulante con acenocumarol (dd/mm/yyyy).
- Fármacos utilizados para control del ritmo (antiarrítmicos):
 - **Amiodarona:** variable binaria cuyo valor es 1 si el paciente toma amiodarona / trangorex, y 0 en caso contrario.
 - **Otros antiarrítmicos:** variable binaria cuyo valor es 1 si el paciente ingiere alguno de los siguientes fármacos: flecainida / propafenona / dronedarona, y 0 en caso contrario.
- Fármacos utilizados para el control de la frecuencia cardiaca:
 - Medicamento bloqueador de canales de calcio:
 - **Verapamilo:** variable binaria cuyo valor es 1 si el paciente toma verapamilo, y 0 en caso contrario.
 - **Diltiazem:** variable binaria cuyo valor es 1 si el paciente toma diltiazem, y 0 en caso contrario.
 - **Digoxina:** cardiotónico digitálico. Es una variable binaria cuyo valor es 1 si el paciente toma digoxina, y 0 en caso contrario.
 - **Sotalol:** medicamento betabloqueante no selectivo. Es una variable binaria cuyo valor es 1 si el paciente toma sotalol, y 0 en caso contrario.
 - **Otros β -bloqueantes:** variable binaria cuyo valor es 1 si el paciente toma alguno de estos fármacos: bisoprolol / atenolol / carvedilol, y 0 en caso contrario.
- Antiagregantes plaquetarios: fármacos utilizados en la prevención de enfermedades cardiovasculares.
 - **Aspirina:** variable binaria cuyo valor es 1 si el paciente toma aspirina, y 0 en caso contrario. En este caso, se consideró como aspirina los siguientes términos: AAS (ácido acetil salicílico) / tromalyt / adiro.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

- **Otros antiagregantes plaquetarios:** variable binaria cuyo valor es 1 si el paciente toma alguno de los siguientes fármacos: triflusal, inhibidores de P2Y12 (ticagrelor, ticlopidina, clopidogrel, prasugrel, cangrelor), dipyridamol, y 0 en caso contrario.
- **Interacciones del acenocumarol con otros fármacos de uso crónico:** según servicio comprobador de interacciones proporcionado por la aplicación digital “iDoctus” a partir de la base de datos farmacológica del Consejo General de Colegios Oficiales Farmacéuticos. Dicha aplicación clasifica la lista de fármacos que interactúan con el acenocumarol en tres categorías: leve, moderada o grave. Estos fármacos se muestran en la Tabla 1.

Interacciones	Fármacos
Leves	Axitinib, Dimetil fumarato, Fostamatinib, Glecaprevir, Inotersen, Ipilimumab, Lactulosa, Ácido obeticolico, Parafina líquida, Pegaspargase, Pibrentasvir, Pixantrone, Sofosbuvir, Tetracosactido, Ácido valproico, Velpatasvir, Volanesorsen, Voxilaprevir, Atorvastatina, Ciclosporina, Ciprofloxacino, Colestiramina, Itraconazol, Levofloxacino, Moxifloxacino, Norfloxacino, Ofloxacino, Oxcarbazepina, Paracetamol, Pravastatina, Raloxifeno, Tocoferol, Tocofersolan.
Moderadas	Aceclofenaco, Alopurinol, Ácido ascórbico, Azitromicina, Benzbromarona, Bosentan, Capecitabina, Cilostazol, Cisplatino, Citalopram, Citarabina, Claritromicina, Clindamicina, Colesevelam, Dapoxetina, Dasatinib, Desvenlafaxina, Dipyridamol, Disopiramida, Disulfiramo, Doxorubicina, Dronedarona, Duloxetina, Entacapon, Enzalutamida, Epoprostenol, Eritromicina, Erlotinib, Escitalopram, Esomeprazol, Espiramicina, Etopósido, Etravirina, Exenatida, Ezetimiba, Fenazona, Fenitoína, Fitomenadiona, Fluorouracilo, Fluoxetina, Fluvoxamina, Fondaparinux, Fosaprepitant, Gefitinib, Glucosamina, Iloprost, Irinotecan, Lactitol, Leflunomida, Liraglutida, Lixisenatida, Lorlatinib, Mitotano, Nevirapina, Noscipina, Omega 3 ésteres etílicos, Orlistat, Paclitaxel, Paroxetina, Piracetam, Procarbazina, Proguanil, Propranolol, Roxitromicina, Serenoa, Sertralina, Sorafenib, Sugammadex, Sunitinib, Tamoxifeno, Tegafur, Tibolona, Tolcapona, Tolterodina, Trastuzumab emtansina, Trazodona, Vemurafenib, Venlafaxina, Vincristina, Vindesina, Vortioxetina, Ácido acetilsalicílico, Amoxicilina, Ampicilina, Atazanavir, Azatioprina, Bencilpenicilina, Betametasona, Bicalutamida, Bifonazol, Bivalirudina, Carbamazepina, Carbimazol, Carnitina, Cefazolina, Cefixima, Cefoxitina, Ceftriaxona, Cefuroxima, Celecoxib, Cetirizina, Danazol, Darunavir, Deflazacort, Dexametasona, Dexibuprofeno, Dexketoprofeno, Diclofenaco, Eptifibatida, Eslicarbazepina, Estradiol, Estriol, Etinilestradiol, Fenilbutazona, Fenobarbital, Fluconazol, Fludrocortisona, Flutamida, Fluvastatina, Fosamprenavir, Gemfibrozilo, Glucagon, Hidrocortisona, Ibuprofeno, Ifosfamida, Indometacina, Isoniazida, Ledipasvir, Levotiroxina, Lopinavir, Lornoxicam, Lovastatina, Ácido mefenámico, Meloxicam, Mercaptopurina, Metamizol, Metilfenidato, Metilprednisolona, Miconazol, Minociclina, Nabumetona, Naproxeno, Ácido niflumico, Parecoxib, Pentoxifilina, Piperacilina, Piroxicam, Pitavastatina, Prednisolona, Prednisona, Propafenona, Ritonavir, Rosuvastatina, Simvastatina, Tenoxicam, Terbinafina, Tetraciclina, Tiamazol, Tigeciclina, Tirofiban, Tramadol, Triamcinolona, Voriconazol.
Graves	Bezafibrato, Clopidogrel, Fenofibrato, Imatinib, Pelargonium, Prasugrel, Rifampicina, Salvia, Treprostnilo, Ácido acetilsalicílico, Amiodarona, Cloranfenicol, Doxiciclina, Metronidazol, Neomicina, Omeprazol, Sulfametoxazol, Ticlopidina.

Tabla 1. Fármacos considerados como leves/moderados/graves en la interacción con acenocumarol



- **Número de pastillas:** variable discreta que representa el número total de pastillas diarias de otros fármacos distintos al acenocumarol.

Posteriormente, se decidió recolectar también aquellos datos relacionados a las analíticas (hallazgos de laboratorio) y ecocardiográficos más cercanos a la fecha del diagnóstico (como máximo ± 1 mes), con el objetivo de extender la base de datos y encontrar nuevas relaciones con el TRT.

4.3.5. Hallazgos de laboratorio

- **ADE** (normal=[11.5-14]): variable continua que indica la variabilidad en el tamaño de los glóbulos rojos en una muestra de sangre (en %).
- **Leucocitos** (normal=[3.8-10.8]): células sanguíneas blancas que forman parte del sistema inmunológico y ayudan a proteger el cuerpo contra infecciones y enfermedades. Variable continua con el recuento de leucocitos expresado en número $\times 10^9$ /Litro (L).
- **Neutrófilos** (normal=[1.8-7.5]): tipo de glóbulo blanco presente en la sangre, que ayudan al cuerpo a combatir infecciones y procesos inflamatorios. Es una variable continua del recuento de neutrófilos expresado en número $\times 10^9$ /L .
- **Linfocitos** (normal=[1.5-4]): tipo de células sanguíneas que forman parte del sistema inmunológico. Es una variable continua que mide el recuento de linfocitos expresado en número $\times 10^9$ /L.
- **Monocitos** (normal=[0.2-0.8]): tipo de glóbulo blanco de la sangre que ayuda al sistema inmunológico a combatir infecciones y eliminar células muertas o dañadas. Es una variable continua que expresa el recuento de monocitos expresado en número $\times 10^9$ /L.
- **Hemoglobina** (normal=[12.5-16]): proteína presente en los glóbulos rojos de la sangre que se encarga de transportar el oxígeno desde los pulmones a los tejidos del cuerpo. Es una variable continua que representa el recuento de hemoglobina expresado en gramos (g)/decilitro (dL).
- **Plaquetas** (normal=[135-350]): pequeñas células sanguíneas que ayudan en la coagulación de la sangre y en la reparación de los vasos sanguíneos dañados. Es una variable continua la cual expresa el recuento de plaquetas en número $\times 10^9$ /L.

4.3.6. Hallazgos ecocardiográficos

- **Disfunción ventricular:** se refiere a un problema en el funcionamiento del ventrículo izquierdo lo que puede resultar en una disminución de la capacidad del corazón. Es una variable binaria cuyo valor es 1 si la ecocardiografía del paciente muestra una fracción de eyección del ventrículo izquierdo (VI) $< 55\%$ y 0 en caso contrario.
- **Aurícula izquierda dilatada:** condición médica en la que la aurícula izquierda (AI) del corazón se agranda. Es una variable binaria cuyo valor es 1 si el volumen de la aurícula izquierda, medida por ecocardiografía, es mayor a 45 ml y 0 en caso contrario.

4.3.7. Variables derivadas

Posteriormente, se calculó el **score SAME-TT₂R₂** (variable discreta entre 0 y 8) como predictor de rango terapéutico en la serie de pacientes del CHGUV. Sin embargo, en este estudio el máximo era de 6 puntos, dado que toda la población era caucásica. Las variables incluidas en dicho *score*, así como su puntuación se muestran en la Tabla 2.

Sexo femenino	1 punto
Edad < 60 años	1 punto
>2 de las siguientes comorbilidades: HTA, DM, cardiopatía isquémica, arteriopatía periférica, IC, ictus previo, enfermedad pulmonar, enfermedad hepática o renal	1 punto
Tratamiento con amiodarona	1 punto
Tabaquismo en los dos últimos años	2 puntos
Etnia no caucásica	2 puntos

Tabla 2. Variables incluidas en el *score* SAME-TT₂R₂ y su aportación

Además, se añadieron otras variables dependientes derivadas con el objetivo de capturar aspectos relevantes, enriquecer la información disponible, explorar relaciones más profundas entre las características del conjunto de datos y ofrecer una perspectiva más completa y detallada del fenómeno a estudio (TRT):

- **Anemia:** según la Organización Mundial de la Salud, se considera anemia una cifra de hemoglobina menor a 13g/dL para el hombre y 12 g/dL para la mujer, por lo que es una variable binaria cuyo valor es 1 si se cumple esta condición, y 0 en caso contrario.
- **Trombocitopenia:** variable binaria cuyo valor es 1 si la cifra de plaquetas es inferior a $100 \times 10^9/L$, y 0 en caso contrario.
- **Edad:** variable discreta calculada a partir de la resta de la fecha de inicio de tratamiento anticoagulante con acenocumarol y la fecha de nacimiento (en años).
- **Ratio neutrófilo/linfocitos:** variable continua calculada a partir de la división entre la cifra de neutrófilos y linfocitos.
- **Cualquier fármaco antiarrítmico:** variable binaria cuyo valor es 1 si el paciente toma amiodarona u otros antiarrítmicos, y 0 en caso contrario.
- **Control frecuencia:** variable binaria cuyo valor es 1 si el paciente toma betabloqueantes o sotalol o verapamilo o diltiazem o digoxina, y 0 en caso contrario.
- **Cualquier antiagregante:** variable binaria cuyo valor es 1 si el paciente toma aspirina u otros antiagregantes, y 0 en caso contrario.
- **Edad<60:** variable binaria cuyo valor es 1 si el paciente es mayor o igual a 60 años, y 0 en caso contrario. Esta variable forma parte en el cálculo del *score* SAME-TT₂R₂, por lo que se ha decidido incorporar como variable propia, dicoyomizando la variable “Edad”.
- **Número de comedificaciones:** variable categórica cuyo valor es 1 si el paciente toma cualquier fármaco para el control de frecuencia o cualquier antiarrítmico, 2 si toma ambos tipos de fármacos, y 0 si no toman ninguno.



- **Ictus/embolia arterial sistémica:** variable binaria cuyo valor es 1 si el paciente tiene antecedentes de ictus o embolia arterial sistémica, y 0 en caso contrario.
- **Enfermedad vascular:** variable binaria cuyo valor es 1 si el paciente tiene antecedentes de cardiopatía isquémica o arteriopatía isquémica, y 0 en caso contrario.
- **Disfunción ventricular / Dilatación aurícula izquierda:** variable binaria cuyo valor es 1 si la ecocardiografía del paciente muestra una de estas dos alteraciones ya definidas previamente, y 0 en caso contrario.
- **Comorbilidades incluidas en SAME-TT₂R₂:** variable binaria cuyo valor es 1 si el paciente tiene más de 2 antecedentes de estas comorbilidades: HTA, DM, cardiopatía isquémica, arteriopatía periférica, IC, ictus previo, enfermedad hepática o renal, y 0 en caso contrario.
- **Doble antiagregación:** variable binaria cuyo valor es 1 si el paciente toma aspirina y otros antiagregantes, y 0 en caso contrario.
- **Enfermedad hepática/renal:** variable binaria cuyo valor es 1 si el paciente tiene antecedentes de enfermedad hepática o enfermedad renal, y 0 en caso contrario.
- **Bloqueadores de calcio:** variable binaria cuyo valor es 1 si el paciente toma verapamilo o diltiazem, y 0 en caso contrario.

4.3.8. Descriptiva de las características clínico-biológicas, analíticas y ecocardiográficas de los pacientes

Las características clínico-biológicas, analíticas y ecocardiográficas de los 1000 pacientes, así como de los 800 utilizados para el entrenamiento del modelo y 200 para su testeo, en el momento del diagnóstico de la FANV se encuentran expresadas en las Tablas 3 y 4:

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

Variables binarias/categorías	Todos n=1000 Nº pacientes (%)	Entrenamiento n= 800 Nº pacientes (%)	Test / Prueba n= 200 Nº pacientes (%)
Características demográficas			
Sexo femenino	474 (47.4)	383 (47.88)	91 (45.5)
Edad<60	75 (7.5)	61 (7.63)	15 (7.5)
Comorbilidades y factores de riesgo			
Dislipemia	599 (59.9)	472 (59)	127 (63.5)
Tabaquismo	388 (38.8)	309 (38.63)	79 (39.5)
Enfermedad hepática/renal	189 (18.9)	149 (18.63)	40 (20)
Enfermedad renal	155 (15.5)	120 (15)	35 (17.5)
Enfermedad hepática	45 (4.5)	38 (4.75)	7 (3.5)
Enfermedad pulmonar	233 (23.3)	176 (22)	57 (28.5)
Neoplasia	112 (11.2)	87 (10.88)	25 (12.5)
Antecedentes	92 (9.2)	70 (8.75)	22 (11)
Activa	20 (2)	17 (2.13)	3 (1.5)
Apnea del sueño	89 (8.9)	68 (8.5)	21 (10.5)
Sangrado mayor	28 (2.8)	24 (3)	4 (2)
Abuso de alcohol	23 (2.3)	20 (2.5)	3 (1.5)
Anemia	203 (20.3)	161 (20.13)	42 (21)
Trombocitopenia	7 (0.7)	6 (0.75)	1 (0.5)
Enfermedad vascular	274 (27.4)	215 (26.88)	59 (29.5)
Cardiopatía isquémica	215 (21.5)	165 (20.63)	50 (25)
Arteriopatía periférica	96 (9.6)	78 (9.75)	18 (9)
HTA	752 (75.2)	598 (74.75)	154 (77)
DM	346 (34.6)	274 (34.25)	72 (36)
IC	229 (22.9)	182 (22.75)	47 (23.5)
Deterioro cognitivo/demencia	13 (1.3)	11 (1.38)	2 (1)
Ictus/embolia arterial	129 (12.9)	104 (13)	25 (12.5)
Ictus previo	110 (11)	89 (11.13)	21 (10.5)
Embolia sistémica	33 (3.3)	27 (3.38)	6 (3)
Comorbilidades en SAME-TT ₂ R ₂ >2	365 (36.5)	289 (36.13)	76 (38)
Fármacos			
Cualquier antiarrítmico	186 (18.6)	141 (17.63)	45 (22.5)
Amiodarona	128 (12.8)	98 (12.25)	30 (15)
Otros antiarrítmicos	60 (6)	45 (5.63)	15 (7.5)
Control frecuencia	545 (54.5)	444 (55.5)	101 (50.5)
Sotalol	2 (0.2)	2 (0.25)	0 (0)
Otros β -bloqueantes	480 (48)	395 (49.38)	85 (42.5)
Bloqueadores de calcio	44 (4.4)	32 (4)	12 (6)
Verapamilo	11 (1.1)	8 (1)	3 (1.5)
Diltiazem	33 (3.3)	24 (3)	9 (4.5)
Digoxina	46 (4.6)	37 (4.63)	9 (4.5)
Nº de comedificaciones			
1	505 (50.5)	413 (51.63)	92 (46)
2	113 (11.3)	86 (10.75)	27 (13.5)
Cualquier antiagregante	198 (19.8)	163 (20.3)	35 (17.5)
AAS	167 (16.7)	138 (17.25)	29 (14.5)
Otros antiagregantes	51 (5.1)	40 (5)	11 (5.5)
Doble antiagregación	20 (2)	15 (1.88)	5 (2.5)
Interacciones con otro fármaco			
Leves	66 (6.6)	48 (6)	18 (9)
Moderadas	209 (20.9)	167 (20.88)	42 (21)
Graves	379 (37.9)	309 (38.63)	70 (35)
Hallazgos ecocardiográficos			
Disfunción VI / Dilatación AI	671 (67.1)	545 (68.13)	126 (63)
Disfunción ventricular	116 (11.6)	96 (12)	20 (10)
Aurícula izquierda dilatada	640 (64)	518 (64.75)	122 (61)

Tabla 3. Descripción de las características de las variables categóricas en el conjunto de datos (total, entrenamiento y prueba)



VARIABLES CONTINUAS/DISCRETAS	Todos n= 1000 media ± DS¹ (rango)	Entrenamiento n= 800 media ± DS (rango)	Test / Prueba n=200 media ± DS (rango)
Características demográficas			
Edad, años	73.45 ± 9.46 (33-95)	73.62 ± 9.47 (33-95)	72.78 ± 9.44 (41-91)
Comorbilidades y factores de riesgo			
IMC, Kg/m ²	30.23 ± 6.28 (15-67.9)	30.16 ± 6.26 (15-67.9)	30.48 ± 6.35 (17.9-60)
Hallazgos de laboratorio			
ADE, %	14.48 ± 1.68 (12-28.9)	14.5 ± 1.77 (12-28.9)	14.37 ± 1.3 (12-20.1)
Leucocitos, x 10 ⁹ /L	8.31 ± 3.02 (3.2-42.9)	8.24 ± 3.08 (3.2-42.9)	8.58 ± 2.78 (3.6-26.1)
Neutrófilos, x 10 ⁹ /L	5.41 ± 2.61 (1.3-32.5)	5.37 ± 2.65 (1.3-32.5)	5.57 ± 2.42 (1.9-19.5)
Linfocitos, x 10 ⁹ /L	2.02 ± 1.37 (0.3-36.4)	1.99 ± 1.46 (0.3-36.4)	2.1 ± 0.95 (0.3-6.4)
Monocitos, x 10 ⁹ /L	0.68 ± 0.26 (0.1-2.3)	0.67 ± 0.26 (0.1-2.3)	0.7 ± 0.24 (0.1-1.5)
Hemoglobina, g/dL	13.75 ± 1.85 (0.1-18.5)	13.75 ± 1.89 (0.1-18.5)	13.77 ± 1.65 (9.3-18.1)
Plaquetas, x 10 ⁹ /L	216.09 ± 71.48 (57-798)	213.86 ± 68.67 (57-725)	225 ± 81.33 (87-798)
Ratio neutrófilo/linfocito	3.47 ± 3.27 (0.11-37)	3.51 ± 3.38 (0.11-37)	3.32 ± 2.79 (0.3-25)
Fármacos			
Número total de pastillas diarias	6.29 ± 5.21 (0-26)	6.26 ± 5.18 (0-23)	6.43 ± 5.35 (0-26)

Tabla 4. Descripción de las características de las variables numéricas en el conjunto de datos (total, entrenamiento y prueba)

Adicionalmente, se aplicó la prueba de Kolmogorov-Smirnov para analizar las diferencias entre las variables numéricas de los conjuntos de datos de entrenamiento y prueba. La prueba de Kolmogorov-Smirnov permitió evaluar si las muestras provenían de la misma distribución o si había variaciones significativas entre ellas. Al obtener un valor $p > 0.05$ para todas las variables, se demostró que los datos de entrenamiento y prueba eran representativos y estaban equilibrados, evitando posibles sesgos y garantizando la validez de los resultados obtenidos.

4.4. Evaluación del *score* SAME-TT₂R₂

Se calculó el *score* SAME-TT₂R₂ en la población incluida en este estudio. Como se refleja en la Tabla 5, el grupo mayoritario (1) contiene 290 pacientes, mientras que el grupo minoritario (5) contiene únicamente 14. Se puede ver una tendencia en los datos, donde el número de individuos crece hasta el grupo 2 y, a partir de este, decrece. El grupo minoritario mostró la

¹ DS = desviación estándar

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

mayor dispersión entre los valores del TRT. También cabe destacar la diferencia entre las medianas de los subgrupos (0, 1, 2) y (3, 4, 5), siendo ésta mayor en los primeros grupos.

SAMe-TT ₂ R ₂	Nº pacientes	TRT Media	TRT DS	TRT 1º Cuartil	TRT Mediana	TRT 3º Cuartil
0	162	58.58	19.35	44.4	57.1	71.4
1	290	57.05	18.53	44.4	55.6	71.4
2	242	58.4	19.08	44.4	55.6	71.4
3	193	55.85	20.42	41.2	50	71.4
4	99	53.8	19.8	40.85	50	66.7
5	14	54.26	29.97	27.3	50	88.03

Tabla 5. Datos estadísticos del TRT por puntuación del *score* SAMe-TT₂R₂

Para evaluar la posible asociación entre el TRT y el *score* SAMe-TT₂R₂ se aplicó la prueba de *Kruskal-Wallis*. Sin embargo, los resultados no mostraron una relación significativa entre ambas variables (Figura 14), ya que el p valor resultó ser 0.25.

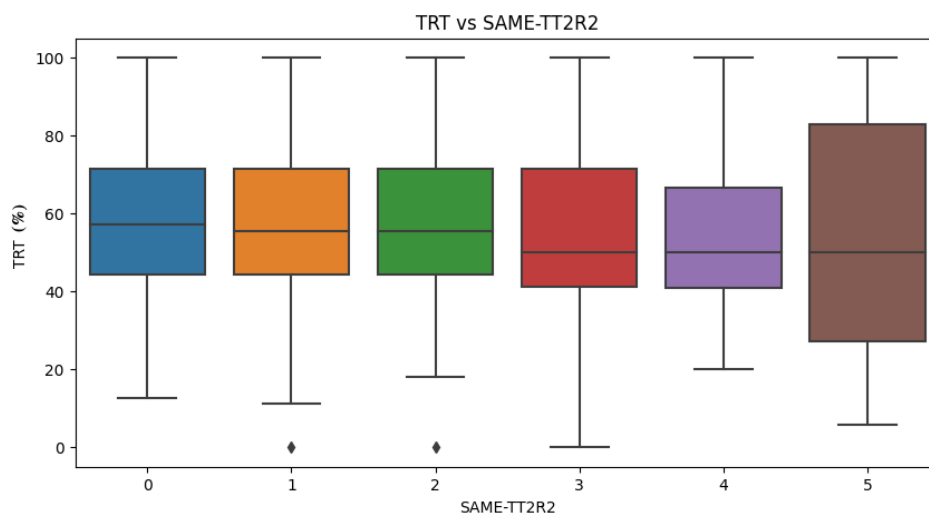


Figura 14. Comparación de la distribución de TRT según cada puntuación del SAMe-TT₂R₂

A pesar de la falta de significación estadística, es importante tener en cuenta que esto no descarta por completo la utilidad del *score* SAMe-TT₂R₂. Por ello, se realizó un análisis más detallado y exhaustivo de ambas variables abordando diferentes enfoques para analizar la relación entre el TRT y el *score* SAMe-TT₂R₂. Se consideraron distintas formas de definir el TRT subóptimo, utilizando tanto valores continuos como criterios de corte específicos (TRT < 60%, TRT < 65% y TRT < 70%). Estas variables se crearon como variables binarias con valor de 0 en caso de cumplirse estos requerimientos y 1 en caso contrario.



Por otro lado, se examinaron diferentes puntos de corte para los valores del score SAME-TT₂R₂, incluyendo SAME-TT₂R₂ ≥ 2, SAME-TT₂R₂ ≥ 3 y SAME-TT₂R₂ ≥ 4. Estas variables se crearon como variables binarias con valor de 1 en caso de cumplirse estos requerimientos y 0 en caso contrario.

En la Tabla 6 se muestran las pruebas efectuadas según los tipos de variables comparadas. La elección de las pruebas estadísticas se basó en la naturaleza de las variables a comparar y en los supuestos requeridos para cada una de ellas. El test de *Kruskal-Wallis* fue aplicado para comparar las medianas de varias muestras independientes que incumplían el supuesto de normalidad. El test de *chi* cuadrado se utilizó para evaluar la independencia entre dos variables categóricas. Además, se empleó el test de *Mann-Whitney U* para comparar las distribuciones de dos muestras independientes cuando los datos no seguían una distribución normal. Por último, se utilizó el test de *Fisher* para analizar la asociación entre dos variables categóricas (binarias) con muestras pequeñas.

	Continua	Binaria
Categórica (>2 clases)	<i>Kruskal-Wallis</i>	Chi-cuadrado
Binaria	<i>Mann-Whitney U</i>	<i>Fisher</i>

Tabla 6. Pruebas de significancia aplicadas según el tipo de variable

Al comparar el SAME-TT₂R₂ con cada uno de los puntos de corte de TRT establecidos (Tabla 7), se pudo corroborar la ausencia de diferencias significativas entre el % de pacientes con TRT subóptimo (< 60%, < 65% y < 70%) y el % de pacientes con TRT óptimo (≥ 60%, ≥ 65% y ≥ 70% respectivamente) para cada una de las puntuaciones del SAME-TT₂R₂, obteniendo valores $p > 0.05$. Cabe destacar que la mediana de TRT para todos los grupos fue inferior al menor punto de corte considerado como subóptimo (60%).

SAME-TT₂R₂ score	TRT < 60 % N° (%)	TRT < 65 % N° (%)	TRT < 70% N° (%)	TRT Mediana (RIQ)
0	87 (53.70)	98 (60.49)	108 (66.66)	57.1 (44.4-71.4)
1	164 (56.55)	198 (68.27)	211 (72.75)	55.6 (44.4-71.4)
2	134 (55.37)	154 (63.63)	168 (69.42)	51.6 (44.4-71.4)
3	118 (61.13)	130 (67.35)	137 (70.98)	50 (41.2-71.4)
4	62 (62.62)	72 (72.72)	75 (75.75)	50 (40.85-66.7)
5	8 (57.14)	8 (57.14)	8 (57.14)	50 (27.3-83.03)
	p= 0.608	p=0.296	p= 0.476	p=0.25

Tabla 7. N° pacientes (%) en rango subóptimo y mediana del TRT por puntuación del SAME-TT₂R₂

Al transformar el SAME-TT₂R₂ en una variable binaria donde 1 es el valor asociado a tener ≥ 2 puntos y 0 el tener menos, el % de pacientes con TRT subóptimo fue mayor y la mediana de TRT menor en el grupo ≥ 2 puntos, pero con valores p aún muy lejos de la significancia (Tabla 8).

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

SAMe-TT ₂ R ₂ score	Nº pacientes	TRT < 60 % Nº (%)	TRT < 65 % Nº (%)	TRT < 70 % Nº (%)	TRT Mediana (RIQ)
<2 puntos	452	251 (55.53)	296 (65.48)	319 (70.57)	55.6 (44.4-71.4)
≥2 puntos	548	322 (58.75)	364 (66.42)	388 (70.80)	54.5 (42.9-71.4)
		p = 0.305	p = 0.789	p = 0.944	p = 0.308

Tabla 8. Nº pacientes (%) en rango subóptimo y mediana del TRT según SAMe-TT₂R₂ < 2 o ≥ 2

Por otra parte, cuando se estableció el punto de corte del SAMe-TT₂R₂ en 3, los resultados mejoraron. En este caso, los valores p fueron mucho más pequeños, incluso siendo estadísticamente significativo con el TRT continuo ($p=0.022 < 0.05$). Como se puede identificar mediante la Tabla 9, se encontraron diferencias en las medianas entre los grupos ≥ 3 y < 3 . Mediante la prueba de *Mann-Whitney U* ratificamos estas disparidades que, aunque puedan parecer insignificantes, el orden de los datos también tiene influencia en estos resultados. En la Figura 15 se observa como en el grupo < 3 hay una mayor cantidad de individuos con un TRT por encima de la mediana comparado con el grupo ≥ 3 . Por otro lado, el grupo ≥ 3 tiene más individuos con TRT bajo, lo que confirma la afirmación de que cuanto mayor puntuación en el SAMe-TT₂R₂, peor es el TRT.

SAMe-TT ₂ R ₂ score	Nº pacientes	TRT <60 % Nº (%)	TRT < 65 % Nº (%)	TRT < 70 % Nº (%)	TRT Mediana (RIQ)
<3 puntos	694	385 (55.47)	450 (64.84)	487 (70.17)	55.6 (44.4-71.4)
≥3 puntos	306	188 (61.43)	210 (68.62)	220 (71.89)	50 (40-71.4)
		p=0.083	p= 0.248	p= 0.598	p = 0.022

Tabla 9. Nº pacientes (%) en rango subóptimo y mediana del TRT según SAMe-TT₂R₂ < 3 o ≥ 3

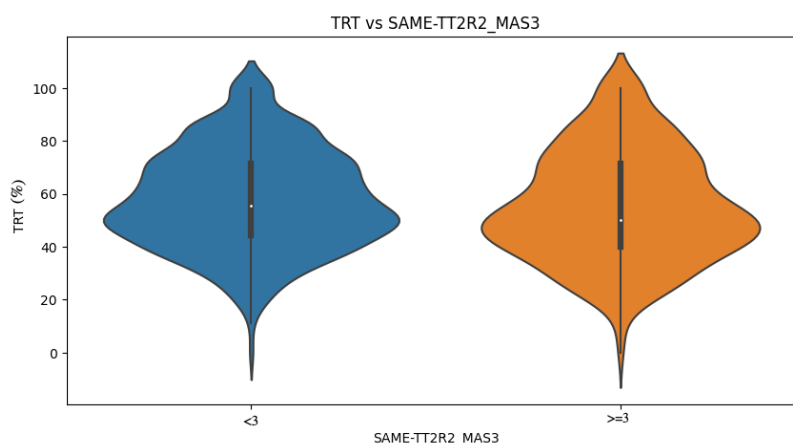


Figura 15. Comparación de la distribución de TRT según SAMe-TT₂R₂ ≥ 3 y < 3

Por último, al establecer como punto de corte 4 puntos para el SAMe-TT₂R₂ (Tabla 10), no se obtuvieron resultados prometedores, ya que los valores de p para todos los enfoques fueron



mayores a 0.05. No obstante, para el TRT continuo se alcanzó un valor p de 0.059, muy cerca de la significancia.

SAMe-TT ₂ R ₂ <i>score</i>	N° pacientes	TRT <60 % N° (%)	TRT < 65 % N° (%)	TRT < 70% N° (%)	TRT Mediana (RIQ)
<4 puntos	887 (88.7)	503 (56.70)	580 (65.38)	624 (70.34)	55.6 (44.4-71.4)
≥4 puntos	113 (11.3)	70 (61.94)	80 (70.79)	83 (73.45)	50 (37.5-71.4)

p= 0.313 p= 0.292 p= 0.583 p = 0.059

Tabla 10. N° pacientes (%) en rango subóptimo y mediana del TRT según SAMe-TT₂R₂ < 4 o ≥ 4

En conclusión, al obtener como único resultado estadísticamente significativo la comparación entre el TRT como variable continua y el *score* SAMe-TT₂R₂ binarizada en < 3 y ≥ 3, se decidió incluir dicho *score* (< 3 y ≥ 3) como parte del grupo de variables potencialmente predictoras en los modelos predictivos desarrollados en este trabajo. Sin embargo, cabe recalcar que las medianas de ambas clases están por debajo de cualquier punto de corte establecido para el TRT óptimo, por lo que, a pesar de ser el único resultado estadísticamente significativo, no se puede interpretar como una buena capacidad discriminativa en estos casos.

4.5. Preparación de datos

4.5.1. Reorganización de datos

A continuación, se procedió a realizar el pretratamiento de los datos. En primer lugar, se estudió el tipo de datos de cada uno de los atributos para garantizar un tratamiento más adecuado de las variables en términos de manipulación, brindando una base sólida para el posterior estudio e interpretación de los resultados obtenidos. Durante esta etapa de investigación, se identificó que las variables fecha de nacimiento, fecha de inicio de tratamiento, sexo, interaccionan con otro fármaco y neoplasia estaban inicialmente clasificadas como tipo "object", el cual generalmente se utiliza para representar datos no numéricos, como nombres, categorías, descripciones o valores alfanuméricos. Con el objetivo de mejorar la calidad y la precisión del análisis, se procedió a transformar dichas variables a un tipo de dato más adecuado, específicamente a un formato de fecha (%d/%m/%Y) en el caso de fecha de nacimiento y fecha de inicio de tratamiento. Para las variables "Interaccionan con otro fármaco" y "Sexo" se realizó una codificación de etiquetas mediante *label encoding* (Interaccionan con otro fármaco: No = 0, Leve = 1, Moderada = 2, Grave = 3; Sexo: H = 0, M = 1). En el caso de "Neoplasia" se efectuó la codificación "*one hot encoding*" al no haber una relación de orden entre las categorías y, así, evitar la introducción de un orden implícito en el modelo.

Al aplicar "*one-hot encoding*", se crea una nueva columna binaria para cada categoría de la variable original. Si una observación pertenece a una categoría específica, se asigna el valor 1 en la columna correspondiente, y se asigna el valor 0 en todas las demás columnas. En este caso, para la variable "Neoplasia" se crean 3 variables *dummy* (Neoplasia_No, Neoplasia_Antecedentes y Neoplasia_Activa). Este enfoque permite capturar de manera explícita la presencia o ausencia de cada categoría, sin hacer suposiciones sobre la relación entre las categorías. Además, "*one-hot encoding*" evita la posibilidad de que el modelo asigne un

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

mayor peso o importancia a una categoría sobre otra debido a una codificación numérica arbitraria.

El resto de variables potencialmente predictoras incluidas en este estudio, se mantuvieron en sus tipos de datos originales (enteros o decimales), sin realizar cambios o transformaciones adicionales.

4.5.2. Imputación de valores nulos o faltantes

Durante el proceso de análisis de datos, se realizó una exhaustiva revisión de los valores nulos presentes en cada variable. La única variable que contenía valores nulos fue el IMC con un total de 36 valores faltantes. En caso de tener el peso (kg) y la altura (m), el IMC se calculó a partir de éstas ($\text{peso} / \text{altura}^2$). En el caso contrario, con el objetivo de preservar la distribución y características originales de esta variable, se optó por imputar los valores faltantes utilizando la media de los valores existentes en el conjunto de datos. Esta estrategia de imputación por media permitió mantener la consistencia y representatividad de la variable IMC, minimizando el impacto en su distribución y evitando distorsiones significativas en los análisis posteriores.

4.5.3. Evaluación de valores anómalos

Se llevó a cabo una detección y evaluación de los valores anómalos en las variables numéricas del conjunto de datos mediante dos enfoques diferentes: el rango intercuartílico (RIQ) y el *Z-score*.

El RIQ se utilizó como una medida de dispersión robusta que se basa en los percentiles 25 y 75 de la distribución. Se identificaron como valores anómalos aquellos que estaban por debajo del percentil 25 menos 1.5 veces el RIQ o por encima del percentil 75 más 1.5 veces el RIQ. Como se puede observar en la Figura 16, en la mayoría de variables solo se encuentran valores anómalos por arriba, excepto para las variables Hemoglobina y Edad, en las cuales estos valores están por debajo. Cabe destacar la gran cantidad de anómalos para la variable Ratio neutrófilo/linfocito (76).



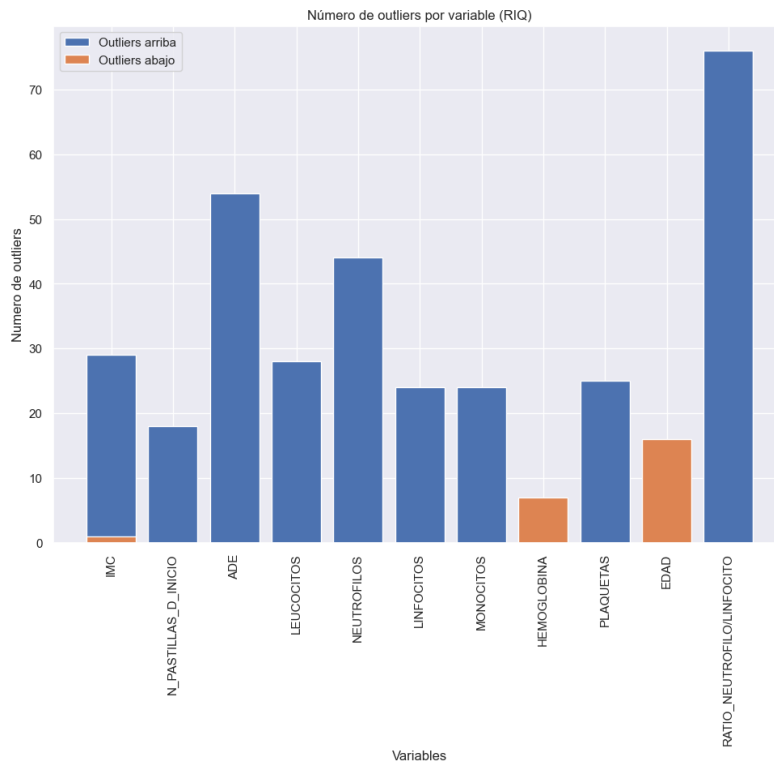


Figura 16. Número de valores considerados anómalos por variable según RIQ

Por otro lado, se aplicó el *Z-score* para evaluar la desviación de cada valor con respecto a la media de la distribución y su desviación estándar. Los valores que superaban un umbral predefinido de *Z-score*, establecido en 3 o -3, fueron considerados como anómalos. En este caso, se sigue observando el mayor número de valores anómalos en Ratio neutrófilo/linfocito y, en contraste, Hemoglobina no contiene ninguno.

Al comparar las técnicas del *Z-score* y el RIQ, se observó que las distribuciones de los datos anómalos en las variables analizadas presentaban similitudes (Figuras 16 y 17). Sin embargo, se puede determinar que el *Z-score* considera un menor número de valores anómalos que el RIQ, aunque estos también están incluidos entre los valores anómalos identificados por el RIQ.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

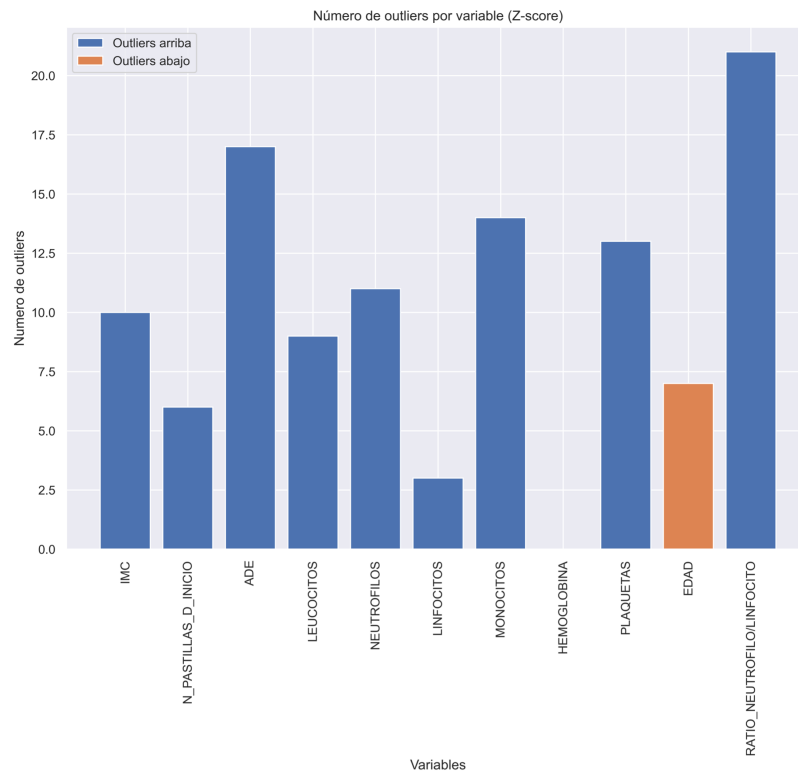


Figura 17. Número de valores considerados anómalos por variable según *Z-score*

Al utilizar tanto el RIQ como el *Z-score*, se obtuvo una evaluación de los valores anómalos en las variables numéricas. Sin embargo, en ese instante no se realizó ningún tratamiento o eliminación de estos datos, ya que se consideró necesario realizar un análisis más profundo y detallado de los datos anómalos para evaluar si estos valores atípicos son el resultado de errores de medición, registros incorrectos o representan información relevante y genuina, y así posteriormente tomar una decisión sobre su manejo.

La Tabla 11 muestra el rango de valores de cada variable no considerados anómalos tanto con RIQ como *Z-score*. Se observa un rango mayor al utilizar el *Z-score*, como se ha dicho previamente.



Variable	RIQ	Z-score
IMC	15.01 - 44.31	11.39 - 49.06
Número de pastillas	-8 - 19	-9 - 21
ADE	11.25 - 17.25	9.42 - 19.53
Leucocitos	2 - 14	-0.76 - 17.38
Neutrófilos	-0.24 - 10.26	-2.4 - 13.22
Linfocitos	-0.1 - 3.9	-2.09 - 6.12
Monocitos	0.05 - 1.25	-0.09 - 1.44
Hemoglobina	9.4 - 18.2	8.54 - 19.02
Plaquetas	54 - 363	2 - 430
Edad	50 - 98	45 - 101
Ratio neutrófilo/linfocito	-1.38 - 7.03	-6.34 - 13.28

Tabla 11. Rangos de valores no considerados anómalos para cada variable según RIQ y Z-score

A fin de comprobar si los valores anómalos en las variables monocitos, linfocitos y neutrófilos podrían ser errores de medición, se verificó si la suma de estos valores se aproximaba al valor de la variable leucocitos, con un margen de error de ± 1 , dado que estas tres variables representan componentes específicos de los leucocitos y se espera que la suma de sus valores sea igual o cercana al valor total de los leucocitos. Para realizar esta comprobación, se sumaron los valores de las variables monocitos, linfocitos y neutrófilos y se comparó el resultado con el valor de la variable leucocitos. Si la diferencia entre estas dos cantidades se encontraba fuera del margen de error establecido (± 1), podría sugerir que los valores anómalos en las variables monocitos, linfocitos y neutrófilos podrían ser errores de transcripción. En este caso, solo había un registro que no se encontraba dentro del margen de error (-1.1), por lo que se revisó la analítica y se decidió preservar el valor original dado que no carecía de coherencia y era completamente válido.

Respecto a los valores anómalos de la Edad, se decidió no modificarlos dada la importancia de preservar la integridad de los datos originales. Es importante tener en cuenta que la edad es una característica intrínseca de los individuos y está determinada por factores biológicos y cronológicos.

En relación al resto de variables numéricas, entre las que se encontraban ADE, IMC, Hemoglobina, Plaquetas, Ratio neutrófilo/linfocitos y Número total de pastillas diarias, se revisaron las analíticas y la HCE de aquellos pacientes en los que se detectó algún valor anómalo en alguna de dichas variables. Esta revisión permitió examinar de cerca los registros médicos y los resultados de las pruebas correspondientes a dichos pacientes, con el objetivo de comprender mejor las posibles causas de las anomalías y determinar si se trataba de errores de

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

medición, registros incorrectos o de condiciones médicas específicas. Los errores de medición y los registros erróneos se corrigieron, mientras que los pacientes con situaciones excepcionales se conservaron tal y como estaban debido a su relevancia en el estudio.

4.5.4. Análisis de distribución de variables numéricas

Posteriormente, se llevó a cabo un análisis minucioso para identificar patrones, tendencias o posibles sesgos en dichas variables. Primeramente, se calculó el sesgo (*skewness*) para evaluar la asimetría de la distribución de cada variable, lo cual indica la tendencia de los datos a concentrarse hacia un extremo de la distribución.

En la Figura 18 se muestran las distribuciones de las variables analizadas. A través de esta representación gráfica, es posible visualizar la forma y la simetría de cada distribución. Se observa que algunas variables presentan un sesgo más pronunciado que otras. Específicamente, se identificaron las variables ADE, Leucocitos, Neutrófilos, Linfocitos, Monocitos, Plaquetas y Ratio neutrófilo/linfocito como las que exhiben un mayor sesgo. Estas variables muestran una tendencia hacia valores más altos o más bajos, lo que sugiere una asimetría en sus distribuciones. Por otro lado, las variables IMC, Número de pastillas, Hemoglobina y Edad parecen tener una distribución más simétrica, sin un sesgo evidente.

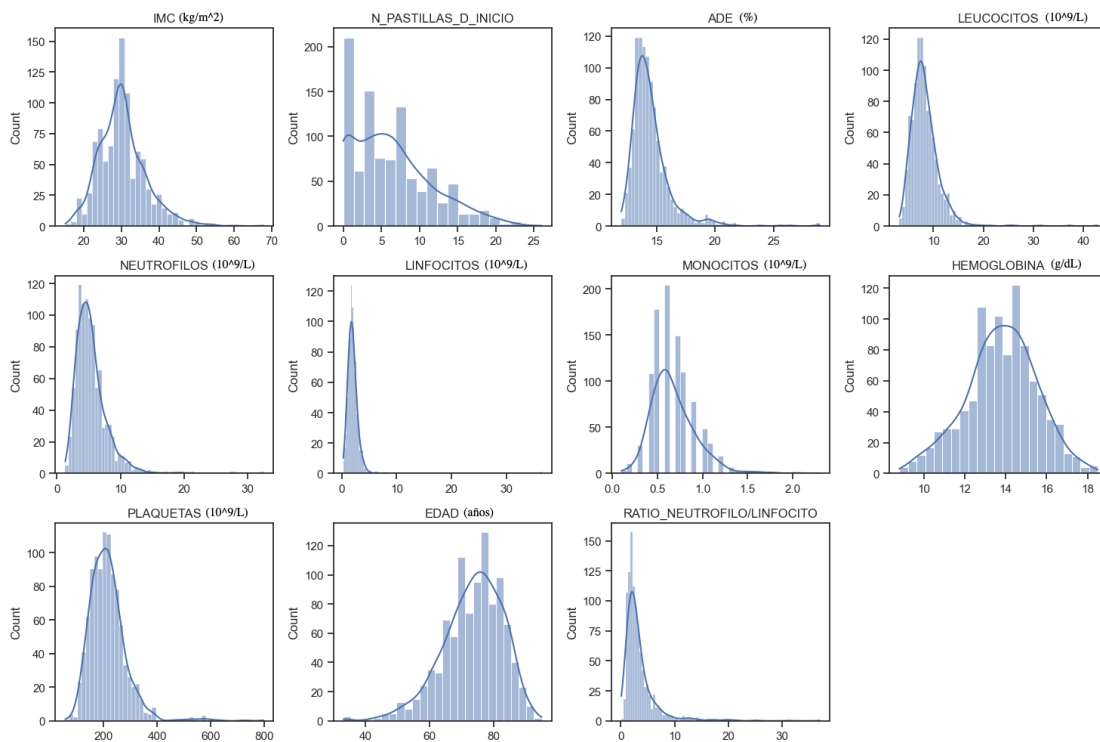


Figura 18. Distribución original de las variables numéricas

Las variables que presentaron un sesgo positivo fueron las siguientes: ADE, Leucocitos, Neutrófilos, Linfocitos, Monocitos, Plaquetas y Ratio neutrófilo/linfocito con un sesgo de 3, 3.65, 3, 16.1, 1.36, 2.22 y 4.12 respectivamente. Como en este caso su sesgo era mayor a 1, se aplicó una transformación logarítmica a la variable con el fin de corregir la asimetría positiva y lograr una distribución más simétrica. En el caso de las variables IMC (0.91) y Número de pastillas (0.79), su sesgo se ubicaba por debajo de 1.



Por otro lado, las variables con sesgo negativo fueron Hemoglobina y Edad con un sesgo de -0.2 y -0.7 respectivamente. Al tener valores por encima de -1, no se les aplicó ninguna transformación.

En la Figura 19 se muestran las distribuciones de las variables numéricas después de aplicarles transformaciones. Se puede observar que, en comparación con las distribuciones originales (Figura 18), las variables han experimentado cambios significativos. Se ha logrado reducir el sesgo y hacer que las distribuciones se asemejen más a una distribución normal.

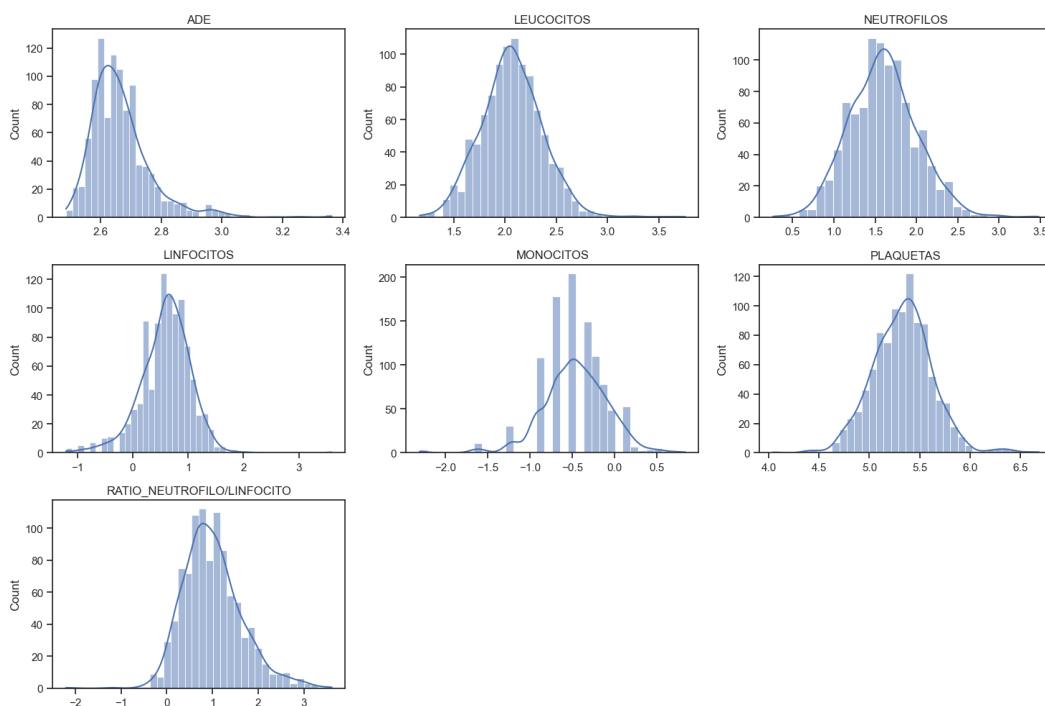


Figura 19. Distribución de las variables numéricas tras las transformaciones

Además de la visualización en la Figura 19, es posible corroborar el efecto de las transformaciones en la aproximación a una distribución normal mediante el valor del sesgo (Tabla 12):

Variable	Sesgo	Variable	Sesgo
ADE	2	Neutrófilos	0.31
Leucocitos	0.49	Monocitos	-0.5
Linfocitos	-0.3	Plaquetas	0.22
Ratio neutrófilo/linfocito	0.56		

Tabla 12. Sesgo de las variables numéricas

4.5.5. Tratamiento de valores anómalos

Dado que la variable ADE seguía teniendo una distribución anormal (sesgo positivo), se imputaron los valores considerados anómalos mediante el *Z-score* por el valor más alto (2.97)/bajo(2.48) no considerado anómalo, dependiendo del valor del *Z-score*, >3 o <-3 respectivamente. A pesar de aplicar el tratamiento para los valores anómalos, el sesgo de la variable ADE se mantuvo en 1.22, lo que indica que aún existe cierto sesgo en la distribución. Sin embargo, se logró corregir parcialmente el sesgo, lo que implica una mejora en la distribución de los valores y una reducción en la concentración de valores extremos.

Respecto a los valores anómalos del IMC, en la mayoría de las clasificaciones del tipo de obesidad, el grado más elevado corresponde a un valor de $IMC \geq 50 \text{ kg/m}^2$ (obesidad extrema o hiperobesidad). Por lo tanto, los valores de IMC superiores a 50 se consideraron atípicos, ya que en la mayoría de los casos, esos valores son extremadamente altos y poco probables. Por ello, se imputaron los valores de IMC mayores a 50 (7 en total) por 50. La imputación de estos valores por 50 se realizó como una estrategia conservadora para evitar la distorsión de la distribución general del IMC y minimizar el impacto de los valores extremos en los análisis posteriores. De esta manera, el sesgo de la variable IMC disminuyó a 0.63, 0.28 por debajo del sesgo original.

4.5.6. Estandarización de variables numéricas

A continuación, se decidió realizar la estandarización de las variables numéricas utilizando el método *Min-Max Scaler* debido a varias razones fundamentales. La estandarización es un paso esencial en el procesamiento de datos para garantizar que las variables numéricas estén en una escala comparable y que ninguna variable tenga un impacto desproporcionado en el análisis.

El *Min-Max Scaler* transforma los datos en un rango específico, generalmente entre 0 y 1, al aplicar una fórmula que escala cada valor de la variable proporcionalmente a su posición en la distribución original. Esta técnica es especialmente útil cuando se desea preservar la interpretación de los datos en términos de sus valores originales y evitar problemas relacionados con la escala de las variables.

4.5.7. Análisis bivariante entre TRT y variables numéricas

Análisis de Componentes Principales

Se aplicó un Análisis de Componentes Principales (PCA) a las variables numéricas con el objetivo de explorar su relación con la variable objetivo. En este caso, se utilizó el TRT como variable binaria con los diferentes *cutoffs* (60%, 65% y 70%). El PCA es una técnica de reducción de dimensionalidad que permite identificar patrones y estructuras en los datos al transformar las variables originales en un nuevo conjunto de variables no correlacionadas llamadas componentes principales.

Se calcularon las componentes principales y se examinó su contribución a la variabilidad total de los datos (Figura 20). Se puede observar que con las 3 primeras componentes solo se explica el 50% de la variabilidad de los datos y con 6 componentes se llega al 80%, por lo que



se puede inferir que existe una alta complejidad o una gran dispersión en los datos, sugiriendo que hay factores adicionales que contribuyen a la variabilidad restante.

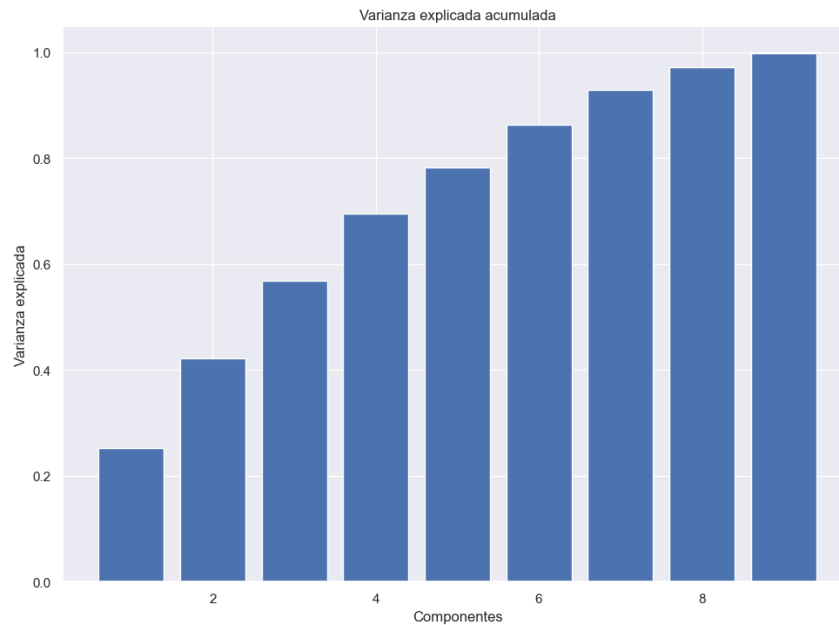


Figura 20. Varianza explicada acumulada del PCA con las variables numéricas

Además, se representó gráficamente el resultado del PCA mediante un *biplot* (Figura 21), donde se mostraron las primeras dos componentes principales junto con las flechas que indican la dirección y la fuerza de asociación de las variables numéricas con dichas componentes. Al interpretar la Figura 21, se puede identificar la gran contribución de las variables ADE y N° de pastillas de forma positiva a la primera componente principal, mientras que Hemoglobina contribuye de forma negativa. Por otro lado, en relación a la segunda componente principal, se atisba una gran importancia de las variables Neutrófilos y Leucocitos de forma positiva. Al estar cerca una de la otra, significa que estas variables tienen una alta correlación positiva en estas dos primeras componentes, lo que no significa que estén realmente correlacionadas. Lo mismo ocurre con Monocitos y Leucocitos, ADE y N° de pastillas, IMC y Neutrófilos, Plaquetas y IMC, Linfocitos y Hemoglobina. Por otro lado, de forma inversa, Hemoglobina/Linfocitos y Edad/ADE.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

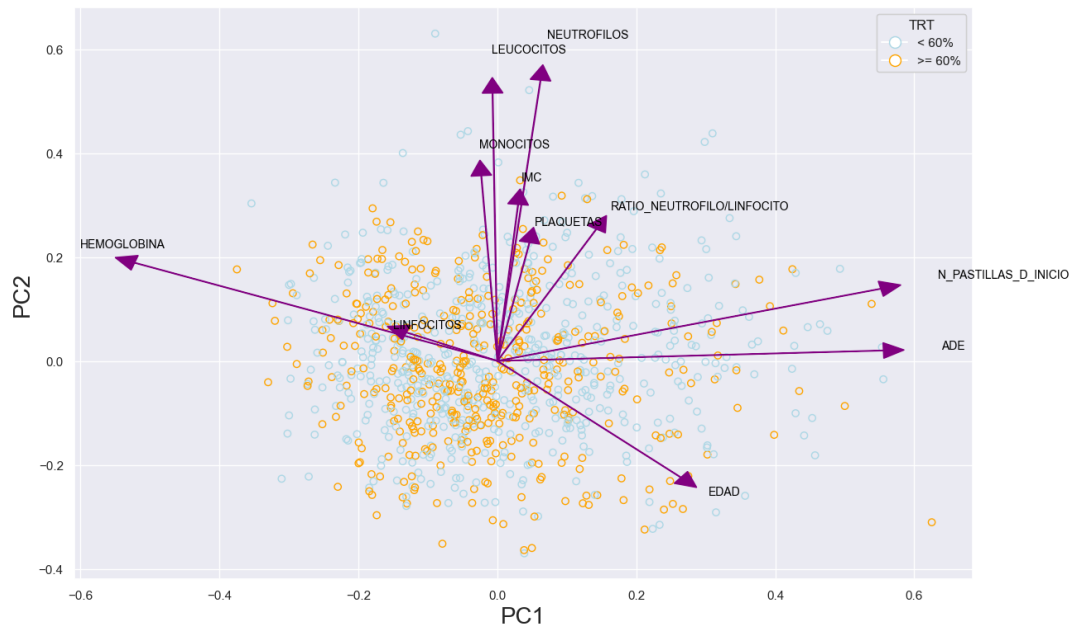


Figura 21. Importancia de las variables numéricas en las primeras 2 componentes principales y TRT < 60%

Asimismo, en el *biplot* generado a partir del PCA, se puede observar que los puntos correspondientes a cada clase no presentan una clara separación. Esto indica que las clases no son fácilmente separables, o al menos utilizando las variables definidas por PCA. Esta falta de separación puede deberse a la alta similitud o superposición de los valores de las variables entre las clases, lo que dificulta la identificación visual de patrones distintivos. Lo mismo ocurre con los puntos de corte 65% y 70%.

Correlación entre variables numéricas

En el análisis de las relaciones entre las variables numéricas y la variable objetivo (TRT), se emplearon dos medidas de correlación: la correlación de *Spearman* y la correlación de *Pearson*. Cada una de estas medidas tiene su utilidad y justificación en función de la naturaleza de las variables y los objetivos del estudio.

La correlación de *Spearman* es especialmente útil cuando se busca evaluar asociaciones monótonas, es decir, cuando los valores de una variable tienden a aumentar o disminuir en conjunto con los valores de la otra variable, pero no necesariamente de forma lineal. La matriz de correlación de *Spearman* se puede observar en la Figura 22, donde se observa correlaciones muy bajas con la variable objetivo (TRT), siendo la correlación positiva más alta con Edad ($r = 0.056$) y la correlación negativa más alta con monocitos ($r = -0.077$).



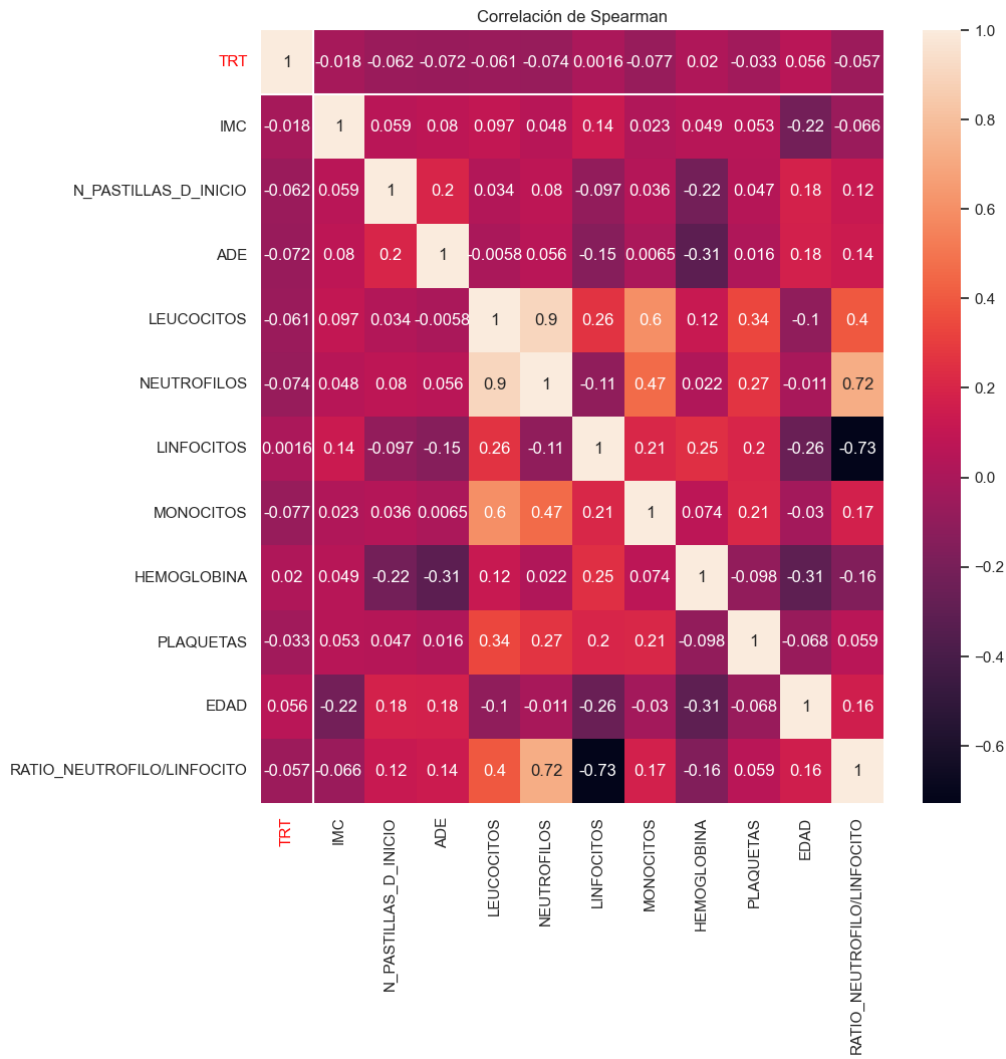


Figura 22. Mapa de calor de la correlación de *Spearman* con las variables numéricas y la variable objetivo TRT

Por otra parte, la correlación de *Pearson* se aplica cuando se busca medir la relación lineal entre dos variables. Los resultados de la correlación de *Pearson* son casi idénticos a los de la correlación de *Spearman*, como se puede identificar en la Figura 23.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

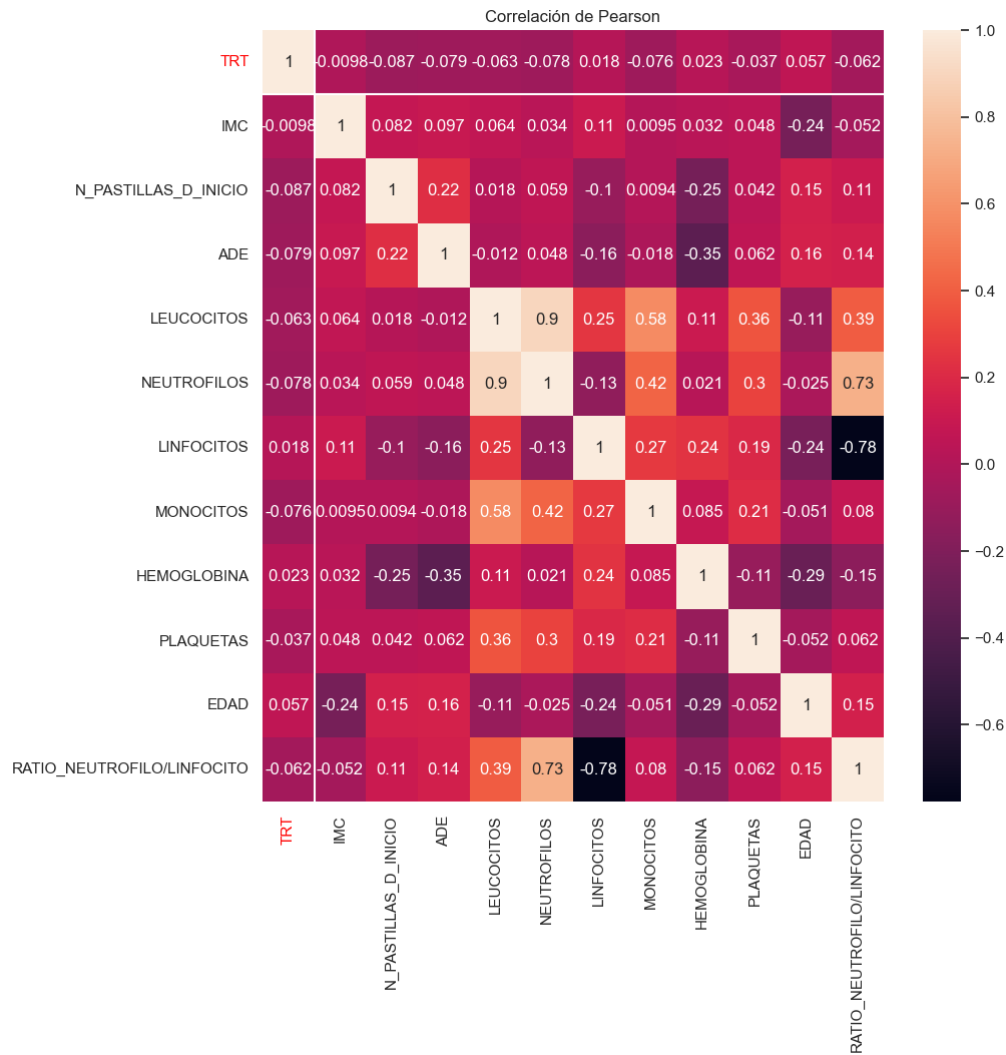


Figura 23. Mapa de calor de la correlación de *Pearson* con las variables numéricas y la variable objetivo TRT

Las correlaciones también desempeñaron un papel crucial en la identificación de multicolinealidad entre las variables. La multicolinealidad se refiere a la alta correlación entre dos o más variables predictoras. Este paso fue fundamental, ya que permitió tomar medidas correctivas para evitar sesgos y problemas de interpretación en el modelo predictivo. La acción tomada para abordar la multicolinealidad fue la eliminación de una de las variables altamente correlacionadas, concretamente la que menos relación tuviese con la variable a predecir (TRT).

En este sentido, se observó una fuerte correlación entre el Ratio neutrófilo/linfocito y las variables Neutrófilos (*Pearson* \rightarrow 0.73, *Spearman* \rightarrow 0.72) y Linfocitos (*Pearson* \rightarrow -0.78, *Spearman* \rightarrow -0.73), dado que el ratio es una variable derivada de estas dos. Además, se encontró una correlación significativa positiva entre las variables Leucocitos y Neutrófilos (*Pearson* \rightarrow 0.9, *Spearman* \rightarrow 0.9). Estas correlaciones se observan en la Figura 24. Posteriormente, analizamos la correlación de cada una de ellas con el TRT. La variable con mayor correlación fue la de Neutrófilos, por lo que se tuvo que eliminar tanto la variable Leucocitos como Ratio neutrófilo/linfocito.



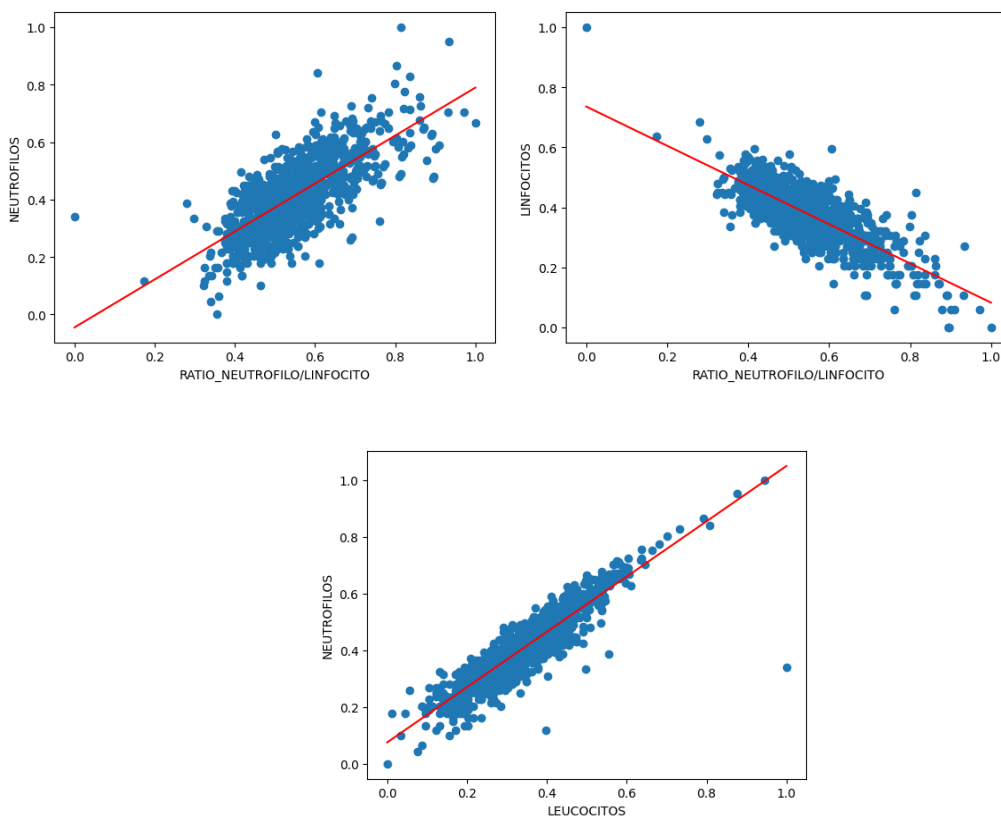


Figura 24. Grado de asociación entre las variables numéricas con alta correlación ($r > 0.7$ o < -0.7)

Además de calcular las medidas de correlación de *Spearman* y *Pearson*, se determinó el valor p para evaluar la significancia estadística de las correlaciones encontradas con el TRT, de las cuales se escogieron aquellas con un valor $p \leq 0.2$ (ADE, N° de pastillas, Monocitos, Neutrófilos y Edad). Cabe destacar que tanto **ADE** como **N° de pastillas**, **Monocitos** y **Neutrófilos** mostraron un valor de $p < 0.05$.

4.5.8. Análisis univariante de variables categóricas

Para las variables categóricas, se analizó la frecuencia o proporción de cada categoría. Si una variable categórica presentaba una categoría con una frecuencia menor al 5% del total de observaciones, se decidió eliminar dicha variable del conjunto de datos. Entre estas variables se incluyeron las siguientes: Enfermedad hepática, Sangrado mayor, Abuso de alcohol, Trombocitopenia, Demencia, Embolia sistémica, Sotalol, Bloqueadores de calcio, Verapamilo, Diltiazem, Neoplasia Activa, Digoxina y Doble antiagregación. Sus proporciones se reflejan en la Figura 25, en la cual se observa que en todas estas variables predomina la clase negativa, especialmente en la variable Sotalol, que contiene únicamente un 0.2% de positivos.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

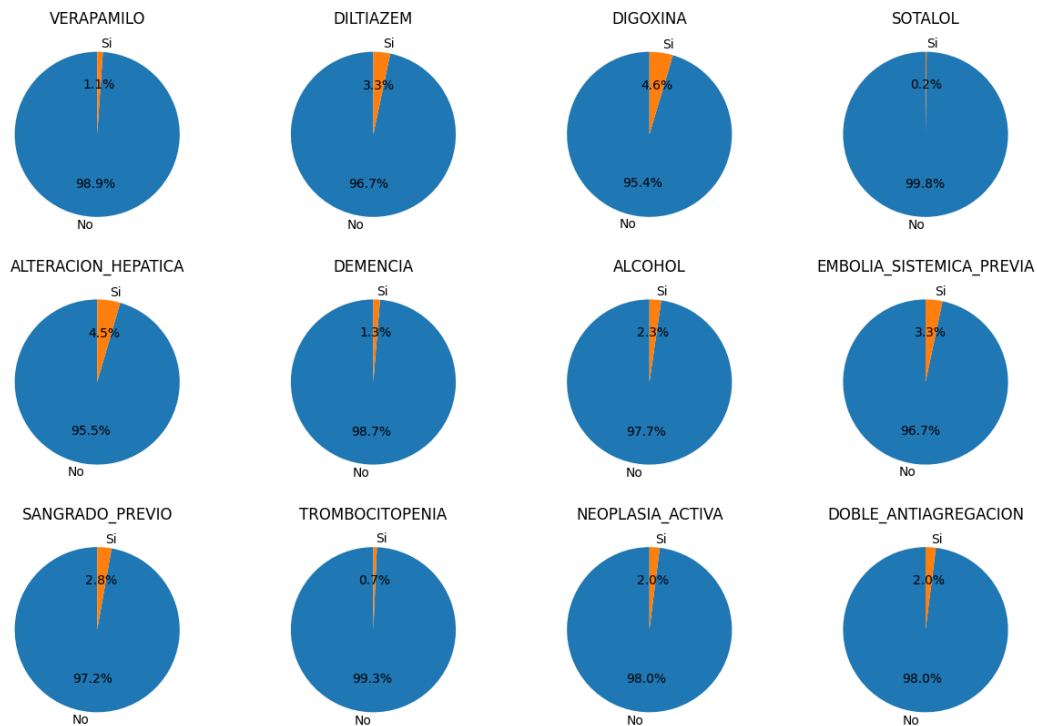


Figura 25. Frecuencia de las variables categóricas con baja representación en alguna de sus clases (<5%)

La decisión de eliminar estas variables categóricas (12 variables en total) se basó en el objetivo de mantener un equilibrio en la distribución de los datos y evitar un sesgo hacia una categoría específica. Asimismo, se buscó garantizar la calidad y la validez de los análisis y modelos posteriores, ya que variables con proporciones desequilibradas podían tener un impacto negativo en la capacidad predictiva y en la generalización del modelo.

En cambio, se decidió mantener otras variables, a pesar de presentar una cantidad significativamente mayor de valores "0" (negativos) en comparación con los valores "1" (positivos), debido a su relevancia y potencial valor predictivo en el análisis. La eliminación de todas las variables desequilibradas podría llevar a una pérdida de información importante y limitar la capacidad de generar resultados significativos.

4.5.9. Análisis bivalente entre TRT y variables categóricas

Se realizaron pruebas de normalidad y homogeneidad en los grupos definidos por cada variable categórica. Para verificar la normalidad de la distribución en cada grupo, se empleó la prueba de *Shapiro*. Esta prueba permitió determinar si los datos seguían una distribución normal o no para cada una de las categorías. Además, se llevó a cabo la prueba de homogeneidad de varianzas mediante el test de *Levene*. Esta prueba evaluó si las varianzas entre los grupos de cada variable categórica eran homogéneas. En el análisis realizado, se observó que ninguna de las variables categóricas cumplía ambas suposiciones de normalidad y homogeneidad de varianza.

Respecto a las variables binarias, estas presentaron una falta de normalidad y homogeneidad de varianza, se utilizó la prueba de *Mann-Whitney U* como una alternativa no paramétrica a la



prueba t de *Student*. Esta prueba permitió evaluar si había diferencias significativas en la distribución de la variable a predecir entre los grupos definidos por las variables categóricas.

Se estableció un umbral de 0.2 para el valor de p con la intención de capturar posibles relaciones sutiles o contextuales que podrían ser relevantes para el análisis. El valor p para cada una de estas variables binarias se muestra en la Tabla 13, donde se pueden observar las variables ordenadas de menor a mayor valor p, siendo Dislipemia la variable con menor valor p y Enfermedad vascular la que mayor p tenía.

Variable	Valor p	Variable	Valor p
Dislipemia	0.019	Neoplasia No	0.11
Amiodarona	0.07	Neoplasia Antecedentes	0.12
IC	0.087	Comorbilidades en SAME-TT ₂ R ₂ >2	0.15
Arteriopatía periférica	0.11	Enfermedad vascular	0.2

Tabla 13. Valor p de la prueba de *Mann-Whitney U* de las variables categóricas (vs TRT) con $p \leq 0.2$

Al fijarse en la Figura 26, se puede denotar la mayor diferencia entre medianas para las variables con un valor p más bajo (Dislipemia, Amiodarona e IC). En dicha figura se muestran las comparaciones entre ambas clases de cada variable binaria cuyo valor p fue ≤ 0.2 . Aunque se observa que el rango intercuartílico y la mediana de ambas clases en cada variable son similares, es importante destacar que la prueba de *Mann-Whitney U* puede dar resultados significativos incluso en presencia de similitudes en estas medidas de tendencia central. Esto se debe a que esta prueba se basa en la comparación de las distribuciones de los datos en su totalidad, teniendo en cuenta el orden de los valores y no solo las medidas de tendencia central. Donde más diferencias se encuentran en cuanto a las medianas son en las variables Dislipemia, Amiodarona e IC. Para IC y Amiodarona, se obtiene una mediana de TRT menor en la clase positiva (tomar amiodarona / padecer IC), mientras que la clase negativa se asemeja a la mediana general del TRT. En consideración a Dislipemia, la mediana del TRT es mayor al padecer de esta enfermedad.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

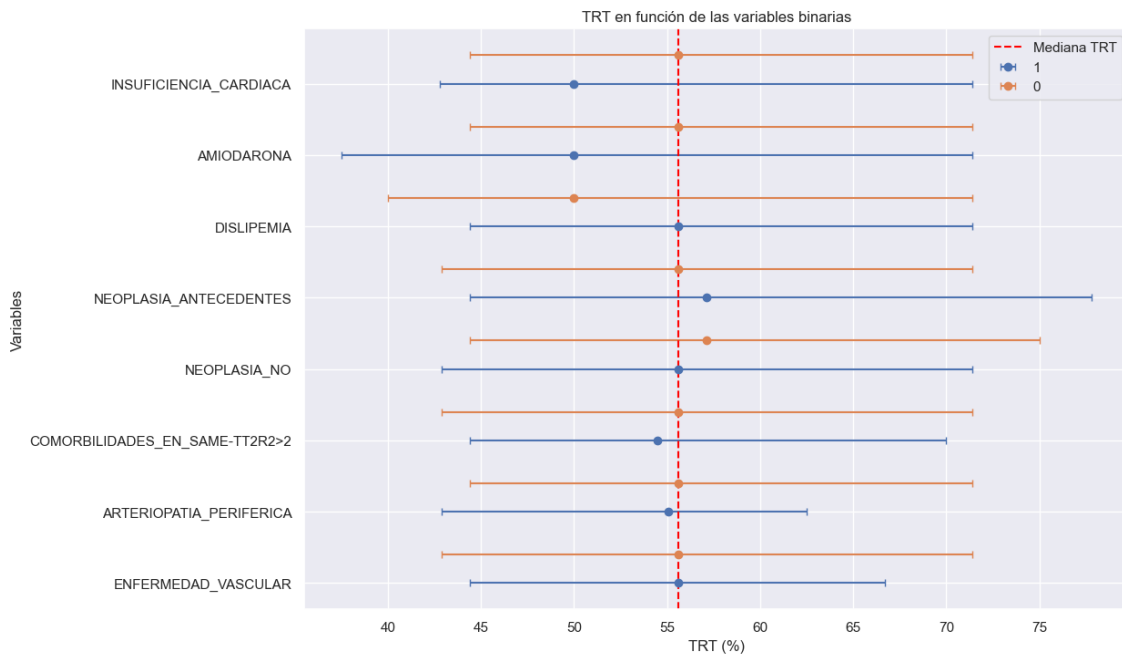


Figura 26. Comparación de medianas entre clases de las variables binarias con valor $p \leq 0.2$

El mismo proceso que se aplicó con las variables binarias, se empleó con las variables categóricas con más de dos clases. Ya que tampoco se cumplían las suposiciones de normalidad y homogeneidad en estas variables, se aplicó la prueba de *Kruskal-Wallis* como alternativa no paramétrica al ANOVA. En este caso, ninguna variable resultó significativa frente a la prueba de *Kruskal-Wallis* ($p > 0.05$).

Correlación entre variables categóricas

Tras realizar el análisis exploratorio de las variables categóricas, se procedió a examinar las correlaciones entre ellas. La correlación de *Pearson* entre variables categóricas puede ayudar a identificar patrones de comportamiento conjunto y determinar si algunas variables están altamente relacionadas entre sí. Esto puede ser especialmente útil para evitar la multicolinealidad en el modelo final y seleccionar las variables más relevantes.

Como se puede observar en la Tabla 14, hay varias variables con altas correlaciones entre sí, por lo que, para cada par de variables, se decidió eliminar la de mayor valor p en las pruebas de significancia con el TRT previamente realizadas. Estas fueron: Neoplasia Antecedentes ($p = 0.12$), Control frecuencia ($p = 0.56$), Cualquier antiagregante ($p = 0.87$), Bloqueadores de calcio ($p = 0.84$), Cualquier antiarrítmico ($p = 0.22$), Alteración hepática/renal ($p = 0.48$), Fuma ($p = 0.31$), Cardiopatía isquémica ($p = 0.5$), Tamaño de la aurícula izquierda ($p = 0.76$), Ictus previo ($p = 0.44$) y N° de comednicaciones ($p = 0.67$).



Variable 1	Variable 2	Correlación
Neoplasia No	Neoplasia Antecedentes	-0.9
Control frecuencia	Betabloqueantes	0.87
Aspirina	Cualquier antiagregante	0.9
Diltiazem	Bloqueadores de calcio	0.86
Amiodarona	Cualquier antiarrítmico	0.8
Enfermedad renal	Enfermedad hepática/renal	0.89
Fuma	SAMe-TT ₂ R ₂ > 3	0.76
Cardiopatía isquémica	Enfermedad vascular	0.85
Tamaño aurícula izquierda	Disfunción / Dilatación	0.93
Ictus previo	Ictus / Embolia	0.91
Nº de comednicaciones	Control frecuencia	0.8
Nº de comednicaciones	Betabloqueantes	0.72

Tabla 14. Correlación de *Pearson* de las variables categóricas con $r > 0.7$

4.5.10. Selección de variables

Después de realizar la preparación y comprensión de datos, se redujo el conjunto de características originales de 58 (incluyendo la variable SAMe-TT₂R₂ ≥ 3) a 33. De las 58 variables iniciales, se eliminaron 12 variables categóricas debido a una baja frecuencia en una de sus categorías (< 5%). Además, se excluyeron 11 variables categóricas y 2 numéricas que mostraban una alta correlación con otras variables existentes, lo que sugiere una redundancia en la información que aportaban.

Por lo tanto, de las 33 variables restantes, las seleccionadas con un valor $p \leq 0.2$ fueron 13, en las que se incluyen 5 variables numéricas (ADE, Monocitos, Neutrófilos, Edad y Número de pastillas) y 7 variables categóricas (Dislipemia, Amiodarona, IC, Arteriopatía periférica, Enfermedad vascular, Comorbilidades en SAMe-TT₂R₂ > 2 y Neoplasia No), además de la variable SAMe-TT₂R₂ ≥ 3 .

5. Entrenamiento y Test de los modelos predictivos

Una vez completado el análisis del TRT en relación a la puntuación del *score* SAMe-TT₂R₂ y la preparación de las otras variables predictoras, se llevó a cabo la etapa de entrenamiento y test de los modelos. Con el fin de evaluar el rendimiento de estos modelos, se dividió el conjunto de pacientes en dos particiones: 80% entrenamiento (800 pacientes) y 20% test (200

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

pacientes), cuya descriptiva de las características clínico-biológicas, analíticas y ecocardiográficas de los pacientes se encuentra detallada en el apartado 4.3.8.

El conjunto de entrenamiento se utilizó para ajustar los parámetros mediante *Grid Search* (*cross validation* = 5) y la estructura de los modelos utilizando algoritmos específicos como ensambles (*AdaBoost* y *Voting*), árboles de decisión, SVM, regresión, redes neuronales (MLP) y KNN de la librería *sklearn* tanto para el problema de regresión (variable objetivo: TRT continuo) como para el de clasificación (variable objetivo: TRT binario). Estos modelos fueron seleccionados debido a su capacidad para capturar patrones complejos y relaciones en los datos. En cuanto al *Voting*, éste se realizó con 3 modelos: SVC, regresión logística y MLP en clasificación; y *AdaBoost*, regresión lineal y KNN en regresión. Posteriormente, los modelos fueron testeados mediante el conjunto de prueba, cuyos resultados se muestran en los apartados 5.1 y 5.2.

Se tomó la decisión de entrenar y testear los modelos tanto con el conjunto completo de las 33 variables obtenidas tras el pretratamiento de los datos (Etapa 1), como con las 13 variables seleccionadas tras los análisis bivariantes con el TRT (Etapa 2). Esto se hizo con el objetivo de comparar el rendimiento de los modelos utilizando diferentes conjuntos de variables y evaluar si la selección de variables mejoraba significativamente las predicciones del TRT.

5.1. Modelos de clasificación binaria

5.1.1. Balanceo de datos

Para abordar el desequilibrio de clases, sobre todo con TRT <65% y TRT <70%, se utilizó el método de *Adaptive Synthetic Sampling* (ADASYN) para realizar un balanceo de los datos. ADASYN es un algoritmo de sobremuestreo que genera de manera adaptativa instancias sintéticas para las clases minoritaria, lo que significa que se presta especial atención a las instancias que son más difíciles de clasificar correctamente. Esto ayuda a mitigar el problema del sobreajuste en el modelo y mejora la capacidad de generalización, lo que puede conducir a mejores resultados.

5.1.2. Métricas de rendimiento

En cuanto al enfoque de clasificación, se utilizó un *scoring* de precisión, *recall*, *AUC ROC* y *f1 score* con *refit* = “*precision*”. Esto se hizo con el objetivo de ajustar el modelo para minimizar los errores de clasificación más costosos en el contexto específico del problema, es decir, minimizar los falsos positivos (considerando positivos los casos con TRT óptimo).

Durante la evaluación de los modelos, se utilizaron las siguientes métricas: *f1 score*, *recall*, precisión, *balanced accuracy* y *AUC ROC* (*Area Under the Receiver Operating Characteristic Curve*).

El *f1 score* es una métrica que combina precisión y *recall*, ofreciendo una medida equilibrada del rendimiento del modelo.



$$F1 \text{ Score} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

El *recall* (también conocido como sensibilidad) mide la proporción de verdaderos positivos clasificados correctamente respecto a todos los valores verdaderos positivos, mientras que la precisión (también conocida como valor predictivo positivo) mide la proporción de verdaderos positivos clasificados correctamente respecto a todos los valores clasificados como positivos.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

La *balanced accuracy* es una métrica que calcula el promedio de los valores de sensibilidad y especificidad.

$$\text{balanced-accuracy} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

Finalmente, se utilizó la métrica AUC ROC para evaluar el rendimiento global del modelo en términos de su capacidad para clasificar correctamente las observaciones de ambas clases. La curva ROC representa la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos a través de diferentes valores de *threshold*.

$$1 - \text{especificidad} = \frac{FP}{VN + FP}$$

5.1.3. Resultados con TRT < 60%

Al analizar los resultados de los modelos de clasificación con TRT < 60% (Tabla 15) y las 33 variables, observamos que, en general, todos los modelos presentaron un desempeño bajo en términos de métricas de evaluación. El modelo *AdaBoost* mostró el peor desempeño en la mayoría de métricas evaluadas, con un *f1 score* de 0.10, una precisión de 0.36 y un *recall* de 0.06. El modelo MLP también presentó un desempeño deficiente, con un *f1 score* de 0.19, una precisión de 0.39 y un *recall* de 0.13. Por otra parte, los mejores modelos en cuanto a estas métricas fueron KNN y árbol de decisión, aún por debajo de 0.5. En términos de AUC, se consiguieron los mejores resultados con el modelo de *AdaBoost*, a pesar de mostrar valores bajos en cuanto a *f1 score*, precisión y *recall*.

Por otro lado, al entrenar los modelos con las 13 variables seleccionadas, se vió un aumento en las medias de precisión y AUC. En este sentido, el mejor modelo fue el de *Voting* con unos valores de 0.62 para AUC y 0.75 en la precisión. Sin embargo, también se deben considerar

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

otras métricas como el *f1 score*, *balanced accuracy* y *recall*. Por lo que cabe destacar el modelo KNN, donde se encontró un AUC y *balanced accuracy* de 0.59 y precisión de 0.52, además de los valores más altos para *f1 score* y *recall* (0.54 y 0.56 respectivamente). Asimismo, la regresión logística también obtuvo valores más equilibrados al compararlo con el resto de modelos. Como peor modelo se destacó MLP, el cual carecía de *f1 score*, precisión y *recall*, y su AUC lo estableció como un modelo completamente aleatorio.

Modelo	F1	Precisión	Recall	Balanced accuracy	AUC
ETAPA 1 (33 variables)					
Media	0.30	0.42	0.26	0.50	0.51
AdaBoost	0.10	0.36	0.06	0.49	0.55
Regresión logística	0.28	0.43	0.21	0.50	0.51
SVC	0.39	0.41	0.36	0.49	0.51
MLP	0.19	0.39	0.13	0.49	0.52
Árbol de decisión	0.40	0.41	0.40	0.49	0.47
KNN	0.44	0.44	0.45	0.51	0.49
Voting	0.27	0.48	0.19	0.52	0.52
ETAPA 2 (13 variables)					
Media	0.26	0.47	0.23	0.52	0.56
AdaBoost	0.11	0.71	0.06	0.52	0.56
Regresión logística	0.38	0.56	0.28	0.56	0.61
SVC	0.36	0.36	0.35	0.45	0.53
MLP	0.00	0.00	0.00	0.50	0.49
Árbol de decisión	0.34	0.39	0.31	0.48	0.48
KNN	0.54	0.52	0.56	0.59	0.59
Voting	0.13	0.75	0.07	0.53	0.62

Tabla 15. Resultados de clasificación para los modelos con 33 y 13 variables con TRT < 60%

5.1.4. Resultados con TRT < 65%

En relación a los resultados con un punto de corte de 65% en la variable objetivo TRT (Tabla 16) e incluyendo las 33 variables post-tratamiento de datos, todos los modelos siguieron teniendo un desempeño modesto, sin ninguna diferencia significativa entre ellos, con un *f1 score* medio de 0.32, una precisión media de 0.30 y un *recall* medio de 0.34. Sin embargo, el *f1 score* y *recall* aumentó en promedio comparado con los modelos con TRT < 60%. La *balanced accuracy* promedio fue de 0.47 y el AUC promedio fue de 0.47.



Entre los modelos con las 13 variables seleccionadas, se observó una mejora en el rendimiento en comparación con la inclusión de las 33 variables. Los modelos SVC y MLP mostraron el mejor rendimiento en la mayoría de métricas, excepto para *recall*. Otros modelos como *Voting* y KNN también presentaron resultados destacables con valores por encima de la media, obteniendo este último el valor más alto para *recall* (0.6).

Modelo	F1	Precisión	Recall	Balanced accuracy	AUC
ETAPA 1 (33 variables)					
Media	0.32	0.30	0.34	0.47	0.47
AdaBoost	0.29	0.29	0.30	0.47	0.50
Regresión logística	0.30	0.27	0.34	0.44	0.44
SVC	0.35	0.35	0.34	0.51	0.47
MLP	0.27	0.28	0.27	0.46	0.48
Árbol de decisión	0.37	0.33	0.42	0.49	0.47
KNN	0.36	0.31	0.42	0.48	0.43
Voting	0.29	0.30	0.28	0.47	0.47
ETAPA 2 (13 variables)					
Media	0.42	0.37	0.48	0.54	0.55
AdaBoost	0.30	0.32	0.28	0.49	0.55
Regresión logística	0.42	0.35	0.51	0.52	0.53
SVC	0.49	0.42	0.58	0.59	0.60
MLP	0.49	0.46	0.52	0.61	0.60
Árbol de decisión	0.28	0.27	0.30	0.45	0.43
KNN	0.48	0.40	0.60	0.58	0.56
Voting	0.45	0.38	0.54	0.55	0.55

Tabla 16. Resultados de clasificación para los modelos con 33 y 13 variables con TRT < 65%

5.1.5. Resultados con TRT < 70%

Por último, analizando los resultados con un punto de corte de 70% (Tabla 17), los modelos se mantuvieron en la misma línea que con TRT < 65%. Éstos tuvieron un desempeño con un *f1 score* medio de 0.32 (0.26 - 0.39), una precisión media de 0.29 (0.25 - 0.35) y un *recall* medio de 0.37 (0.27 - 0.44). La *balanced accuracy* promedio fue de 0.5 (0.45 - 0.55) y el AUC promedio de 0.49 (0.40 - 0.55).

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

Entre los modelos con las 13 variables seleccionadas, se observó una mejora en el rendimiento. El modelo SVC mostró el mejor rendimiento con un *f1 score* de 0.44, una precisión de 0.36, un *recall* de 0.54, una *balanced accuracy* de 0.57 y un AUC de 0.56. Otros modelos como MLP y KNN también presentaron resultados remarcables en todas sus métricas, con valores por encima de la media.

Modelo	F1	Precisión	Recall	Balanced accuracy	AUC
ETAPA 1 (33 variables)					
Media	0.32	0.29	0.37	0.50	0.49
AdaBoost	0.26	0.25	0.27	0.47	0.49
Regresión logística	0.30	0.26	0.36	0.47	0.45
SVC	0.35	0.34	0.37	0.53	0.52
MLP	0.35	0.32	0.39	0.52	0.55
Árbol de decisión	0.39	0.35	0.44	0.55	0.53
KNN	0.31	0.25	0.41	0.45	0.40
Voting	0.31	0.30	0.32	0.50	0.52
ETAPA 2 (13 variables)					
Media	0.37	0.33	0.45	0.53	0.53
AdaBoost	0.27	0.33	0.24	0.52	0.56
Regresión logística	0.40	0.32	0.53	0.53	0.53
SVC	0.44	0.36	0.54	0.57	0.56
MLP	0.41	0.35	0.47	0.56	0.54
Árbol de decisión	0.28	0.27	0.29	0.48	0.48
KNN	0.42	0.34	0.54	0.55	0.52
Voting	0.39	0.32	0.51	0.52	0.52

Tabla 17. Resultados de clasificación para los modelos con 33 y 13 variables con TRT < 70%

5.1.6. Curvas ROC (*Receiver Operating Characteristic Curve*)

Dado que los valores de AUC, para todos los modelos, fueron más altos cuando se incluyeron solo las 13 variables seleccionadas, se procedió a analizar sus curvas ROC (Figura 27) en cada uno de los enfoques (TRT < 60%, TRT < 65% y TRT < 70%). De esta forma, se pudo corroborar gráficamente los resultados vistos en las tablas anteriores (Tabla 15, 16 y 17). Para TRT < 60% se obtuvo un mayor AUC con los modelos de regresión logística y *Voting*. Por otro lado, en cuanto a TRT < 65%, el mejor AUC viene dado por SVC y MLP, destacando también el



bajo rendimiento del árbol de decisión. Con $TRT < 70\%$, todas las curvas ROC se encuentran alrededor de 0.5, sobresaliendo en pequeña medida los modelos de *AdaBoost* y SVC.

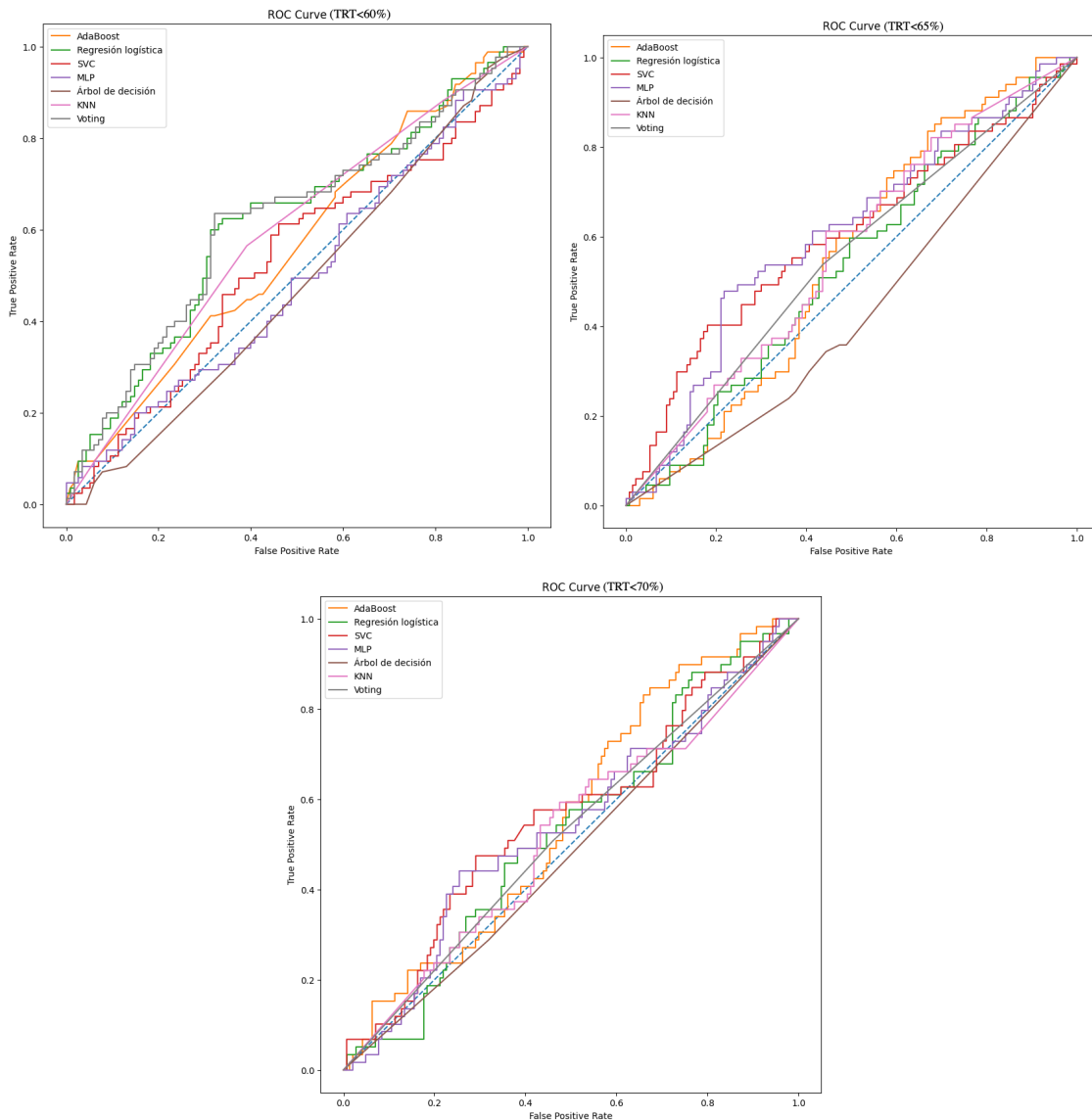


Figura 27. Curvas ROC para todos los modelos con puntos de corte de TRT: 60%, 65% y 70%

5.1.7. Selección de modelos de clasificación

Para la elección del mejor modelo, primero se analizaron las medias de cada una de las métricas en los distintos enfoques. Consideramos la precisión y la *balanced accuracy* nuestras principales métricas, al tener la necesidad de reducir los falsos positivos así como maximizar las tasas de verdaderos positivos de ambas clases. Por ello, nos decantamos por el punto de corte de 60% con las 13 variables seleccionadas, el cual mostró un mayor valor en la precisión (0.47), a pesar de una *balanced accuracy* similar al de los otros enfoques (0.52). Seguidamente, dentro de los resultados de $TRT < 60\%$, se buscó un equilibrio entre la precisión, *f1 score* y *recall*, además de AUC, por lo que se eligieron los modelos KNN y regresión logística como mejores modelos, aunque este último con valores de *f1 score* y *recall* aún por debajo de 0.5.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

Para estos dos modelos también se evaluó su calibración, que muestra la relación entre las probabilidades de predicción y las proporciones observadas de los eventos positivos (Figura 28). En dicha figura se percibe una calibración imperfecta en ambos modelos, ya que ésta va cambiando en diferentes rangos de probabilidad. Se puede observar que hasta cierto punto de probabilidad predicha (0.4), los modelos muestran una tendencia a subestimar las probabilidades reales de los eventos positivos. Después de ese punto de probabilidad predicha, los modelos muestran una tendencia a sobreestimar las probabilidades reales de los eventos positivos. No obstante, la regresión logística muestra más irregularidad en cuanto a la relación de las probabilidades predichas y la frecuencia de las observaciones.

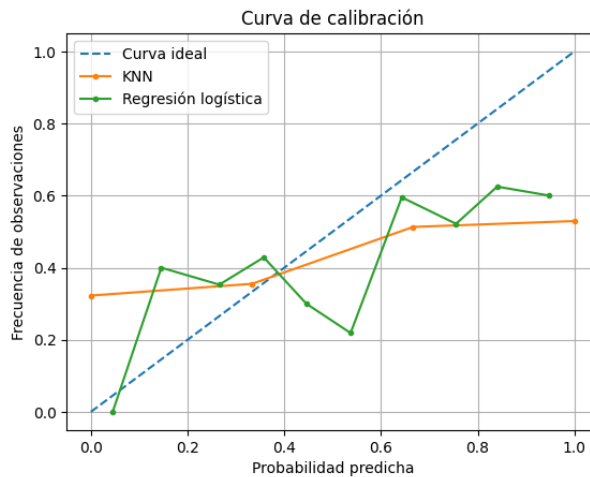


Figura 28. Curvas de calibración para KNN y regresión logística con TRT < 60%

Con el objetivo de maximizar la precisión y encontrar un equilibrio entre ésta y el *recall*, se examinó la curva de *precision-recall* de ambos modelos (Figura 29). Se pudo presenciar la alta precisión de los modelos con un umbral de *recall* casi nulo. Posteriormente, la precisión empeora drásticamente, a medida que el umbral de *recall* aumenta. Tras sobrepasar 0.2 de *recall*, el descenso de las curvas se estabiliza, encontrando el punto óptimo de equilibrio entre estas dos métricas en un *recall* de aproximadamente 0.58, mejorando el *recall* de ambos modelos. De este modo, la precisión del modelo KNN quedaría en 0.52 y 0.58 para la regresión logística, sin cambios representativos respecto a los valores originales.

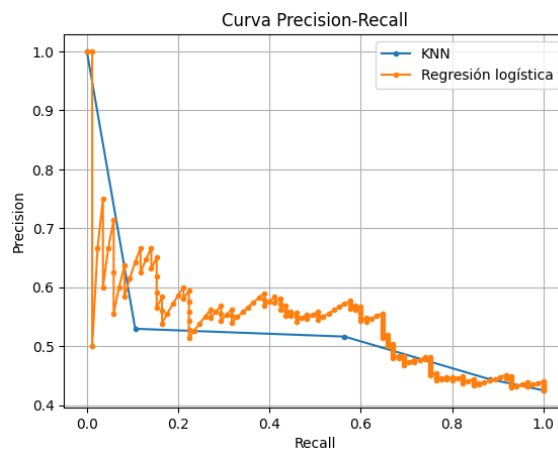


Figura 29. Curvas de *precision-recall* para KNN y regresión logística con TRT < 60%



5.1.8. Valor del *score* SAME-TT₂R₂ ≥ 3 para predecir TRT < 60%

Ajustando el modelo con una única variable (SAME-TT₂R₂ ≥ 3) y un punto de corte de TRT en 60% (óptimo), se obtuvo un valor de 0.5 tanto en AUC como en *balanced accuracy* (Tabla 18). De esta manera, se concluye que los modelos tienen un rendimiento similar al azar al clasificar las clases. Asimismo, excepto para *AdaBoost* y KNN, los valores de *f1 score*, precisión y *recall* fueron nulos, por lo que indica que el modelo no ha sido capaz de clasificar correctamente ningún caso positivo. En cambio, en *AdaBoost* y KNN la precisión permaneció por debajo de 0.5, cometiendo una cantidad considerable de errores al clasificar los casos positivos.

Adicionalmente, aunque en el caso de una sola variable algunos de los modelos pueden ser matemáticamente equivalentes, se optó por incluirlos en el estudio para poder comparar los resultados obtenidos con los experimentos que involucran varias variables de entrada.

Comparando estos resultados con los de los modelos entrenados a partir de las variables seleccionadas ($p \leq 0.2$), se concluyó que la inclusión de estas variables al SAME-TT₂R₂ ≥ 3 mejoraron el rendimiento de los modelos y, por lo tanto, la capacidad de discriminar entre ambas clases.

Modelo	F1	Precisión	Recall	Balanced accuracy	AUC
Media	0.14	0.12	0.19	0.50	0.50
AdaBoost	0.60	0.43	1.00	0.50	0.50
Regresión logística	0.00	0.00	0.00	0.50	0.50
SVC	0.00	0.00	0.00	0.50	0.50
MLP	0.00	0.00	0.00	0.50	0.50
Árbol de decisión	0.00	0.00	0.00	0.50	0.50
KNN	0.36	0.42	0.32	0.50	0.50
Voting	0.00	0.00	0.00	0.50	0.50

Tabla 18. Resultados de clasificación para los modelos con el *score* SAME-TT₂R₂ ≥ 3 y TRT < 60%

5.2. Modelos de regresión

5.2.1. Métricas de rendimiento

Durante la evaluación de los modelos de regresión, se utilizaron varias métricas para medir su rendimiento. Estas métricas incluyeron:

1. *Root Mean Squared Error* (RMSE): es una medida de la diferencia entre los valores reales y las predicciones del modelo. Se calcula como la raíz cuadrada de la media de los errores al cuadrado. El RMSE penaliza los errores más grandes de manera más significativa y proporciona una medida de la dispersión de los errores.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

2. *Mean Absolute Error* (MAE): es una medida del promedio de las diferencias absolutas entre los valores reales y las predicciones del modelo. El MAE es menos sensible a los valores atípicos que el RMSE y proporciona una medida del error promedio.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

3. *Mean Squared Error* (MSE): el MSE calcula el promedio de los errores al cuadrado y proporciona una medida de la variabilidad de los errores.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

4. Coeficiente de determinación (R^2): es una medida de cuánta variabilidad de la variable dependiente puede explicar el modelo. R^2 se calcula como la proporción de la varianza total de la variable dependiente que puede ser explicada por el modelo. Un valor de R^2 cercano a 1 indica un buen ajuste del modelo. Esta métrica será decisiva en la elección del mejor modelo.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

A parte, también se utilizó el R^2 ajustado, métrica que tiene en cuenta el número de variables predictoras en un modelo de regresión y proporciona una medida del porcentaje de variabilidad explicada por el modelo ajustada por el número de variables.

$$R_{ajust}^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$$

donde n = tamaño muestral y p = número de variables.

5.2.2. Valor predictivo del *score* SAME-TT₂R₂ ≥ 3

Entrenando los modelos solo con la variable SAME-TT₂R₂ ≥ 3 , única variable binaria relacionada con el SAME-TT₂R₂ con diferencias significativas con respecto a la variable objetivo TRT, se obtuvieron los resultados de la Tabla 19. Todos los modelos obtuvieron unas métricas de error elevadas, así como un valor negativo tanto en el R^2 como en el R^2 ajustado, lo que indica que el modelo de regresión no se ajustó bien a los datos y tuvo un rendimiento deficiente.



Modelo	MAE ± DS	RMSE	MSE	R²	R² ajustado
SVR	15.50 ± 19.09	19.09	364.41	-0.009	-0.01
Árbol de decisión	15.61 ± 19.02	19.02	361.88	-0.002	-0.007
KNN	15.98 ± 19.41	19.41	376.8	-0.04	-0.048
MLP	15.61 ± 19.02	19.02	361.87	-0.002	-0.007
Regresión lineal	15.61 ± 19.02	19.02	361.88	-0.002	-0.007
AdaBoost	15.63 ± 19.04	19.04	362.38	-0.003	-0.008
Voting	15.60 ± 19.02	19.02	361.76	-0.001	-0.006

Tabla 19. Resultados de regresión para los modelos con SAME-TT₂R₂ ≥ 3

5.2.3. Resultados incluyendo las 33 variables

En la Tabla 20, se observan los resultados obtenidos al utilizar todo el conjunto de variables post-tratamiento (33 en total). Estos fueron similares entre los modelos en términos de las métricas de error como el MAE, RMSE y MSE. Esto indica que las diferencias en la precisión de las predicciones fueron mínimas entre los modelos en esta configuración. Además, los valores de R² fueron muy bajos en todos los modelos, lo que muestra una dificultad para explicar la variabilidad en los datos de salida. El modelo *Voting* mostró el valor más alto de R², pero sigue siendo bajo en general, lo que sugiere una capacidad limitada para explicar los datos. Asimismo, el modelo *Voting* tuvo el menor RMSE y MSE, lo que indica una mayor precisión en las predicciones en comparación con otros modelos.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

Modelo	MAE ± DS	RMSE	MSE	R²	R² ajustado
SVR	15.48 ± 19.06	19.06	363.19	-0.005	-0.20
Árbol de decisión	15.83 ± 19.41	19.41	376.70	-0.043	-0.24
KNN	15.91 ± 19.59	19.59	383.86	-0.063	-0.27
MLP	15.5 ± 19.01	19.01	361.27	0	-0.19
Regresión lineal	15.3 ± 18.86	18.86	355.77	0.015	-0.17
AdaBoost	15.8 ± 18.84	18.84	354.98	0.017	-0.17
Voting	15.38 ± 18.83	18.83	353.71	0.019	-0.17

Tabla 20. Resultados de regresión para los modelos con 33 variables

5.2.4. Resultados con las 13 variables seleccionadas

Al utilizar únicamente las 13 variables seleccionadas ($p \leq 0.2$ en análisis bivariante), los resultados mostraron la misma tendencia (Tabla 21), aunque con una ligera disminución en los valores de todas las métricas estudiadas excepto para el R^2 , el cual aumentó; siendo el modelo de *Voting* capaz de capturar mayor variabilidad de los datos (5.7%) y menor MAE (15.05).

Modelo	MAE ± DS	RMSE	MSE	R²	R² ajustado
SVR	15.48 ± 19.04	19.04	362.66	-0.004	-0.068
Árbol de decisión	15.45 ± 18.84	18.84	355.09	0.017	-0.046
KNN	15.08 ± 18.63	18.63	347.14	0.039	-0.023
MLP	15.49 ± 19.01	19.01	361.29	0	-0.064
Regresión lineal	15.11 ± 18.54	18.54	343.73	0.049	-0.012
AdaBoost	15.31 ± 18.74	18.74	351.22	0.03	-0.035
Voting	15.05 ± 18.45	18.45	340.52	0.057	-0.003

Tabla 21. Resultados de regresión para los modelos con 13 variables

En resumen, tanto en la Tabla 20 como en la Tabla 21, los modelos de regresión muestran resultados similares en términos de las métricas de error, indicando una precisión de predicción similar. Sin embargo, todos los modelos presentaron un ajuste deficiente, como se evidencia por



los valores bajos de R^2 , lo que sugiere dificultades para explicar la variabilidad en los datos de salida. También cabe destacar los valores negativos del R^2 ajustado para los 7 modelos y ambos conjuntos de datos estudiados, lo que puede indicar la presencia de predictores irrelevantes en los modelos. Estos predictores no estarían aportando información adicional, ya que no están suficientemente relacionados con la variable respuesta. Por ello, se entrenaron los modelos utilizando únicamente aquellas variables con un valor p estadísticamente significativo ($p < 0.05$). Éstas fueron: SAME- $TT_2R_2 \geq 3$, N° pastillas diarias, ADE, Monocitos, Neutrófilos y Dislipemia (resultados en el apartado 5.2.5)

Al analizar los residuos de los diferentes modelos incluyendo las 13 variables seleccionadas (Figura 30), se observaron diferentes características en su tendencia central y dispersión. En términos de la mediana, el modelo SVR tuvo una mediana de 0.51, mientras que los modelos de árbol de decisión, MLP, regresión lineal, *AdaBoost* y *Voting* tuvieron valores medianos negativos, variando entre -2 y -1.7. Esto indica que estos modelos tuvieron tendencia a generar residuos negativos. Por otro lado, el modelo KNN mostró una mediana de -2.02, sugiriendo una tendencia aún más negativa en sus residuos.

La dispersión de los residuos fue observada a través del RIQ. En este caso, el modelo de árbol de decisión presentó el RIQ más alto de 30.79, lo que indica una mayor variabilidad en los residuos generados, y KNN con el RIQ más bajo (27.74). El resto modelos mostraron una dispersión similar en sus residuos, con RIQs en el rango de 28.02 a 29.58.

También se pueden percibir varios valores atípicos de los residuos correspondientes a los modelos SVR (-55.24), KNN (-58.71) y MLP (-57.05), los cuales fueron analizados y se detectó que pertenecían al mismo paciente, el cual tenía un TRT muy por debajo de la media y fuera del RIQ.

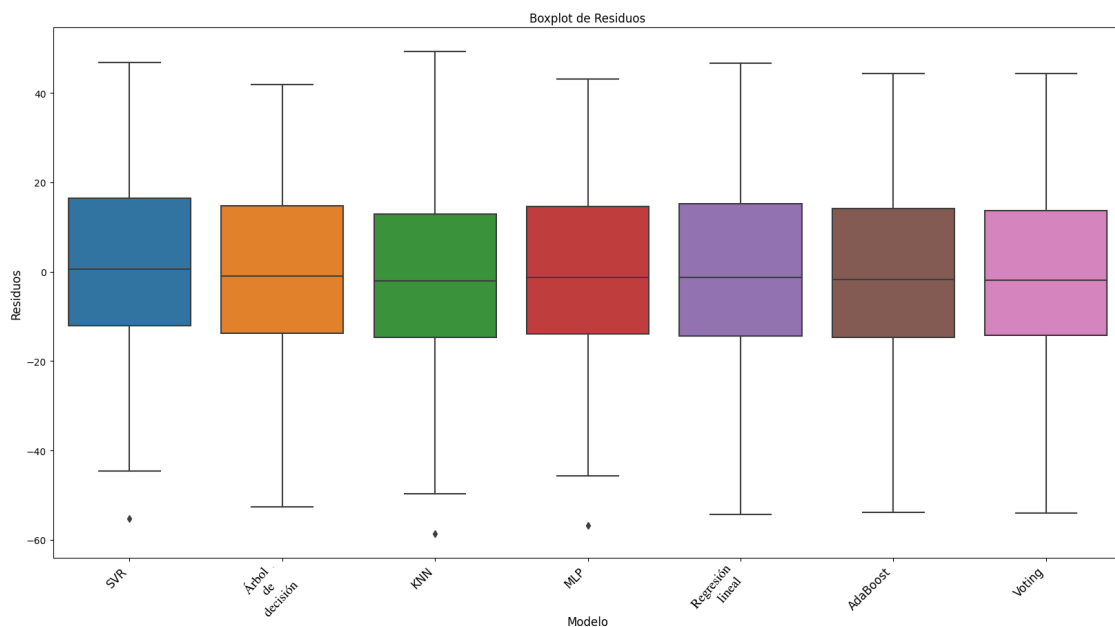


Figura 30. Comparación de residuos de los modelos de regresión con 13 variables

También se llevaron a cabo pruebas de normalidad y homogeneidad de los residuos. Estas pruebas permitieron evaluar si los residuos seguían una distribución normal y si las varianzas

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

eran similares entre los modelos. Los resultados indicaron que los residuos de todos los modelos cumplían con el supuesto de normalidad ($p > 0.05$) y que las varianzas eran homogéneas ($p > 0.05$).

Dadas estas suposiciones, se realizó un análisis de varianza (ANOVA) para comparar los modelos en términos de sus rendimientos. Los resultados del ANOVA revelaron un valor de p de 0.87, lo cual indica que no hay diferencias significativas entre los modelos en cuanto a sus efectos en la variable objetivo.

5.2.5. Resultados con las 6 variables significativas

Los resultados incluyendo $SAME-TT_2R_2 \geq 3$ y las otras 5 variables significativas: N° pastillas diarias, ADE, Monocitos, Neutrófilos y Dislipemia (Tabla 22), mostraron mejoría sobre todo en los modelos MLP y regresión lineal en cuanto a variabilidad explicada de los datos (únicos modelos con valores positivos para R^2 y R^2 ajustado). No obstante, en consideración al MAE, RMSE y MSE, estos valores no presentaron cambios destacables con los resultados previos.

Modelo	MAE \pm DS	RMSE	MSE	R^2	R^2 ajustado
SVR	15.34 \pm 18.91	18.91	357.74	0.01	-0.021
Árbol de decisión	15.61 \pm 19.02	19.02	361.88	-0.002	-0.032
KNN	16.21 \pm 19.95	19.95	398.2	-0.1	-0.14
MLP	15.25 \pm 18.67	18.67	348.63	0.035	0.005
Regresión lineal	15.21 \pm 18.62	18.62	346.67	0.04	0.01
AdaBoost	15.48 \pm 18.89	18.89	356.92	0.012	-0.02
Voting	15.40 \pm 18.81	18.81	353.89	0.02	-0.01

Tabla 22. Resultados de regresión para los modelos con 6 variables

Al examinar los residuos (Figura 31), se observaron similitudes en comparación con los residuos de los modelos empleando 13 variables. En relación a la mediana, el modelo SVR exhibió una mediana de 0.56, mientras que los modelos de árbol de decisión, MLP, regresión lineal, *AdaBoost* y *Voting* mostraron medianas negativas, oscilando entre -2.25 y -1.41. Esto sugiere que estos modelos tendieron a generar residuos negativos, especialmente los modelos *AdaBoost* y KNN, con medianas inferiores a -2.

La dispersión de los residuos se evaluó mediante el RIQ. En este caso, el modelo de árbol de decisión exhibió una mayor variabilidad en los residuos (RIQ = 31.4), mientras que MLP registró el RIQ más bajo (27.89). Los modelos restantes presentaron RIQs en el rango de 28.68 a 29.96. Cabe denotar la ausencia de valores atípicos en los residuos de todos los modelos.



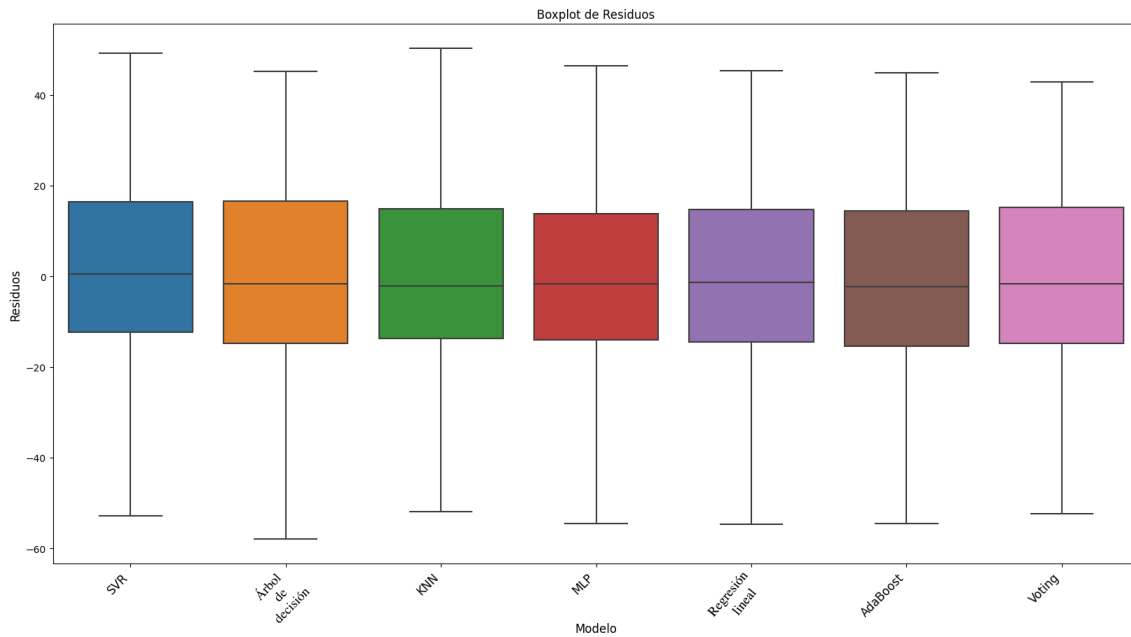


Figura 31. Comparación de residuos de los modelos de regresión con 6 variables

En cuanto a las pruebas de normalidad y homogeneidad de los residuos, los resultados indicaron que éstos cumplieran con el supuesto de normalidad y que las varianzas eran homogéneas en todos los modelos. Además, los resultados del ANOVA revelaron que no había diferencias significativas entre ellos en términos de su impacto en la variable objetivo ($p = 0.89$).

5.2.6. Selección de modelos de regresión

Después de evaluar todos los modelos incluyendo únicamente las 6 variables significativas, se determinó que MLP y regresión lineal fueron los más adecuados para este estudio. Esta elección se basó en considerar el valor positivo del R^2 ajustado. Además, se tuvo en cuenta el MAE y el R^2 , que también respaldaron la elección de estos modelos como los más efectivos.

Al comparar los resultados de nuestros modelos utilizando un conjunto de 6 variables con los modelos que utilizan únicamente el $SAME-TT_2R_2 \geq 3$, no se observó una mejora significativa en términos de MAE, RMSE y MSE, cuyos rangos fueron de 15.21 - 15.98, 18.62 - 19.95 y 346.67 - 398.2 respectivamente. Sin embargo, vale la pena destacar que dos modelos en particular, MLP y regresión lineal, obtuvieron valores positivos para R^2 y R^2 ajustado, lo cual indica que la inclusión de estas 5 variables, al $SAME-TT_2R_2 \geq 3$, ayudaron a explicar una mayor parte de la variabilidad de los datos.

5.3. Tiempos de ejecución

Al analizar los tiempos de ejecución de test (en segundos) de los modelos en el problema de clasificación (Tabla 23), calculados mediante la media de $TRT < 60\%$, $TRT < 65\%$ y $TRT < 70\%$, se observó que el uso de un conjunto reducido de variables no necesariamente conlleva una disminución en el tiempo de ejecución. En particular, los modelos *AdaBoost* y *SVC* destacaron como los más costosos en términos de tiempo de ejecución. Estos modelos suelen requerir un mayor poder computacional debido a su complejidad y la cantidad de cálculos involucrados en

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

su predicción. Por otro lado, los modelos KNN, regresión logística y árbol de decisión tendieron a ser más rápidos en comparación. Estos modelos, en general, tienen una estructura más simple y computacionalmente eficiente, lo que se traduce en tiempos de ejecución más cortos. Es importante tener en cuenta que el tiempo de ejecución de un modelo puede depender de diversos factores, como el número de parámetros utilizados. En este contexto, MLP fue el modelo con más parámetros libres y regresión logística la que menos.

Modelo	Nº parámetros libres	Tiempo de ejecución (33 variables)	Tiempo de ejecución (13 variables)
SVC	3	0.027	0.021
Árbol de decisión	3	0.0013	0.004
KNN	3	0.0057	0.0083
MLP	4	0.0036	0.0037
Regresión logística	0	0.001	0.0017
AdaBoost	3	0.035	0.055
Voting	2	0.016	0.015

Tabla 23. Comparación del nº de parámetros libres y tiempos de ejecución (s) para ambas etapas y todos los modelos (clasificación)

En el problema de regresión (Tabla 24), se observaron las mismas tendencias que en el problema de clasificación en cuanto a los tiempos de ejecución y la disminución de variables. En este caso, el modelo *Voting* con 33 variables destacó como el más costoso en términos de tiempo de ejecución, ya que se construyó a partir de la combinación de 3 modelos. Por otro lado, los modelos de regresión lineal, árbol de decisión y MLP fueron los que menos tiempo necesitaron para realizar las predicciones, teniendo en cuenta la falta de parámetros libres en la regresión lineal. Asimismo, al reducir la cantidad de variables utilizadas en los modelos de regresión, no se observó una mejora destacable en los tiempos de ejecución.



Modelo	Nº parámetros libres	Tiempo de ejecución (33 variables)	Tiempo de ejecución (13 variables)	Tiempo de ejecución (6 variables)
SVR	3	0.01	0.007	0.009
Árbol de decisión	3	0.002	0.002	0.001
KNN	3	0.015	0.007	0.005
MLP	4	0.002	0.002	0.002
Regresión lineal	0	0.001	0.001	0.001
AdaBoost	3	0.018	0.04	0.01
Voting	1	0.066	0.02	0.02

Tabla 24. Comparación del nº de parámetros libres y tiempos de ejecución (s) para los 3 conjuntos de variables y todos los modelos (regresión)

6. Validación de los modelos predictivos óptimos

Para validar los modelos seleccionados previamente (Regresión: MLP y regresión lineal, Clasificación: KNN y regresión logística), se utilizó un conjunto de validación independiente a los de *training* y test compuesto por 50 pacientes que no formaban parte del conjunto de datos original. Este conjunto de validación proporcionó una evaluación imparcial y objetiva del rendimiento de los modelos finales.

6.1. Modelos de clasificación

Los modelos elegidos para la validación en el problema de clasificación (TRT<60%) fueron KNN (parámetros: *algorithm* = 'auto', *n_neighbors* = 3, *weights* = 'uniform') y regresión logística. Esta selección ha sido justificada previamente basándose en su mayor nivel de precisión, *balanced accuracy* y AUC, así como en su equilibrio entre *f1 score* y *recall*.

La regresión logística solo predijo la clase positiva o TRT óptimo (Figura 32), obteniendo un *recall* de 1 y un *f1 score* de 0.65, pero una precisión y *balanced accuracy* bajas (0.48 y 0.5 respectivamente) para dicha clase. Aunque se consiguió un AUC de 0.59, debido a la alta tasa de verdaderos positivos, los valores de las métricas para la clase negativa son nulas, lo que hace de este modelo un rendimiento deficiente.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

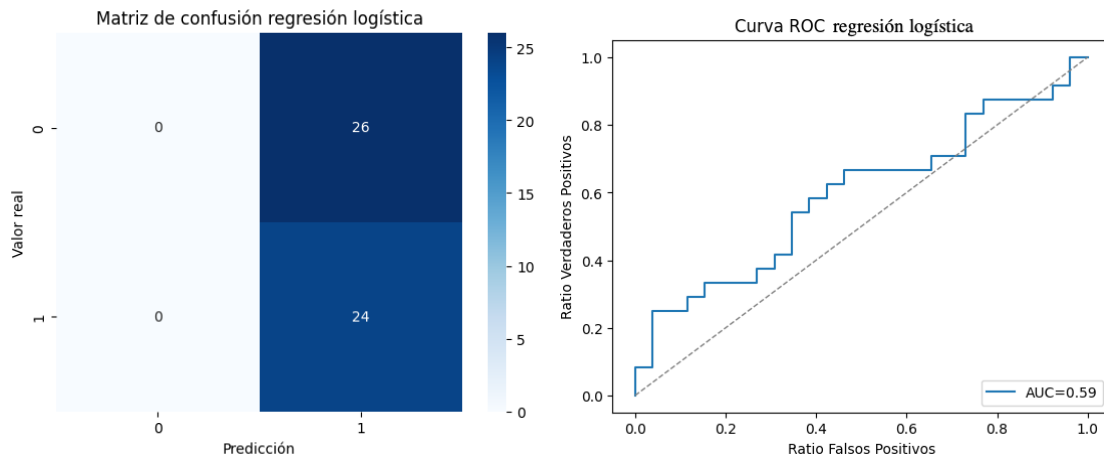


Figura 32. Matriz de confusión y curva ROC para regresión logística y TRT < 60%

En cuanto a los resultados del KNN, se obtuvo un modelo más equilibrado con un AUC (0.57) y *f1 score* (0.61) menor que con la regresión logística, pero unos valores de precisión y *recall* por encima de 0.60 para ambas clases. De esta manera, este modelo discriminó mucho mejor las clases, mejorando así la *balanced accuracy* (0.62). Esto se puede distinguir en la matriz de confusión y curva ROC de la Figura 33.

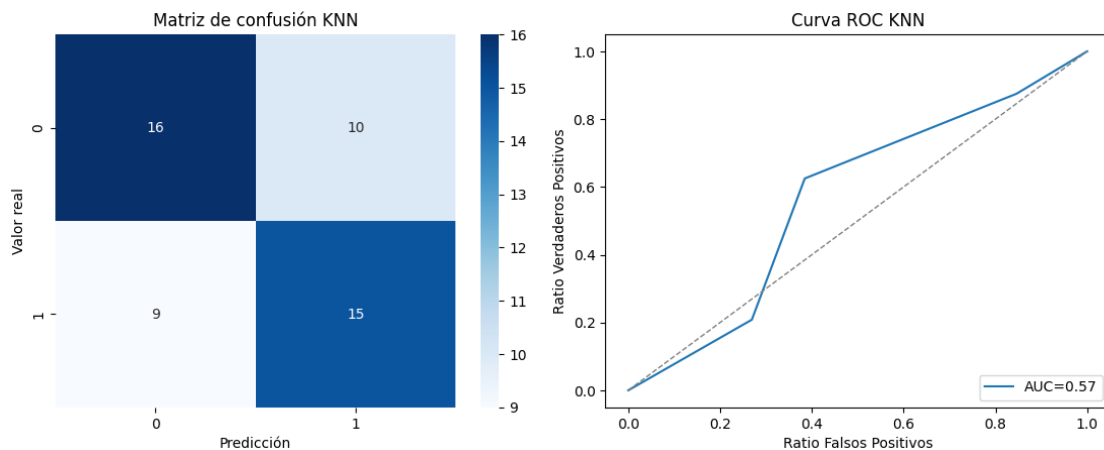


Figura 33. Matriz de confusión y curva ROC para KNN y TRT < 60%

6.2. Modelos de regresión

Los modelos elegidos para la validación en el problema de regresión fueron MLP con los siguientes parámetros: *'activation': 'relu', 'hidden_layer_sizes': (100,)*, *'learning_rate': 'constant', 'solver': 'sgd'*, y la regresión lineal, que mostró el mejor rendimiento en términos de R^2 ajustado, R^2 y MAE en comparación con los otros modelos utilizando las 6 variables significativas. Se aplicaron los modelos, ya previamente entrenados, a los datos del conjunto de validación y se calcularon las métricas de rendimiento correspondientes (Tabla 24).



Modelo	MAE ± DS	RMSE	MSE	R²	R² ajustado
MLP	21.03 ± 14.61	25.61	655.76	-0.35	-0.54
Regresión lineal	21.58 ± 14.67	26.09	680.55	-0.41	-0.60

Tabla 24. Resultados de regresión para MLP y regresión lineal con 6 variables

Al realizar la validación de estos modelos, se observó que los resultados fueron peores en comparación con el conjunto de prueba. En este caso, todas las métricas de error tuvieron valores mucho más altos, con un R² y R² ajustado por debajo de 0, lo cual significa que los modelos no fueron capaces de detectar la variabilidad de los datos. Esto puede deberse a la limitada cantidad de datos en el conjunto de validación, lo cual puede conducir a una estimación menos precisa del rendimiento de los modelos.

Al estudiar los residuos de ambos modelos (Figura 34 y Figura 39), se pudo observar que los modelos no predecían correctamente. Mediante los gráficos de valores reales vs predichos, se percibe que no existe ninguna relación lineal entre ambas, ya que todas las predicciones se encuentran alrededor de una línea diagonal en torno a 40%. En los gráficos de residuo vs id se ve que la mayoría de errores no se encuentran cercanos a 0, es más, casi todos estos están por encima de 0, lo que indica que los valores predichos tienden a ser menores que los valores reales. Esto también se puede identificar en los gráficos de distribución de los residuos, donde parecen estar centrados en 10-20. Con los gráficos Q-Q (*Quantile-Quantile*) se presencia que los puntos siguen aproximadamente una línea diagonal, indicando que los residuos se ajustan a una distribución normal. Para finalizar, el mal ajuste de los modelos se puede corroborar mediante los gráficos de residuos vs predicción, donde se contempla la evidencia de subestimación por parte del modelo.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitaminas K

Diagnóstico residuos MLP

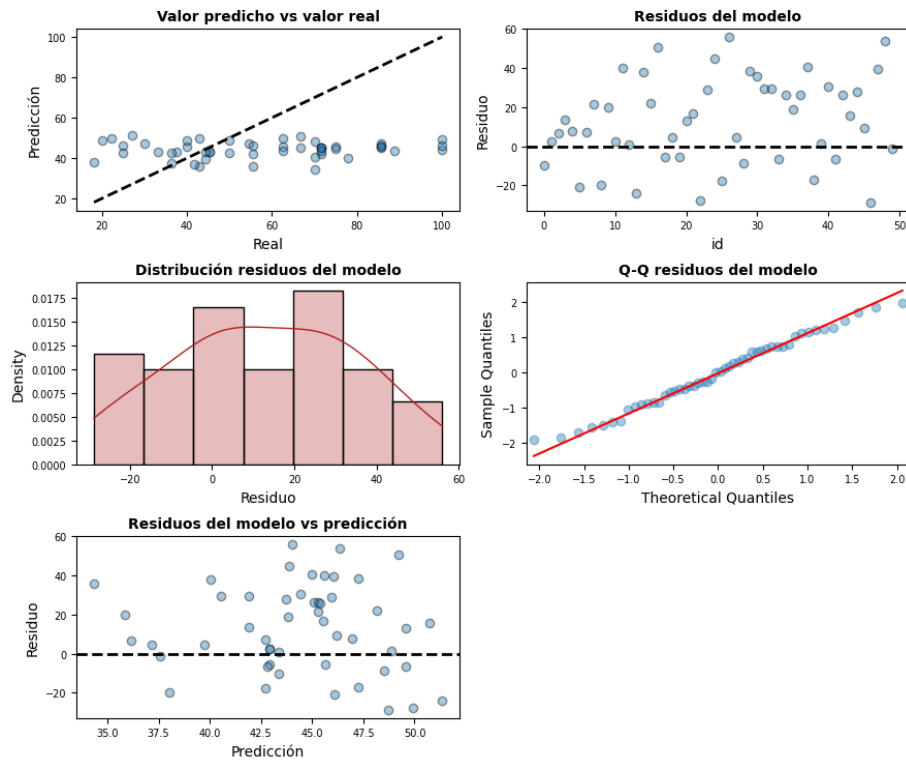


Figura 34. Análisis de residuos para MLP con 6 variables

Diagnóstico residuos regresión lineal

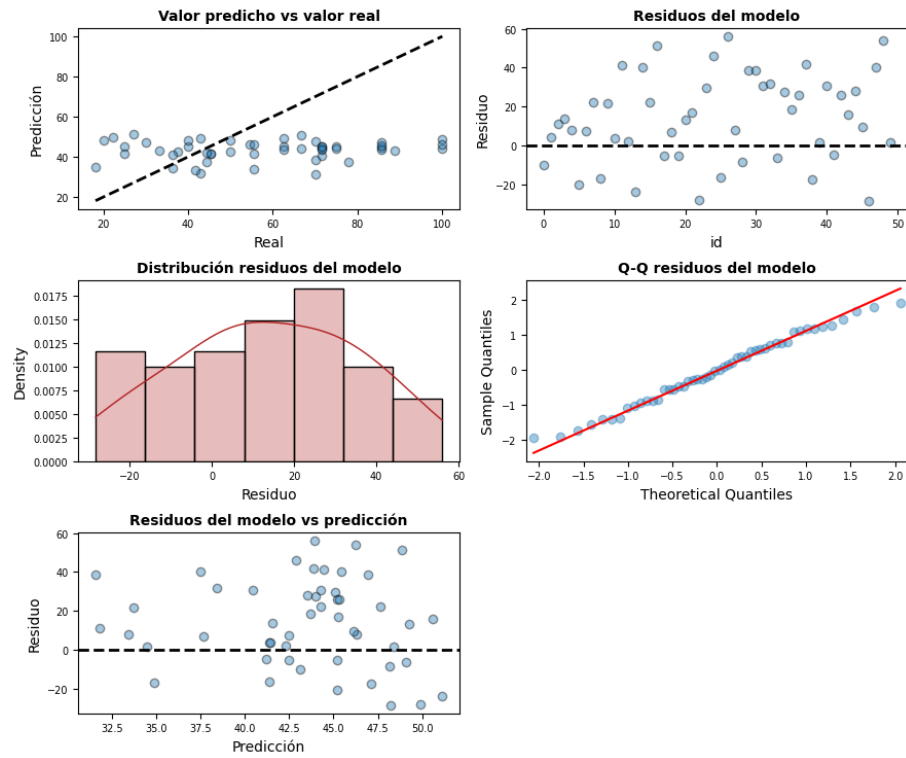


Figura 35. Análisis de residuos para regresión lineal con 6 variables



7. Conclusiones

Respecto al problema planteado, la IA podría optimizar el manejo del paciente con anticoagulación en la consulta de fibrilación auricular. Un modelo predictivo de TRT sería una herramienta efectiva para predecir la calidad de la anticoagulación siendo de gran utilidad para médicos de familia, neurólogos, cardiólogos y hematólogos. En este sentido, la IA nos podría ayudar a analizar todas las variables al diagnóstico y factores de riesgo de TRT subóptimo para poder ofrecer a los pacientes una medicina de precisión en el abordaje de la anticoagulación de la FANV, aplicando el tratamiento más óptimo en cada caso. Para ello se necesitaría de una coordinación estrecha, estable y constante entre conocimientos médicos y de ciencia de datos, muestra de ello es que yo he tenido el problema de comprensión de datos médicos para una adecuada recogida y tratamiento, requiriendo el apoyo constante del tutor en la empresa (CHGUV).

Como principales conclusiones del estudio destacar:

- El *score* SAME-TT₂R₂ no ha sido un factor predictor del TRT en la muestra poblacional retrospectiva del CHGUV, contando con 1000 casos de FANV anticoagulados con acenocumarol. Esta información es muy relevante, ya que pone en evidencia problemas de generalización en el uso de este *score*.
- La inclusión de variables adicionales y métodos más sofisticados de modelado de los datos ha permitido mejorar la capacidad predictiva del *score* SAME-TT₂R₂, aunque los valores de rendimiento obtenidos siguen siendo bajos para conllevar un impacto en el manejo de este tipo de pacientes.
- Para obtener mejoras significativas se debería evaluar la inclusión de variables más relevantes para la predicción del TRT, así como la falta de homogeneidad en la implementación de los diagnósticos clínicos utilizados habitualmente en la redacción de la historia clínica electrónica en la atención sanitaria del día a día.

Por lo que en el momento actual según lo ya existente y los resultados de este TFG, se debería apoyar la restricción terapéutica con ACOD a ciertos subgrupos de pacientes (de difícil control del INR o con ciertas situaciones clínicas) y se recomienda a los profesionales seguir introduciendo tratamiento anticoagulante con AVK durante 6 meses siguientes al primer mes de ajuste de dosis y tras este periodo, en función del TRT, decidir mantener AVK o cambio a ACOD si TRT subóptimo (TRT <60-65%), tal y como indica el IPT UT_ACOD/VS/21112016 de la agencia española de medicamento y productos sanitarios [5]. También se debería apoyar la posibilidad de buscar formas de mejorar el control del INR para incrementar la efectividad del acenocumarol a menor coste (educación de los pacientes, automanejo, etc.).

En conclusión, los resultados han evidenciado las limitaciones del *score* SAME-TT₂R₂ y que los modelos de ML pueden ayudar a mejorar la calidad de las predicciones si se les alimenta con las variables correctas. Por ello, espero poder continuar mi formación en el ámbito de ML con el desarrollo de las ideas expuestas en el apartado de trabajos futuros (apartado 8) dentro del ámbito de la FANV. Además, es evidente que en la próxima década, gran parte de los sistemas médicos dependerán cada vez más de las herramientas de apoyo a las decisiones basados en el aprendizaje automático, lo que aumentará la seguridad y eficacia de los tratamientos. No

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

obstante, en ningún caso, estas herramientas sustituirán al médico, que seguirá siendo quien tome las decisiones finales.

7.1. Legado

El legado del proyecto radica en su contribución al avance del campo de la predicción del tiempo en rango terapéutico en el tratamiento de anticoagulante. A través de esta investigación, se ha logrado obtener conocimientos valiosos sobre las limitaciones y desafíos asociados a este problema específico.

En primer lugar, los resultados permiten identificar áreas de mejora y enfocar los esfuerzos futuros en aspectos específicos. Estos resultados indican la necesidad de investigar y comprender mejor los factores subyacentes que afectan el tiempo en rango terapéutico en pacientes anticoagulados. Esto puede incluir la exploración de nuevas variables predictoras, la consideración de diferentes enfoques metodológicos o la realización de estudios más amplios y diversificados.

Es importante destacar que los resultados sirven como un recordatorio de la complejidad de los problemas médicos y la necesidad de un enfoque multidisciplinario y continuo en la investigación.

En conclusión, el proyecto deja un legado valioso en términos de aprendizaje, identificación de áreas de mejora y fomento de la investigación continua en la predicción del tiempo en rango terapéutico. A través de este legado, se espera que futuros estudios y proyectos se beneficien de las lecciones aprendidas y logren avances más significativos en la mejora del tratamiento anticoagulante y el manejo del paciente en la consulta de FA.

En el siguiente repositorio de GitHub ([TFG-GCD](#)), se encuentra disponible todo el código utilizado en este proyecto de investigación para replicar los resultados obtenidos. Sin embargo, por razones de seguridad y protección de datos, no es posible compartir directamente el conjunto de datos utilizado en este estudio. Para acceder a los datos, es necesario ponerse en contacto con el CHGUV, quienes poseen la autoridad y los permisos necesarios para su distribución. A través de su sitio web o por medio de información de contacto proporcionada por el consorcio, los interesados pueden obtener acceso a los datos necesarios para llevar a cabo el análisis y replicar los resultados presentados en este trabajo.

7.2. Relación del trabajo desarrollado con los estudios cursados

Durante mi formación en ciencia de datos, he adquirido conocimientos y habilidades en el manejo y análisis de grandes volúmenes de datos, así como en la aplicación de aproximaciones y algoritmos de aprendizaje automático. Estas competencias son fundamentales para abordar los objetivos de este proyecto.

Además, mi formación me ha proporcionado los conocimientos necesarios para realizar un análisis exploratorio de datos de manera eficiente y aplicar técnicas de preprocesamiento y limpieza de datos. Esto es esencial para garantizar la calidad de los datos, y obtener resultados precisos y confiables.

Asimismo, esta capacitación me ha brindado una comprensión profunda de los modelos predictivos y algoritmos de aprendizaje automático. Esto me ha permitido investigar y comparar



diferentes enfoques para predecir el tiempo en rango terapéutico subóptimo, utilizando variables candidatas y evaluando la eficacia de cada modelo.

En general, las 13 competencias transversales (Comprensión e integración de conocimientos, Aplicación y pensamiento práctico, Análisis y resolución de problemas, Innovación, creatividad y emprendeduría, Diseño y proyecto, Trabajo en equipo y liderazgo, Responsabilidad ética, medioambiental y profesional, Comunicación efectiva, Pensamiento crítico, Conocimiento de problemas contemporáneos, Aprendizaje permanente, Planificación y gestión del tiempo, Instrumentación específica) han sido importantes durante el transcurso del proyecto. Sin embargo, destacan 6 que podría considerarse especialmente relevantes:

1. Comprensión e integración de conocimientos: Esta competencia es esencial para comprender los fundamentos teóricos y prácticos del campo de estudio, así como para integrar diferentes áreas de conocimiento en la investigación.
2. Análisis y resolución de problemas: El proyecto requerirá un enfoque analítico y la capacidad de identificar y abordar desafíos y problemas en el proceso de análisis de datos y desarrollo de modelos predictivos.
3. Innovación, creatividad y emprendeduría: La aplicación de técnicas avanzadas de análisis de datos y modelos predictivos implica la necesidad de ser innovador, creativo y estar abierto a nuevas ideas y enfoques.
4. Trabajo en equipo y liderazgo: El trabajo en equipo es esencial para colaborar con otros investigadores, profesionales de la salud y científicos de datos. Además, la capacidad de liderazgo será importante para coordinar y dirigir el proyecto de manera efectiva.
5. Comunicación efectiva: La comunicación clara y efectiva será fundamental para transmitir los hallazgos, resultados y recomendaciones del proyecto tanto a audiencias especializadas como a profesionales de la salud y público en general.
6. Pensamiento crítico: La capacidad de analizar de manera crítica los resultados, evaluar las limitaciones y tomar decisiones fundamentadas es esencial en todo el proceso de investigación y desarrollo del modelo de predicción.

8. Trabajos futuros

Una de las limitaciones de nuestro estudio podría ser el número de pacientes. Además, se trata de un estudio unicéntrico, por ello resultaría útil futuros estudios que incluyeran más centros y mayor número de pacientes. Esto permitiría obtener resultados más robustos y generalizables, así como explorar posibles subgrupos de pacientes que podrían tener diferentes características y respuestas al tratamiento anticoagulante.

Se podría considerar la inclusión de otros tipos de variables como por ejemplo parámetros sociodemográficos (analfabetismo funcional y situación ocupacional) e información farmacogenética. Estas variables podrían aportar información adicional sobre la respuesta al tratamiento anticoagulante para así optimizar la capacidad predictiva del modelo.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

Sería interesante probar otros modelos de predicción. Se podrían explorar algoritmos de aprendizaje automático más avanzados o modelos basados en reglas. Esto permitiría evaluar si otros modelos presentan un mejor desempeño en la predicción del tiempo en rango terapéutico. Además de los modelos supervisados, se podría explorar un enfoque descriptivo utilizando modelos no supervisados, a parte del PCA. Estos modelos podrían identificar patrones y grupos de pacientes basados en características comunes, lo que podría ayudar a comprender mejor la variabilidad en la respuesta al tratamiento anticoagulante.

Para evaluar la generalización y aplicabilidad del modelo de predicción, sería recomendable realizar una validación externa en otro departamento de salud o en un centro clínico independiente. Esto permitiría verificar la capacidad del modelo para predecir el tiempo en rango terapéutico en diferentes contextos clínicos y poblaciones de pacientes.

En caso de lograr un modelo predictivo con suficiente capacidad discriminativa, se podría considerar el desarrollo de una aplicación (*app*) que permita a los médicos ingresar los datos de los pacientes y obtener una predicción del tiempo en rango terapéutico. Esto facilitaría la toma de decisiones clínicas y la optimización del tratamiento anticoagulante.

También sería relevante llevar a cabo estudios observacionales prospectivos para evaluar tanto la detección temprana de un TRT subóptimo mediante la implementación de sistemas de monitoreo continuo a tiempo real como la evolución de estos pacientes tras el cambio a un ACOD, lo que permitiría evaluar la efectividad y seguridad de los ACOD en relación con la incidencia de complicaciones tromboembólicas y hemorrágicas.

En conjunto, la exploración de estas líneas de investigación contribuiría al avance en el campo de la predicción del tiempo en rango terapéutico y la mejora del manejo de los anticoagulantes en la práctica clínica diaria.



9. Referencias

- [1] Lippi, G., Sanchis-Gomar, F., & Cervellin, G. (2020). Global Epidemiology of Atrial Fibrillation: An Increasing Epidemic and Public Health challenge. *Int. J. Stroke*, 16, 217–221.
- [2] Lip, G.Y.H., et al. (2017). Stroke Prevention in Atrial Fibrillation: Past, Present and Future. *Thromb Haemost*, 117 (7), 1230–1239.
- [3] Apostolakis, S., et al. (2013). Factors affecting quality of anticoagulation control among patients with atrial fibrillation on warfarin: The SAME-TT₂R₂ score. *Chest*, 144 (5), 1555-1563.
- [4] Abumuailleq, R.R., et al. (2015). Evaluation of SAME-TT₂R₂ risk score for predicting the quality of anticoagulation control in a real-world cohort of patients with non-valvular atrial fibrillation on vitamin-K antagonists. *Europace*, 17 (5), 711-717.
- [5] Informe de posicionamiento terapéutico UT/V5/21112016. Criterios y recomendaciones generales para el uso de anticoagulantes orales directos (ACOD) en la prevención del ictus y la embolia sistémica en pacientes con fibrilación auricular no valvular. Disponible en: <https://www.aemps.gob.es/medicamentosUsoHumano/informesPublicos/docs/criterios-anticoagulantes-orales.pdf>
- [6] Williams, B. A., et al. (2017). Clinical Prediction Model for Time in Therapeutic Range While on Warfarin in Newly Diagnosed Atrial Fibrillation. *Journal of the American Heart Association*, 6 (10).
- [7] Korenstra, J., et al. (2016). Effectiveness and safety of dabigatran versus acenocoumarol in 'real-world' patients with atrial fibrillation. *Europace*, 18 (9), 1319-27.
- [8] Carles, M., et al. (2015). Cost-effectiveness analysis of dabigatran and anticoagulation monitoring strategies of vitamin K antagonists. *BMC Health Services Research*, 15, 289.
- [9] Ball, J., et al. (2013). Atrial Fibrillation: Profile and burden of an Evolving Epidemic in the 21st century. *Int. J. Cardiol*, 167 (5), 1807–1824.
- [10] January, C.T., et al. (2019). AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration With the Society of Thoracic Surgeons. *J Am Coll Cardiol*, 140 (2), 125-151.
- [11] Cea-Calvo, L., et al. (2007). Prevalence of atrial fibrillation in the Spanish population aged 60 years or more. The PREV-ICTUS study. *Study Rev Esp Cardiol*, 60 (6), 616-24.
- [12] Gamero, M.T., et al. (2017). Factores asociados a la anticoagulación óptima en el adulto mayor con fibrilación auricular. *Revista Médica Herediana*, 28 (2), 84-92.

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

- [13] Pastori, D., et al. (2015). Incidence of myocardial infarction and vascular death in elderly patients with atrial fibrillation taking anticoagulants: Relation to atherosclerotic risk factors. *Chest*, 147, 1644-50.
- [14] Lip, G.Y., et al. (2010). Refining clinical risk stratification for predicting Stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the Euro Heart Survey on atrial fibrillation. *Chest*, 137, 263–272.
- [15] Anguita-Sánchez, M., et al. (2019). Anticoagulantes orales directos frente a antagonistas de la vitamina K en pacientes del «mundo real» con fibrilación auricular no valvular. Estudio FANTASIIA. *Rev Esp Cardiol*, 73 (1), 14-20.
- [16] Florez, J. (2016). Farmacología Humana. 12ª Ed. Madrid: *Editorial Elsevier-Masson*.
- [17] Alania, E.M., et al. (2013). Tiempo en rango terapéutico de los pacientes que siguen tratamiento con derivados cumarínicos en la práctica clínica diaria. *Cardiocoore*, 48 (4), 158-161.
- [18] Hindricks, G., et al. (2021). 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J*, 42 (5), 373-498.
- [19] Kirchhof, P., et al. (2016). 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J*, 37, 2893-962.
- [20] Falcone, M., et al. (2022). Tiempo en rango terapéutico e incidencia de complicaciones por warfarina en una cohorte retrospectiva de 20 años. *Atención Primaria*, 54.
- [21] Asunción Esteve-Pastor, M., et al. (2018). Quality of oral anticoagulation with vitamin K antagonists in 'real-world' patients with atrial fibrillation: a report from the prospective multicentre FANTASIIA registry. *Europace*, 20, 1435-1441.
- [22] Joshua Lin, K., et al. (2017). Prediction Score for Anticoagulation Control Quality Among Older Adults. *Journal of the American Heart Association*, 6.
- [23] Björck, F., Kadhim, H., & Själander, A. (2019). Predictors for INR-control in a well-managed warfarin treatment setting. *J Thromb Thrombolysis*, 47(2), 227–232.
- [24] Wasniewski, S., et al. (2018). Low Performance of a Clinical-Genetic Model in the Estimation of Time in Therapeutic Range in Acenocoumarol-Adherent Patients with Nonvalvular Atrial Fibrillation: The Quality of Anticoagulation Challenge. *Hindawi BioMed Research International*, 9.
- [25] Hurwitz, J., & Kirsch, D. (2018) *Machine Learning for dummies*. IBM Limited Edition.
- [26] Müller, A.C., & Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media.



- [27] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20 (3), 273-297.
- [28] Furey, T.S., et al. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16, 906-14.
- [29] Sarkar, D., & Natarajan, V. (2019). *Ensemble Machine Learning Cookbook: Over 35 Practical Recipes to Explore Ensemble Machine Learning Techniques Using Python*. Packt.
- [30] Pastur-Romay, L.A., et al. (2016). Deep Artificial Neural Networks and Neuromorphic Chips for Big Data Analysis: Pharmaceutical and Bioinformatics Applications. *International Journal of Molecular Sciences*, 17 (8), 1313.
- [31] Lemon, S.C., et al. (2003). Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med*, 26 (3), 172-181.
- [32] Stolfus, J.C. (2011). Logistic regression: a brief primer. *Acad Emerg Med*, 18 (10), 1099-104.
- [33] Sandoval-Riveros, C.L., et al. (2019). Evaluación del SAME-TT2R2 score en la calidad de la anticoagulación con warfarina. *Acta Médica Colombiana*, 44 (2), 75-81.
- [34] Krittayaphong, R., et al. (2019). SAME-TT2R2 score for prediction of suboptimal time in therapeutic range in a Thai population with atrial fibrillation. *Singapore Medical Journal*, 61 (12), 641-646.
- [35] Bernaitis, N., et al. (2019). The SAME-TT2R2 Score Predicts Warfarin Control in an Australian Population with Atrial Fibrillation. *J. Clin. Med*, 8, 882.

10. Anexos: ODS

Como se ha especificado en el apartado 1.3, este proyecto de investigación tiene como objetivo predecir el TRT de los anticoagulantes orales en pacientes con FANV. Al mejorar la precisión en la predicción del TRT, se busca una mejor gestión del tratamiento y una mejora en la salud y el bienestar de los pacientes. Además, el uso de técnicas avanzadas de análisis de datos y modelos de predicción en este proyecto se alinea con el objetivo de promover la innovación y mejorar la calidad de la atención médica. La Tabla 25 muestra el grado de asociación entre los Objetivos de Desarrollo Sostenible y el proyecto.

Objetivo de Desarrollo Sostenible	Alto	Medio	Bajo	N.C.
1. Fin de la pobreza				✗
2. Hambre cero				✗
3. Salud y bienestar	✗			
4. Educación de calidad				✗
5. Igualdad de género				✗
6. Agua limpia y saneamiento				✗
7. Energía asequible y no contaminante				✗
8. Trabajo decente y crecimiento económico				✗
9. Industria, innovación e infraestructura		✗		
10. Reducción de las desigualdades				✗
11. Ciudades y comunidades sostenibles				✗
12. Producción y consumo responsables				✗
13. Acción por el clima				✗
14. Vida submarina				✗
15. Vida de ecosistemas terrestres				✗
16. Paz, justicia e instituciones sólidas				✗
17. Alianzas para lograr los objetivos				✗

Tabla 25. Asociación de los Objetivos de Desarrollo Sostenible con el TFG



11. Glosario de abreviaciones

AAS = ácido acetilsalicílico o aspirina

ACOD= anticoagulante oral de acción directa

ACODs = anticoagulantes orales de acción directa

ADASYN = *Adaptative Synthetic Sampling*

ADE = ancho de distribución eritrocitaria

AEMPS = Agencia Española del Medicamento

AI= aurícula izquierda

AUC ROC = *Area Under the Receiver Operating Characteristic Curve* o área bajo la curva característica operativa del receptor

AVAC = años de vida ajustados por calidad

AVK = antivitamina K o antagonistas de la vitamina K

CHGUV = Consorcio Hospital General Universitario de Valencia

dL= decilitros

DM= diabetes mellitus

DS = desviación estándar

FA = fibrilación auricular

FANV = fibrilación auricular sin patología valvular

g= gramos

HCE = historia clínica electrónica

HTA= hipertensión arterial

IA = inteligencia artificial

IC= insuficiencia cardiaca

IMC= índice de masa corporal

Predicción de tiempo en rango terapéutico para pacientes con fibrilación auricular no valvular anticoagulados con antivitamina K

INR = Índice Normalizado Internacional

IPT = informe de posicionamiento terapéutico

Kg= kilogramos

KNN = *K-Nearest Neighbors* o algoritmo de los K Vecinos más Cercanos

L= litro

m= metro

mg= miligramos

min= minuto

ml= mililitros

ML = *Machine Learning*

MLP = *Multilayer Perceptron*

NHC= número de historia clínica

ODS = Objetivo de Desarrollo Sostenible

PCA= Principal Component Analysis o análisis de Componentes Principales

Q= Quantile

RIQ = rango intercuartílico

s = segundos.

SVM = *Support Vector Machines* o Máquinas de Vectores de Soporte

SVR = *Support Vector Regression*

SVC = *Support Vector Classification*

TFG = Trabajo Fin de Grado

TRT = tiempo en rango terapéutico

VI= ventrículo izquierdo

