

FACULTAD DE MEDICINA

Departamento de Medicina

Predicción de eventos clínicos a través de aprendizaje automático



VNiVERSIDAD
D SALAMANCA

TESIS DOCTORAL

Pablo José Antúnez Muiños

DIRECTORES:

Ignacio Cruz González

Víctor Vicente Palacios

Salamanca, 2023

FACULTAD DE MEDICINA

PROGRAMA DE DOCTORADO:

Biociencias: Biología y Clínica del Cáncer y Medicina Traslacional

Predicción de eventos clínicos a través de aprendizaje automático



DOCTORANDO:

PABLO JOSÉ ANTÚNEZ MUIÑOS

D. IGNACIO CRUZ GONZÁLEZ

**Profesor Titular de Medicina del Departamento de Medicina de la Universidad de Salamanca
hago constar que la tesis titulada**

PREDICCIÓN DE EVENTOS CLÍNICOS A TRAVÉS DE APRENDIZAJE AUTOMÁTICO

Que presenta el doctorando PABLO JOSÉ ANTÚNEZ MUIÑOS, licenciado en Medicina, para optar al grado de doctor, ha sido realizada bajo mi dirección y reúne los requisitos para su presentación, lectura y defensa.

Y para que así conste y en cumplimiento de la legislación vigente, se firma la presente en Salamanca a 26 de octubre de 2023.

Fdo. Dr. Ignacio Cruz González

D. VÍCTOR VICENTE PALACIOS

Doctor por la Universidad de Salamanca hago constar que la tesis titulada

PREDICCIÓN DE EVENTOS CLÍNICOS A TRAVÉS DE APRENDIZAJE AUTOMÁTICO

Que presenta el doctorando PABLO JOSÉ ANTÚNEZ MUIÑOS, licenciado en Medicina, para optar al grado de doctor, ha sido realizada bajo mi dirección y reúne los requisitos para su presentación, lectura y defensa.

Y para que así conste y en cumplimiento de la legislación vigente, se firma la presente en Salamanca a 26 de octubre de 2023.

Fdo. Dr. Víctor Vicente Palacios

A todos los que me habéis ayudado en cada momento,
permitiéndome llegar hasta aquí.

AGRADECIMIENTOS:

Al Dr. Ignacio Cruz por su inestimable liderazgo, su orientación, su apoyo constante y su sabiduría académica. Su guía experta ha sido fundamental en el desarrollo de este proyecto, y su compromiso con mi crecimiento académico es inmensamente apreciado. Además, quiero agradecerle sinceramente su compromiso y apoyo constante a lo largo de toda mi formación como cardiólogo intervencionista. Su influencia perdurará en mi desarrollo académico y profesional, y estoy agradecido por la oportunidad de aprender de su experiencia.

Al Dr. Víctor Vicente por su incalculable apoyo y guía durante este proyecto, cuya experiencia y conocimientos en campos complejos han sido de inestimable valor en este viaje académico. Su dedicación y orientación en áreas que eran previamente más desconocidas para mí han sido esenciales para el desarrollo de esta tesis doctoral.

A los Dres. Javier Martín Moreiras, Javier Rodríguez Collado, Alejandro Diego Nieto y Jesús Herrero Garibi, por su inestimable colaboración, enseñanza y apoyo durante mi formación durante estos últimos años. También al resto del equipo de la sección de hemodinámica, la enfermería, los técnicos y el resto de los componentes de la unidad, por su compañerismo y respeto hacia mi persona desde el primer día.

A todo el servicio de Cardiología del Hospital de Salamanca, que desde el principio me acogieron en mi nueva casa y, en especial al Dr. Pedro Luis Sánchez, por su liderazgo y por haber confiado en mí.

Al Dr. Sergio López Tejero, por este viaje juntos, por su amistad y su apoyo, y por ser un ejemplo de tesón y bondad.

A Pablo Pérez y Andrea Vázquez, por su valioso trabajo y colaboración en este proyecto.

Al servicio de Cardiología del Hospital de Santiago de Compostela, en especial al Dr. José Ramón González Juanatey, por permitirme formar parte de ese grupo y completar mi formación como cardiólogo tanto en la faceta clínica, como investigadora y docente.

A mis familiares y amigos, por su respaldo, comprensión y aliento.

A mis padres, Ana y Pepe, por su amor, apoyo incondicional y sacrificio a lo largo de toda mi vida que han sido los cimientos sobre los cuales se ha construido mi educación. Sin su constante aliento y respaldo, este logro no habría sido posible.

A mis hermanos, Fernando y Marta, quienes siempre han estado ahí para animarme y motivarme, les agradezco por ser una fuente constante de inspiración y alegría en mi vida.

A Alejandra, por su apoyo inquebrantable, paciencia y comprensión en todas las facetas de la vida. Mi compañera en este viaje, tu presencia y aliento constante han sido mi mayor fortaleza.

Salamanca, 26 de octubre de 2023

LISTADO DE ABREVIATURAS:

ACP: Amplatz Cardiac Plug

AIT: ataque isquémico transitorio

CSV: formato de archivo "Comma Separated Values", o valores separados por comas

DRT: trombosis del dispositivo o *device-related thrombosis*

ETE: ecocardiograma transesofágico

GB: *gradient boosting*

IA: inteligencia artificial

ML: *machine learning*

OMS: Organización Mundial de la Salud

RF: *random forest*

ROC: *receiver operating characteristic* o característica operativa del receptor

RL: regresión logística

TC: tomografía computarizada

t-SNE: *t-distributed stochastic neighbor embedding*

UMAP: *uniform manifold approximation and projection*

ÍNDICE

1.	Introducción	1
1.1.	Fibrilación auricular.....	3
1.1.1.	Epidemiología.....	3
1.1.2.	Fisiopatología	5
1.1.3.	Riesgo tromboembólico	6
1.1.4.	Estrategias para la prevención de fenómenos tromboembólicos.....	9
1.2.	Trombosis sobre el dispositivo de cierre de orejuela.....	15
1.2.1.	Epidemiología, fisiopatología y diagnóstico	15
1.2.2.	Problemas en los análisis estadísticos previos sobre trombosis sobre dispositivo de cierre de orejuela	17
1.3.	Inteligencia artificial y <i>machine learning</i> o aprendizaje automático.....	22
1.3.1.	Nacimiento y desarrollo de la inteligencia artificial	22
1.3.2.	Ramas de la inteligencia artificial	23
1.3.3.	Tipos de aprendizaje automático	26
1.3.4.	Tipos de algoritmos	31
1.3.5.	Construcción de modelos de <i>machine learning</i>	34
1.4.	Educación en aprendizaje automático	36
1.4.1.	Situación de la formación en inteligencia artificial y aprendizaje automático en los trabajadores en el ámbito sanitario.....	36
2.	Justificación	40
3.	Hipótesis.....	44
4.	Objetivos	47

5. Resultados	50
5.1. Aplicación del aprendizaje automático en la medicina. Comparación de la potencia de predicción de trombosis sobre dispositivo de cierre de orejuela entre los métodos estadísticos clásicos y las técnicas de <i>machine learning</i>	52
5.1.1. Objetivos	52
5.1.2. Métodos	53
5.1.3. Resultados	59
5.2. Evaluación del conocimiento de los profesionales de la salud y desarrollo de una plataforma para facilitar el uso de modelos de inteligencia artificial en el ámbito sanitario. Proyecto KOOPA ML	67
5.2.1. KoopaML	69
5.2.2. Métodos	70
5.2.3. Resultados	75
6. Discusión	82
6.1. Aplicación del aprendizaje automático en la medicina. Comparación de la potencia de predicción de trombosis sobre dispositivo de cierre de orejuela entre los métodos estadísticos clásicos y las técnicas de <i>machine learning</i>	84
6.1.1. Limitaciones.....	89
6.1.2. Resumen de los hallazgos.....	90
6.2. Problemática de los automatismos matemáticos	91
6.3. Evaluación del conocimiento de los profesionales de la salud y desarrollo de una plataforma para facilitar el uso de modelos de inteligencia artificial en el ámbito sanitario. Proyecto KOOPA ML	94
6.3.1. Discusión general	94

6.3.2.	Discusión desde el punto de vista médico	97
6.3.3.	Perspectivas futuras	99
6.3.4.	Resumen de los hallazgos.....	100
6.4.	Ventajas e inconvenientes del uso del machine learning.....	101
6.4.1.	Ventajas o potenciales usos	101
6.4.2.	Desventajas o retos pendientes de clarificar.....	102
6.5.	Retos éticos y aspectos legales de la IA.....	104
6.5.1.	Responsabilidad	106
6.5.2.	Obligaciones	107
6.5.3.	Aspectos legales y culpabilidad	107
6.5.4.	Retos éticos	108
6.5.5.	Plagio o derechos de autor.....	109
7.	Conclusiones.....	111
8.	Bibliografía.....	113
9.	Anexos	124
9.1.	Predictive power for thrombus detection after atrial appendage closure: machine learning vs. classical methods	126

ÍNDICE DE FIGURAS

Figura 1. Epidemiología de la FA.	4
Figura 2. Fisiopatología de la formación de trombos en la orejuela izquierda en la fibrilación auricular.	6
Figura 3. Indicaciones de cierre percutáneo de orejuela	12
Figura 4. Dispositivos de cierre percutáneo de orejuela izquierda.	14
Figura 5. Trombosis de dispositivo de cierre de orejuela.	16
Figura 6. Inteligencia artificial y su división en las diferentes ramas.	24
Figura 7. Técnicas aplicadas en la inteligencia artificial.	25
Figura 8. Esquema aprendizaje supervisado.	27
Figura 9. Esquema aprendizaje no supervisado.	29
Figura 10. Esquema aprendizaje semisupervisado.	29
Figura 11. Esquema aprendizaje por refuerzo.	30
Figura 12. Tipos de algoritmos.	32
Figura 13. Diseño y flujo de trabajo de los experimentos.	55
Figura 14. Resultados del modelo de regresión logística.	62
Figura 15. Peso de cada variable en el experimento II al utilizar la regresión logística.	63
Figura 16. Ratio de variación del predictor en cada remuestreo en el experimento III (A) y en el IV (B) utilizando la regresión logística.	66
Figura 17. Ejemplo de un flujo de trabajo con KoopaML.	70
Figura 18. Recomendaciones adicionales.	72
Figura 19. Nodo correspondiente a la subida de datos.	76

Figura 20. Nodo para la división de los datos en el subconjunto para entrenamiento del modelo y otro para el test posterior.....	77
Figura 21. Nodo del algoritmo seleccionado para confeccionar el modelo, en este caso el Random Forest.	78
Figura 22. Nodo para la validación del modelo.....	79
Figura 23. Mapas de utilización de la aplicación KoopaML.....	96

ÍNDICE DE TABLAS

Tabla 1. Escala de riesgo tromboembólico CHA2DS2-VASc.....	7
Tabla 2. Escala de riesgo hemorrágico HAS-BLED.....	8
Tabla 3. Estudios comparativos de cierre de orejuela izquierda frente a anticoagulación oral para la prevención de eventos cardioembólicos en FA.	11
Tabla 4. Estudios que analizaron posibles predictores de trombosis de dispositivo de cierre de orejuela.	21
Tabla 5. Características basales.	59
Tabla 6. Resultados de los cuatro experimentos.....	60
Tabla 7. Selección de variables predictoras en el experimento I.....	63
Tabla 8. Coincidencia como predictores de trombosis sobre dispositivo de cierre de orejuela en cada método.	64
Tabla 9. Sanitarios participantes del estudio y sus conocimientos sobre inteligencia artificial..	71
Tabla 10. Ejercicios y tareas propuestos a los participantes.	73
Tabla 11. Resultados de las tareas completadas por los participantes en función del grupo.....	81

1. Introducción

1. Introducción

1.1. Fibrilación auricular

La fibrilación auricular es una arritmia supraventricular que se caracteriza por una despolarización no coordinada de las aurículas, que se debe a fenómenos de microrreentrada que se generan y perpetúan. Esto conlleva que no exista una contracción auricular adecuada y que desaparezcan las ondas P en el electrocardiograma. Además, esta activación no coordinada de la aurícula, va a provocar que el pulso cardíaco sea irregular ya que la despolarización llega al ventrículo también de forma no constante, provocando intervalos R-R no regulares¹.

Para establecer su diagnóstico, es necesario que se objete, bien en un ECG de 12 derivaciones o bien en un registro de una sola derivación durante más de 30 segundos, un ritmo con intervalos R-R irregulares y en ausencias de ondas P².

1.1.1. Epidemiología

La fibrilación auricular es la arritmia más frecuente en el mundo, y su prevalencia está en aumento, especialmente en los países más desarrollados (Figura 1). Actualmente, la prevalencia de fibrilación auricular, ronda entre el 2-4% de la población³, aunque, debido al aumento de la esperanza de vida, se espera que pueda aumentar a más del doble en los próximos años. De hecho, se estima que el riesgo de presentar fibrilación auricular a lo largo de toda su vida en la población europea de 55 años es de más de un tercio. Sin embargo, a pesar de que su prevalencia aumenta considerablemente con la edad, esta arritmia no se asocia solamente al envejecimiento de la población. Existen múltiples factores asociados que se han demostrado tener mayor predisposición para presentar fibrilación auricular, como la hipertensión arterial, la diabetes mellitus, el sobrepeso, la enfermedad renal crónica o el síndrome de apnea/hipopnea del sueño, entre otros².

1. Introducción

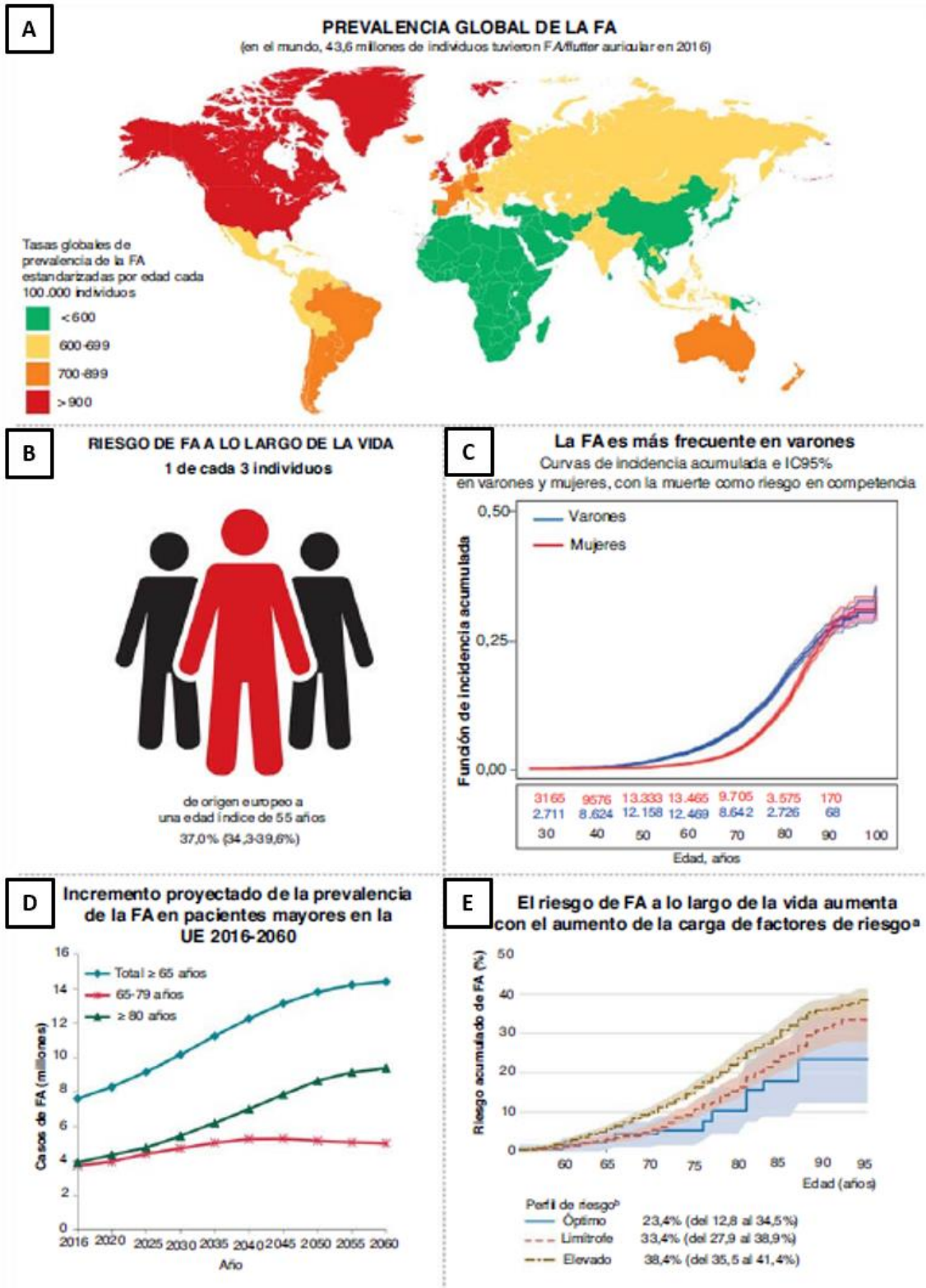


Figura 1. Epidemiología de la FA. A: Prevalencia global de FA. B: Riesgo de presenta FA a lo largo de la vida. C: Incidencia acumulada por sexo. D: Expectativa de prevalencia de FA en los próximos años en la UE. E: Riesgo acumulado según factores de riesgo. Adaptado de Hindricks et al². FA: fibrilación auricular, UE: Unión Europea.

^aTabaquismo, consumo de alcohol, índice de masa corporal, diabetes, y antecedente de infarto agudo de miocardio o insuficiencia cardíaca.

^bPerfil de riesgo: óptimo (ningún factor de riesgo), límitrofe (más de un factor de riesgo en el límite), elevado (más de un factor de riesgo importante).

1. Introducción

Es por todo esto, que las sociedades científicas inciden en la necesidad de implementar programas multidisciplinarios para la prevención de esta arritmia. La mayor longevidad, el sedentarismo y el aumento de la prevalencia de factores de riesgo para el desarrollo de la fibrilación auricular, hacen que el riesgo de padecerla a lo largo de la vida sea cada vez mayor con la repercusión que esto conlleva. Se ha demostrado que la fibrilación auricular empeora el pronóstico de los pacientes, aumentando la mortalidad, el riesgo de ictus isquémico^{4,5}, enfermedades cardíacas como la insuficiencia cardíaca⁶, e incluso el deterioro cognitivo⁷ o la depresión. Además, más del 60% de los pacientes ven reducida su calidad de vida, lo que releja claramente el impacto de esta arritmia en la sociedad⁸.

1.1.2. Fisiopatología

La aparición de fibrilación auricular se basa en tres pilares: el sustrato, que proporcionará una base sobre la que se desarrolle la arritmia; los *triggers*, que provocarán que la arritmia se inicie; y los catalizadores, que facilitarán este proceso⁹. El sustrato puede ser tanto anatómico como electrofisiológico, aunque la principal causa siempre será la dilatación auricular y la presencia de áreas de fibrosis. Esto podrá ser secundario a diferentes cardiopatías (como la cardiopatía hipertrófica o la estenosis mitral) o podrá estar relacionado con factores de riesgo (como la hipertensión arterial). Una vez se da el sustrato adecuado que comentamos, la presencia de extrasístoles supraventriculares puede actuar como *triggers* y desencadenar la arritmia. Estas suelen producirse a nivel de las venas pulmonares o de la cava superior. Finalmente, el sistema nervioso autónomo, el hipertiroidismo o el uso de drogas pueden actuar como catalizadores y favorecer este mecanismo.

Una vez se desencadena la arritmia, debido a los múltiples focos de microrreentrada que se generan, se produce una despolarización no coordinada de la aurícula, y esto perpetúa la arritmia. Esto conlleva la pérdida de la contracción auricular, con la consiguiente caída del gasto cardíaco, especialmente en las cardiopatías con alteración de la relajación ventricular. Todo este

1. Introducción

proceso genera una retroalimentación deletérea que provoca que la miopatía atrial continúe progresando. La fibrilación auricular genera más dilatación y más fibrosis auricular, provocando entonces un aumento también de la propia fibrilación auricular.

1.1.3. Riesgo tromboembólico

Todos estos cambios, tanto morfológicos como funcionales, promueven un estado protrombótico (Figura 2). La pérdida de la contracción auricular aumentará la estasis sanguínea, que promoverá la formación de trombos, especialmente a nivel de la orejuela izquierda. Además, se produce una alteración del endotelio vascular debido al remodelado auricular, que aumentará también el riesgo de formación de trombos en esta cavidad.

Esto provoca que los pacientes con fibrilación auricular presenten un riesgo hasta 5 veces mayor de presentar accidentes cerebrovasculares isquémicos¹⁰. Se considera que hasta el 15-20% de los ictus isquémicos son secundarios a fibrilación auricular y, además, presentan una mayor morbimortalidad con los ictus no cardioembólicos¹¹.

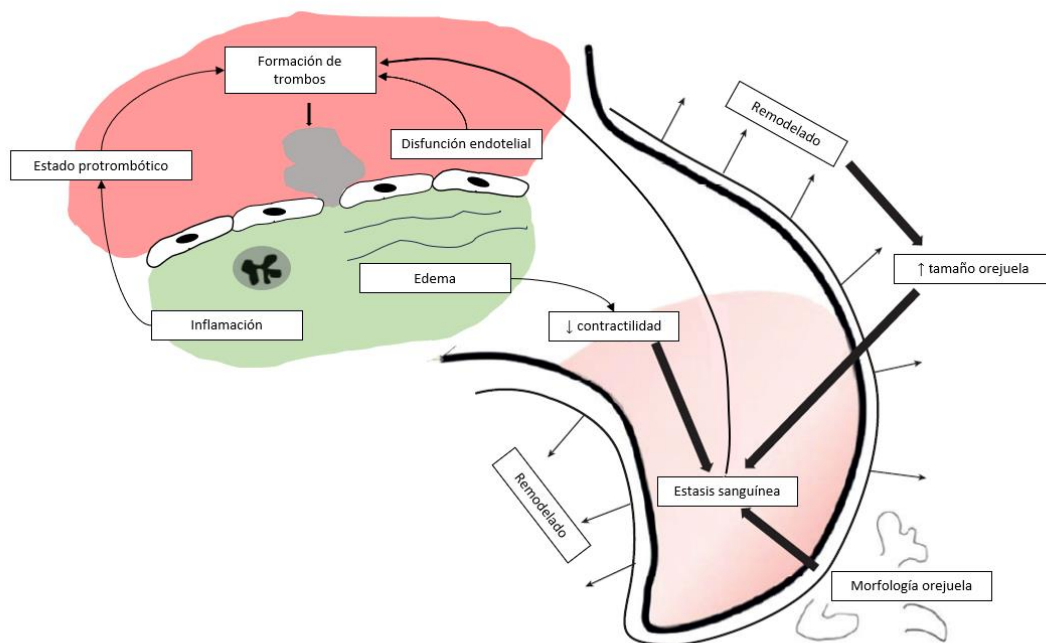


Figura 2. Fisiopatología de la formación de trombos en la orejuela izquierda en la fibrilación auricular. *Adaptado de Glikson et al¹².*

1. Introducción

La estratificación del riesgo tromboembólico secundario a la presencia de fibrilación auricular no valvular supone un aspecto clave a la hora de plantear las diferentes opciones terapéuticas. Aquellos pacientes con un elevado riesgo tromboembólico podrían beneficiarse del tratamiento anticoagulante, sin embargo, esta decisión debe estar balanceada también por el riesgo hemorrágico que presenten. Para evaluar el riesgo tromboembólico de estos pacientes se instauró en primer lugar una escala de riesgo denominada CHADS₂¹³. Esta escala no parecía identificar de forma clara aquellos pacientes con bajo riesgo que no se beneficiaban del tratamiento anticoagulante. Posteriormente se instauró una nueva escala, CHA₂DS₂-VASC¹⁴, que adaptó la previa y mejoró la estratificación de estos pacientes. Esta última es hoy en día la recomendada para descartar aquellos pacientes que no se benefician del inicio de la terapia anticoagulante (Tabla 1). Las guías europeas de práctica clínica², recomiendan el inicio de la anticoagulación en pacientes con una puntuación de 2 o más en hombre y 3 o más en mujeres (clase I, nivel de recomendación A) y en varones con una puntuación de 1 o en mujeres de 2 (clase IIa, nivel de recomendación B).

Tabla 1. Escala de riesgo tromboembólico CHA₂DS₂-VASC

Escala CHA ₂ DS ₂ -VASC	Factores de riesgo	Puntuación
C	Insuficiencia cardíaca	1
H	Hipertensión arterial	1
A	Edad ≥ 75 años	2
D	Diabetes	1
S	Ictus, AIT o fenómeno embólico	2
V	Enfermedad vascular (coronaria o periférica)	1
A	Edad 65-74 años	1
Sc	Sexo (mujer)	1

AIT: ataque isquémico transitorio.

1. Introducción

En el lado contrario, como comentamos previamente, la decisión ha de ser tomada teniendo en cuenta de forma balanceada, tanto el riesgo tromboembólico como el hemorrágico. Para evaluar el segundo, surgió una nueva escala denominada HASBLED¹⁵, recomendada por las sociedades científicas para identificar aquellos pacientes con elevado riesgo en los que la anticoagulación puede ser deletérea (Tabla 2). Aquellos pacientes con un riesgo hemorrágico elevado (HASBLED ≥ 3) deberán de ser reevaluados de una forma más estrecha para intentar controlar o modificar los factores de riesgo de sangrado. En caso de no poder controlarlos, deberán plantearse alternativas a la anticoagulación para la prevención de eventos tromboembólicos.

Tabla 2. Escala de riesgo hemorrágico HAS-BLED.

Escala HAS-BLED	Factores de riesgo	Puntuación
H	Hipertensión no controlada (TAS > 160 mmHg)	1
A	Función renal alterada (Diálisis, trasplante o creatinina > 200 mmol/L)	1
	Función hepática alterada (cirrosis, bilirrubina $\times 2$ LSN, AST/ALT/ALP $\times 3$ LSN)	1
S	Ictus hemorrágico o isquémico previo	1
B	Antecedentes o predisposición al sangrado	1
L	INR lábil (tiempo en rango terapéutico < 60%)	1
E	Edad > 65 años	1
D	Consumo semanal de alcohol excesivo	1
	Uso concomitante de antiagregantes y AINES	1

AINES: antiinflamatorios no esteroideos; ALP: fosfatasa alcalina; ALT: alanina transaminasa; AST: aspartato aminotransferasa; LSN límite superior de la normalidad; TAS: tensión arterial sistólica

1. Introducción

1.1.4. Estrategias para la prevención de fenómenos tromboembólicos

1.1.4.1. Tratamiento farmacológico

Tanto las guías de práctica clínica europeas como las americanas recomiendan el inicio del tratamiento anticoagulante en pacientes con fibrilación auricular, excepto en aquellos pacientes con bajo riesgo tromboembólico (puntuación en la escala CHA₂DS₂-VASc de 0 en varones o 1 en mujeres)^{2,16}. En un primer lugar, se comenzaron a utilizar los fármacos anti vitamina K para la prevención del ictus isquémico, ya que se objetivó que reducía el riesgo más que el tratamiento antiagregante, reduciendo las tasas de ictus en hasta un 60%¹⁷. Sin embargo, cabe destacar también que el riesgo de hemorragia extracraneal era más elevado. Ante estos resultados, surgieron nuevas opciones terapéuticas, como los anticoagulantes de acción directa que son actualmente los de primera elección, ya que han demostrado mejores resultados a la hora de la prevención de eventos tromboembólicos y un menor riesgo de sangrado que los fármacos anti vitamina K¹⁸⁻²¹.

A pesar del mejor perfil de seguridad de los nuevos anticoagulantes de acción directa, persisten pacientes con elevado riesgo hemorrágico que presentan contraindicación para el tratamiento anticoagulante. Por este motivo, las guías de práctica clínica recomiendan el cierre de orejuela ya como una alternativa a la anticoagulación en estos pacientes (clase IIb, nivel de evidencia B).

1.1.4.2. Tratamiento no farmacológico: El cierre percutáneo de la orejuela izquierda

A pesar de la llegada de los anticoagulantes de acción directa, la ratio de pacientes que tienen indicación de anticoagulación por elevado riesgo tromboembólico que están bajo este tratamiento apenas supera el 60%²². Esto significa que un alto porcentaje de estos pacientes no se encuentran tratados, con el riesgo que ello conlleva. O'Brien et al²³ han observado en su registro, que el 13,1% presentan alguna contraindicación para el inicio de la terapia anticoagulante. Por otro lado, en un estudio que incluía 86671 pacientes, Steinberg et al²⁴ han

1. Introducción

detectado que hasta el 2.2% presentaban alguna contraindicación absoluta, siendo la mayoría el antecedente de hemorragia intracraneal.

Es en este contexto el cierre percutáneo de la orejuela emerge como una alternativa al tratamiento anticoagulante. El aumento de la trombogenicidad en la fibrilación auricular se debe a múltiples factores, pero uno de los más importantes es la miopatía atrial que se produce. La progresiva dilatación y fibrosis de la aurícula, además de la pérdida de una contracción auricular eficiente, y las alteraciones a nivel endotelial, acaban provocando un escenario protrombótico, con el consiguiente aumento del riesgo isquémico. Dentro de la anatomía de la aurícula izquierda, se ha objetivado que la inmensa mayoría (>90%) de trombos se forma a nivel de la orejuela¹². La propia anatomía de este apéndice, que produce un aumento de la estasis sanguínea sobre todo en la fibrilación auricular, la hace proclive para la formación de trombos. Incluso, se ha estudiado de forma objetiva, como la reducción de la velocidad pico dentro de la orejuela en pacientes con fibrilación auricular es un marcador independiente de ictus isquémico²⁵.

Desde hace más de 20 años se ha utilizado el abordaje percutáneo para el cierre de la orejuela, con una evidencia cada vez más sólida. Se han publicado diversos estudios aleatorizados en los que se comparó el cierre de orejuela frente a la anticoagulación, bien con fármacos anti-vitamina K o bien con anticoagulantes de acción directa (Tabla 3). Estudios como el PROTECT AF^{26,27} o el PREVAIL²⁸ demostraron la no inferioridad del cierre de orejuela frente a los anticoagulantes anti-vitamina K para la prevención del evento combinado de ictus, embolismo sistémico y muerte por causa cardiovascular. Incluso, un metaanálisis²⁹ posterior de ambos estudios con un seguimiento de 5 años confirmó la no inferioridad frente a warfarina para la prevención de ictus isquémico, además de resultar superior para sangrados mayores y para mortalidad. Por otro lado, el estudio PRAGUE-17³⁰ demostró la no inferioridad frente a anticoagulantes de acción directa para el combinado de eventos cardiovasculares o neurológicos adversos y sangrados. También, se

1. Introducción

objetivó un descenso significativo en los eventos hemorrágicos no relacionados con el procedimiento. En un metaanálisis posterior, incluyendo estos tres ensayos clínicos aleatorizados, Turagam et al³¹ ratificaron los hallazgos previos. No se encontraron diferencias significativas en la tasa de ictus isquémico o de cualquier hemorragia mayor.

Tabla 3. Estudios comparativos de cierre de orejuela izquierda frente a anticoagulación oral para la prevención de eventos cardioembólicos en FA.

Estudio	Autores	Comparación	Objetivo	Resultado
Protect AF²⁷	Holmes et al	Cierre de orejuela (Watchman) vs warfarina	<ul style="list-style-type: none"> - Compuesto de eficacia (ictus, embolismo periférico, muerte CV/desconocida) - Combinado de seguridad (eventos relacionados con el procedimiento y sangrados relevantes) 	RR 0.60 (IC 95% 0.41-1.05) RR 1.17 (IC 95% 0.78-1.95)
Prevail²⁸	Holmes et al	Cierre de orejuela (Watchman) vs warfarina	<ul style="list-style-type: none"> - Compuesto de eficacia (ictus, embolismo periférico y muerte CV/desconocida) - Objetivo de seguridad (eventos adversos relacionados con el procedimiento) 	RR 1.07 (IC 95% 0.57-1.89) sin cumplir criterio de no inferioridad de 2.2% (Margen de referencia para no inferioridad 95% unilateral <2.652%)
Prague-17³⁰	Osmancik et al	Cierre de orejuela (Amulet o Watchman) vs	<ul style="list-style-type: none"> - Evento combinado (fenómenos cardioembólicos, muerte CV, 	HR 0.81 (IC 95% 0.56-1.18) p = 0.27; p para no

1. Introducción

		ACODs	sangrado relevante, complicaciones periprocedimiento)	inferioridad = 0.006)
CV: cardiovascular, HR: hazard ratio, IC: intervalo de confianza, RR: <i>rate ratio</i> o razón de tasas				

Ante la evidencia de los estudios publicados hasta la fecha, se han establecido diferentes indicaciones para el cierre percutáneo de la orejuela izquierda, entre los que destacan el alto riesgo hemorrágico y el fracaso documentado del tratamiento anticoagulante para la prevención de eventos tromboembólicos (Figura 3). Incluso se empieza a plantear también su indicación en pacientes con enfermedad renal avanzada, en diálisis o no, ya que no se pueden prescribir los anticoagulantes de acción directa, quedando como única alternativa los fármacos anti-vitamina K³².

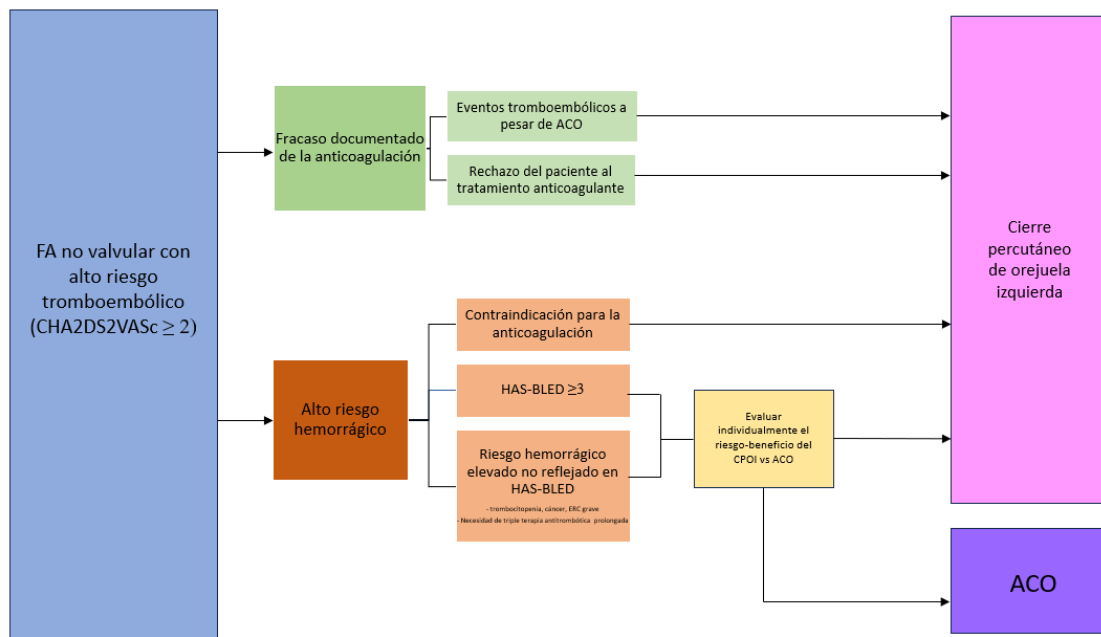


Figura 3. Indicaciones de cierre percutáneo de orejuela. Adaptado de Cruz-González et al[45]

El éxito de este procedimiento, se basa no solo en su eficacia, sino en una baja tasa de complicaciones periprocedimiento y durante el seguimiento³³. La evolución de la técnica, como

1. Introducción

de los materiales, ha posibilitado estos resultados, con una tasa de éxito del procedimiento muy elevada³⁴. El primer paso debe ser siempre la correcta selección de pacientes, con un estudio previo con imagen (mediante ETE -ecocardiograma transesofágico- o TC -tomografía computarizada- cardíaco) para evaluar la morfología, las dimensiones de la orejuela y descartar así la presencia de trombo intracavitario. Durante el procedimiento, este debe guiarse mediante imagen cardíaca, bien sea a través de ETE o mediante ecografía intracardiaca. Diferentes pasos son necesarios para llevar a cabo la intervención de forma segura, comenzado por la canalización de un acceso femoral venoso^{12,35}. Seguidamente, se accederá a la aurícula derecha, y se procederá a la realización de la punción transeptal para acceder a la aurícula izquierda. Para facilitar el acceso posterior a la orejuela, deberíamos intentar pinchar en la parte posterior e inferior de la fosa oval, ya que este apéndice se sitúa anterior y superior al septo interauricular. Es importante guiar este paso con la ecografía para intentar no dañar estructuras adyacentes, como la raíz aórtica. Una vez en la aurícula izquierda, se administrará la heparina a dosis plenas (100 UI por Kg de peso), e intentaremos alojar la guía en la vena pulmonar superior izquierda. En esta posición donde se realizará el intercambio por la vaina de liberación del dispositivo. Desde la vena pulmonar, se caerá a la aurícula y se avanzará hacia la orejuela. Finalmente, se procederá al implante del dispositivo seleccionado. Si se confirma una buena posición, con correcta estabilidad y sin fugas peridispositivo residuales, se liberará y se finalizará el procedimiento.

Para este procedimiento se han ido diseñando a lo largo del tiempo diferentes dispositivos que han ido evolucionando y mejorando las características para un mejor anclaje y mayor sellado del apéndice. Estos se dividen en tres tipos: dispositivos de tapón, de disco y de ligadura. Los primeros presentan un lóbulo que produce el sellado a nivel del cuello de la orejuela (Figura 4). En contrapartida, los dispositivos de disco presentan un lóbulo o paraguas con el que se fija, y un disco cubre el orificio del apéndice. Los dispositivos más utilizados y testados son el Watchman (Boston Scientific®), con su evolución Watchman FLX, y el Amulet (Abbott®) con su

1. Introducción

predecesor el dispositivo ACP (Amplatzer Cardiac Plug). Además, existen otros modelos cuyo uso también se está extendiendo como el Lambre (Lifetech®) o el Omega (Eclipse®). Por otro lado, existen otros dispositivos como el Lariat (SentreHeart®), en el que se excluye de forma completa y permanente la orejuela izquierda desde un acceso percutáneo epicárdico.

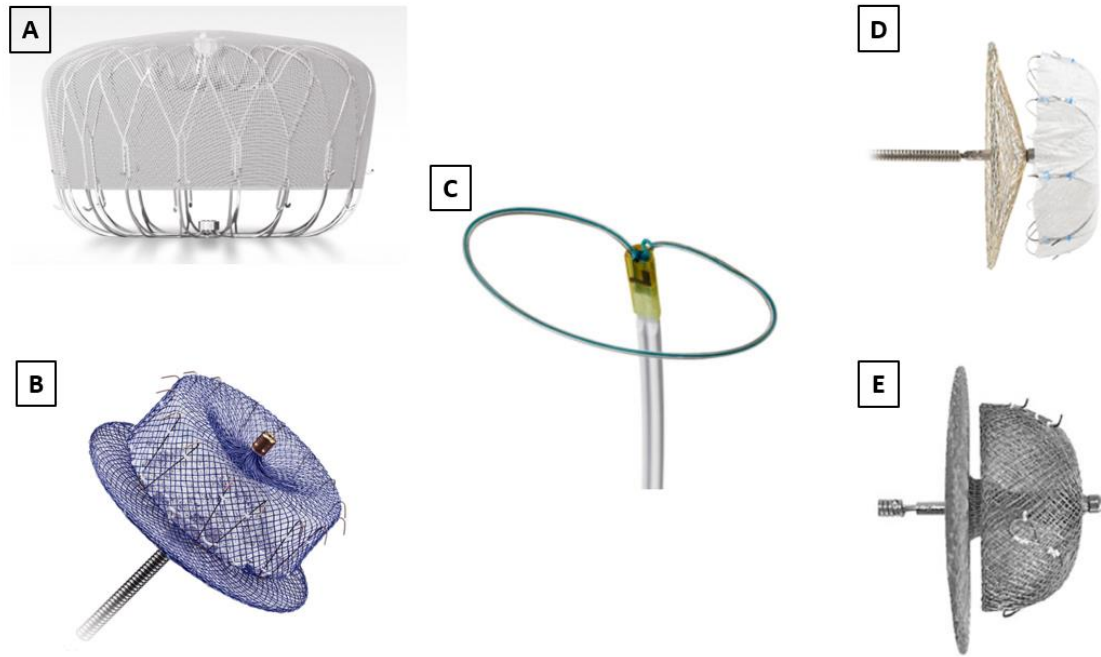


Figura 4. Dispositivos de cierre percutáneo de orejuela izquierda. A: *Watchman* (Boston Scientific); B: *Amulet* (Abbott); C: *Lariat* (SentreHeart); D: *Lambre* (Lifetech); E: *Omega* (Eclipse)

Al tratarse de un procedimiento de prevención para evitar fenómenos cardioembólicos, es de vital importancia reducir las potenciales complicaciones al mínimo. Con los nuevos dispositivos, y la mayor experiencia adquirida en los profesionales, la tasa de éxito y de oclusión completa es cada vez mayor, reduciéndose a su vez las complicaciones a niveles muy bajos. Sin embargo, sigue habiendo discusión y dudas sobre el manejo antitrombótico los primeros meses tras el procedimiento, antes de que se produzca la completa endotelización del dispositivo, ya que se utilizaron pautas diferentes en los estudios publicados. Actualmente se están llevando a cabo diversos ensayos aleatorizados que podrán aportar mayor evidencia al respecto, planteando la posibilidad de utilizar monoantiagregación, doble antiagregación, o anticoagulación, incluso con anticoagulantes de acción directa a dosis bajas. Por otro lado, surgen también las dudas respecto

1. Introducción

a la actitud a mantener en pacientes que presentan fugas peridispositivo residuales en el seguimiento, sin exclusión completa de la orejuela. Se ha establecido un límite de tamaño entre aquellos pequeños o grandes, que podría ser el punto de corte para mantener o no el tratamiento anticoagulante si el paciente no presenta contraindicación. Finalmente, también se ha visto durante el seguimiento con técnicas de imagen (mediante ETE o TC) otra complicación como es la trombosis sobre el dispositivo de cierre de orejuela. Sobre la cara auricular del dispositivo en ocasiones se forman trombos que pueden llegar a provocar eventos tromboembólicos si se sueltan al torrente sanguíneo³⁶.

1.2. Trombosis sobre el dispositivo de cierre de orejuela

1.2.1. Epidemiología, fisiopatología y diagnóstico

La trombosis sobre el dispositivo de cierre de orejuela es una de las complicaciones que se pueden dar durante el seguimiento tras la intervención. Su aparición tras el implante de dispositivo es poco frecuente, pero parece que puede tener un impacto clínico relevante. Tanto el ETE como el TC cardíaco son fundamentales para su diagnóstico durante el seguimiento (Figura 5). En los diferentes registros la frecuencia durante el seguimiento, varía entre el 3-5% de los casos aproximadamente³⁶⁻³⁹. En cuanto a la trascendencia de este fenómeno, en los primeros registros no se evidenció que aquellos pacientes que desarrollaban trombo sobre la cara auricular del dispositivo presentaran una mayor incidencia de eventos tromboembólicos durante el seguimiento³⁹. Sin embargo, en registros más actuales, se ha objetivado que estos pacientes presentan un mayor porcentaje de ictus o AIT (ataque isquémico transitorio) durante el seguimiento⁴⁰⁻⁴², e incluso en algún caso una mayor mortalidad o eventos cardiovasculares adversos^{36,38}.

1. Introducción

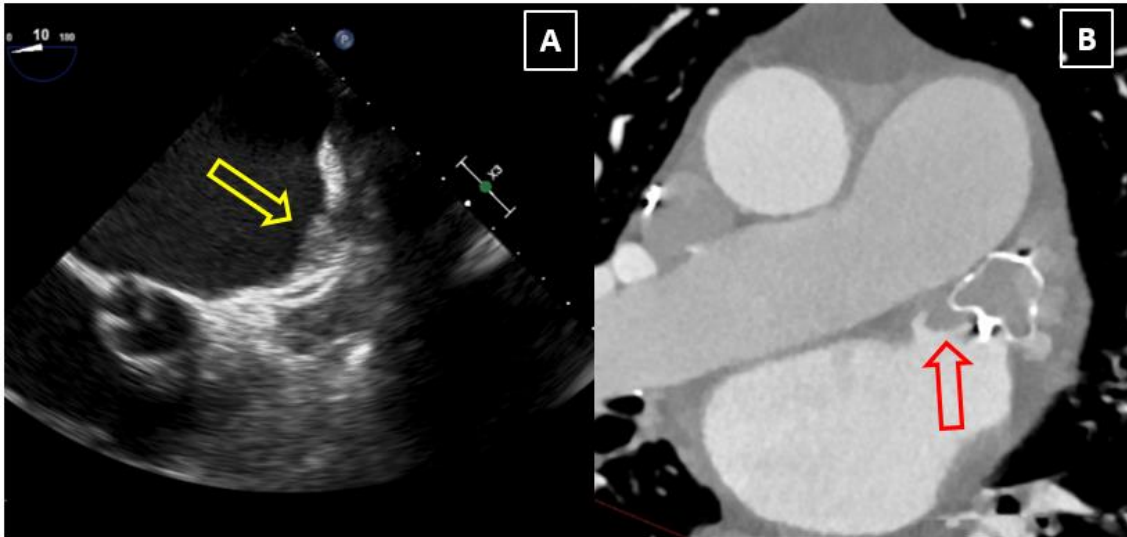


Figura 5. Trombosis de dispositivo de cierre de orejuela. Diagnóstico mediante ecocardiograma transesofágico (A) o mediante TC (B)

Hasta el momento no se conoce exactamente la fisiopatología de este fenómeno. En un primer momento, se creía que se debía fundamentalmente a una falta de endotelización del dispositivo durante los primeros meses. Por este motivo, los primeros registros abogaron por alargar el tratamiento antitrombótico los primeros meses, incluso hasta 6 meses tras el procedimiento. Este tratamiento se basa mayoritariamente en la doble antiagregación o la anticoagulación asociada o no a monoantiagregación. Actualmente, se utilizan protocolos con menor duración, debido a que, en la mayoría de los casos, la indicación de cierre de orejuela es debido al alto riesgo hemorrágico.

Sin embargo, actualmente se han publicado diferentes registros en los que un porcentaje considerable de los eventos de trombosis del dispositivo se produce después de los primeros seis meses desde procedimiento índice^{38,40,42,43}. Estos hallazgos sugieren que pueden existir otros factores que jueguen un papel determinante en la formación de trombos sobre el dispositivo durante el seguimiento.

Conseguir seleccionar aquellos casos con un riesgo elevado de desarrollar esta complicación parece de gran importancia. El impacto clínico que puede tener en estos pacientes, ya de por sí con un riesgo tromboembólico inherentemente elevado, parece muy relevante. Cabe destacar

1. Introducción

que este hallazgo durante el seguimiento supondría, muy probablemente, tener que reintroducir o mantener la anticoagulación, asociada o no a monoantiagregación, en pacientes que muchas veces presentan elevado riesgo de sangrado o numerosos episodios de sangrado previo. En este contexto, diferentes estudios han intentado dilucidar cuáles son los factores de riesgo para desarrollar un trombo sobre el dispositivo, y así poder realizar una medicina más personalizada, por ejemplo, con un seguimiento más cercano con técnicas de imagen más frecuentes.

1.2.2. Problemas en los análisis estadísticos previos sobre trombosis sobre dispositivo de cierre de orejuela

Diversos estudios en relación con la trombosis sobre dispositivo de cierre de orejuela se han publicado hasta la fecha con resultados dispares. En muchos de ellos, se buscan factores predictores de trombosis del dispositivo, pero, en general, no se detalla cuál es la potencia predictiva de los modelos que se crean. Poder encontrar esos factores predisponentes para presentar trombosis sobre el dispositivo tras el cierre de orejuela es de vital importancia, por este motivo los esfuerzos se centran en ello. Sin embargo, para ello es necesario discernir cuánta potencia poseen estos modelos, ya que una baja capacidad predictiva asignaría a estas variables poca relevancia, tanto estadística como clínica. Por otro lado, en varias ocasiones existe un error en la terminología utilizada, ya que se habla de análisis multivariante cuando probablemente sea más razonable utilizar el multivariable. Así como el segundo sí que se utiliza para hacer predicciones de un evento como variable dependiente, a través de modelos constituidos por una o más variables independientes, bien sean cualitativas o cuantitativas, el primero no tiene este uso. En el análisis multivariante se reconocen patrones y las relaciones existentes entre diferentes variables. De esta forma, se utiliza con frecuencia para realizar inferencias de los resultados de una muestra a la población, generando modelos que tienen en cuenta la posible aleatoriedad de los datos observados inherentes a la propia selección de la muestra estudiada.

1. Introducción

En los siguientes párrafos comentaremos los resultados de cada uno de ellos, así como las limitaciones que, desde nuestro punto de vista, presentan (Tabla 4).

En primer lugar, Saw et al³⁹ analizan en su estudio más de mil pacientes del registro multicéntrico de la experiencia con el dispositivo ACP. En este análisis, pretenden objetivar cuáles son las variables predictoras. Para ello realizan, lo primero, un análisis univariable para seleccionar aquellas que incluirán, posteriormente, en la RL (regresión logística) de comprobación. En este caso, encuentran que tanto el sexo femenino como el antecedente de tabaquismo son predictores de trombosis del dispositivo durante el seguimiento. Por otro lado, Vij et al⁴¹ utilizan también el método de RL para la predicción de un único evento, como es la trombosis sobre el dispositivo de cierre de orejuela. Asimismo, realizan también un primer análisis univariable para la selección de pacientes, y hallan como predictores independientes tres variables diferentes a las del estudio previo. Tanto la edad, como los antecedentes de ictus o AIT y el ecocontraste espontáneo en la aurícula izquierda surgen como factores independientes. Incluso, posteriormente realizan una comparación entre los dos grupos de trombosis y no trombosis, emparejados por el puntaje de propensión, determinando que existen diferencias significativas en la profundidad de implante del dispositivo durante el procedimiento. Sin embargo, falta reportar cómo ajusta el modelo y tampoco se realiza una validación externa del mismo.

Esta confusión entre los diferentes tipos de regresión multivariable o multivariante, es frecuente en muchos estudios. Además, en este caso se crearán modelos para realizar un análisis inferencial de la población en función de los patrones detectados en la muestra. Por este motivo, cuándo el objetivo es predecir únicamente la trombosis del dispositivo tras el cierre de orejuela, es importante establecer la diferencia entre ambos, y escoger el adecuado, que este caso sería un análisis multivariable.

Más allá del detalle sobre el tipo de RL, y las diferencias entre los análisis multivariado y multivariable, en ninguno de los dos estudios se hace referencia al poder predictivo de los

1. Introducción

modelos creados. Esto supone que no tenemos información sobre si estos presentan una capacidad predictiva al menos aceptable. Por este motivo, conocer los factores que pueden ser predictores, si no sabemos la potencia que poseen estos modelos, carece de relevancia a la hora de poder predecir eventos futuros.

Para esto existen diferentes herramientas en estadística. Una de ellas es el área bajo la curva ROC (*receiver operating characteristic*). La curva ROC se trata de una representación gráfica de la sensibilidad frente a la especificidad de una prueba. En ella se enfrentan en un eje la sensibilidad frente a 1-especificidad. El área bajo la curva resultante se utiliza muchas veces en medicina como método para valorar la potencia de predicción de un modelo, o el rendimiento diagnóstico de una prueba. Por otro lado, el estadístico c proporciona la misma información y coincide con el área bajo la curva de los modelos de RL. En ambos casos, valores próximos a 0.5 indicarán un pésimo poder predictivo, ya que se asemejará a un modelo azaroso, como puede ser lanzar una moneda de dos caras al aire. Por el contrario, un modelo que obtenga un área bajo la curva próximo a 1.0 aportará una capacidad predictiva excelente. Rangos intermedios pueden ser aceptables, considerando un valor superior a 0.7 como una moderada potencia, y por encima de 0.8 un poder predictivo bueno. Existen además otros coeficientes que se utilizan en estadística en determinados contextos, como el R^2 , que trata de explicar la variabilidad de un factor causada por otro factor independiente. Esta correlación se conoce como bondad de ajuste, y el valor se representa desde 0.0 a 1.0, siendo 1.0 un ajuste perfecto.

Otros estudios han realizado análisis similares, aunque también sin llevar a cabo un análisis sobre la potencia estadística de los modelos resultantes. Fauchier et al⁴² realizaron un análisis multivariable en una muestra de 487 pacientes, detectándose como predictores independientes la mayor edad y el antecedente de ictus o AIT. En contraposición, se establecieron como factores protectores la doble antiagregación o el tratamiento anticoagulante oral al alta. Sorprendentemente, el tratamiento al alta fue analizado en todos los demás estudios, sin

1. Introducción

evidenciarse en ningún otro esta asociación. Por otro lado, Kaneko et al³⁷ utilizaron el registro de Watchman para realizar un análisis multivariable. Sin embargo, en este modelo se incluyeron aquellas variables en las que se encontraron diferencias significativas entre los grupos con y sin trombosis del dispositivo. En este caso, se hallaron la escala CHA2DS2VASc y la profundidad del implante como posibles predictores. Aminian et al³⁶ realizaron un análisis multivariable en el que encontraron solamente la anchura del orificio de la orejuela como predictor independiente. Este es el único que presentaba un carácter prospectivo, aunque se trataba de un estudio observacional también. Simard et al³⁸ realizaron un análisis multivariable en una muestra más equilibrada, en la que incluyeron hasta 237 casos con trombosis del dispositivo frente a 474 controles. Esta muestra más equilibrada podría conseguir una potencia estadística mayor en un evento tan poco frecuente, sin embargo, tampoco en este caso se mostró ningún parámetro que así la valorara. Dukkipati et al⁴⁰ incluyeron en su estudio todos los pacientes de los brazos de intervención de dos ensayos aleatorizados como el PROTECT⁴⁴ y el PREVAIL²⁸, y dos estudios no randomizados, consiguiendo hasta 1739 paciente para su análisis. En el análisis multivariable, se detectaron algunos predictores ya mencionados previamente, como el antecedente de ictus o AIT, o el diámetro de la orejuela, y otros diferentes a los descritos, como la FA permanente o la arteriopatía periférica. Por último, Cepas-Guillén et al⁴⁵, analizaron la profundidad de implantación, tanto de dispositivos con lóbulo y disco como aquellos solamente con lóbulo, en más de 1000 pacientes de diferentes centros. En su estudio encontraron que una profundidad mayor se asociaba de forma significativa con mayor tasa de episodios de DRT (trombosis del dispositivo *-device-related thrombosis-*) documentado, aunque sin un aumento en la tasa de eventos tromboembólicos. Además de una posición más distal de implantación del dispositivo, en el análisis multivariable, encontraron que el tratamiento tras el procedimiento, bien sin ningún tratamiento antiagregante o bien con antiagregación simple, eran predictores de DRT. Sin embargo, en estos dos últimos estudios, como en los previamente mencionados, en ningún momento se realiza un análisis de la potencia predictiva de estos hallazgos.

1. Introducción

Tabla 4. Estudios que analizaron posibles predictores de trombosis de dispositivo de cierre de orejuela.

Estudio	Muestra	Predictores de DRT
Saw et al³⁹	344 pacientes del registro del dispositivo ACP	<ul style="list-style-type: none"> - Sexo femenino - Fumadores
Vij et al⁴¹	Registro de 537 pacientes	<ul style="list-style-type: none"> - Edad - Ictus o AIT previo - Ecocontraste en aurícula izquierda
Aminian et al³⁶	Estudio prospectivo y observacional de 1088 pacientes	<ul style="list-style-type: none"> - Diámetro del ostium
Fauchier et al⁴²	Registro de 487 pacientes	<ul style="list-style-type: none"> - Edad - Ictus o AIT previo
Kaneko et al³⁷	Registro 78 pacientes con implante de Watchman	<ul style="list-style-type: none"> - CHA2DS2VASc - Profundidad del implante
Simard et al³⁸	Estudio de 237 pacientes con DRT vs 474 controles	<ul style="list-style-type: none"> - Hipercoagulabilidad - Derrame pericárdico - Enfermedad renal crónica - Profundidad del implante > 10mm - FA no paroxística
Dukkipati et al⁴⁰	1739 casos de ensayos aleatorizados (PREVAIL ²⁸ y PROTECT ⁴⁴) y estudios observacionales	<ul style="list-style-type: none"> - Ictus o AIT previo - Diámetro de ostium - FA permanente

		- Arteriopatía periférica
--	--	---------------------------

Ante esta falta de homogeneidad en los hallazgos de los estudios publicados hasta la fecha, y todos ellos sin demostrar que se trata de modelos con un alto, o al menos moderado, poder predictivo surge la posibilidad de utilizar los métodos de ML (*machine learning*) para este propósito.

1.3. Inteligencia artificial y *machine learning* o aprendizaje automático

1.3.1. Nacimiento y desarrollo de la inteligencia artificial

En los últimos años se ha producido una masiva digitalización de la sociedad debido fundamentalmente a la evolución de la tecnología. Desde mediados del siglo XX, el desarrollo de las computadoras permitió iniciar el análisis y procesamiento de datos. En los últimos años ha habido un aumento exponencial, tanto de la capacidad de almacenamiento de datos, como de la velocidad de procesamiento de estos. Algunos autores incluso definen esta nueva etapa como la cuarta revolución industrial, la de los sistemas inteligentes⁴⁶⁻⁴⁸. Esto se basa en un aumento de la capacidad de procesamiento de los datos gracias a los nuevos microprocesadores, la capacidad de almacenamiento de datos cada vez mayor, y la facilidad actual para el intercambio de información debido fundamentalmente al desarrollo del internet. Estos aspectos han sido claves en el uso masivo de datos que estamos experimentando en la última década. Ante esta situación, surgen nuevas oportunidades para optimizar el rendimiento de estos datos, como son el *big data*, también llamado inteligencia de datos, y la IA (inteligencia artificial). El fin principal de esta nueva tecnología es extraer la máxima información posible de estos grandes bancos de datos. Además, ofrecen la posibilidad de realizar análisis mucho más complejos y eficientes.

A lo largo del último siglo diferentes definiciones de IA se han postulado. Una de las primeras fue la capacidad de las máquinas para actuar y realizar razonamientos como el ser humano⁴⁹.

1. Introducción

Actualmente, una de las más aceptadas es la capacidad de tomar decisiones y aprender en base a sus resultados^{50,51}.

1.3.2. Ramas de la inteligencia artificial

Dentro de la IA surgen diferentes ramas a lo largo de la historia. La primera de ellas es la que denominamos simbólica. Esta se basa en las reglas lógicas clásicas, ya utilizadas en su época por Aristóteles. Sin embargo, su uso ha ido quedando en un plano secundario con la incorporación de los modelos probabilísticos a la IA. Por otro lado, surgieron también los algoritmos de optimización. Estos tratan de buscar las formas más eficientes para llevar a cabo diferentes tipos de trabajo. Al igual que la IA simbólica, se utilizan preferiblemente modelos híbridos que incorporan técnicas del aprendizaje automático actual.

El aprendizaje automático ML es la rama de la IA que más ha contribuido al auge de esta disciplina, desde el inicio de este último siglo XXI (Figura 6). La principal función del ML es la detección de patrones, aprender a identificarlos y poder clasificarlos en diferentes grupos de datos. Con esta tecnología es posible predecir eventos futuros, gracias a los patrones detectados en los conjuntos de datos y a la información que se le proporciona al algoritmo. Además, el propio algoritmo va mejorando los modelos de predicción conforme aumenta el número de datos incorporados. Dentro de este tipo de IA, se encuentra el aprendizaje profundo (*deep learning*), en el que las redes neuronales son su máximo exponente y gracias a las cuales se pueden realizar análisis mucho más complejos.

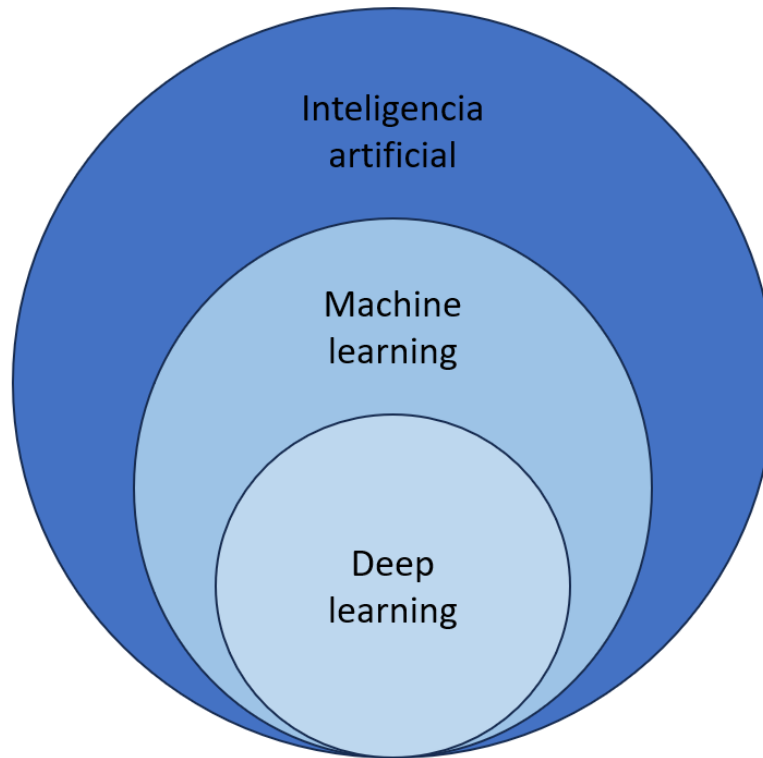


Figura 6. *Inteligencia artificial y su división en las diferentes ramas.*

Dentro de la IA se incluyen diferentes técnicas que puede ser aplicadas en el ámbito sanitario (Figura 7):

- El procesado del lenguaje natural: estas técnicas permiten a las máquinas analizar el lenguaje del ser humano. Para ello además de analizar las palabras, deben aplicar una serie de reglas probabilísticas y así interpretar su significado en función del contexto. Esto puede aplicarse para la búsqueda o extracción de información de documentos clínicos. Dentro de este campo se puede incluir la transcripción por voz, que puede transcribir mensajes orales, o dar voz a aquellos escritos⁵².
- Sistemas de planificación: estos sistemas automatizados permiten generar programas asistenciales más eficientes mediante la secuenciación de los procesos y la optimización de los recursos.
- Análisis predictivos: permiten generar modelos de predicción que consiguen crear planes de salud pública que mejoren los actuales programas de prevención y

1. Introducción

monitorización de la salud. Además, se puede estratificar el riesgo de los pacientes y así poder realizar una medicina más personalizada⁵³.

- Análisis de la imagen: existen ya sistemas que ayudan al diagnóstico tanto en radiología como en anatomía patológica. Estos convierten las imágenes en conjuntos de datos para poder así hacer inferencias y reconocer o diferenciar imágenes⁵⁴.
- Robótica: estas técnicas se están desarrollando en la cirugía, sobre todo para procesos repetitivos y automatizados, consiguiendo que la combinación entre máquinas y seres humanos mejore los resultados ⁵⁵.

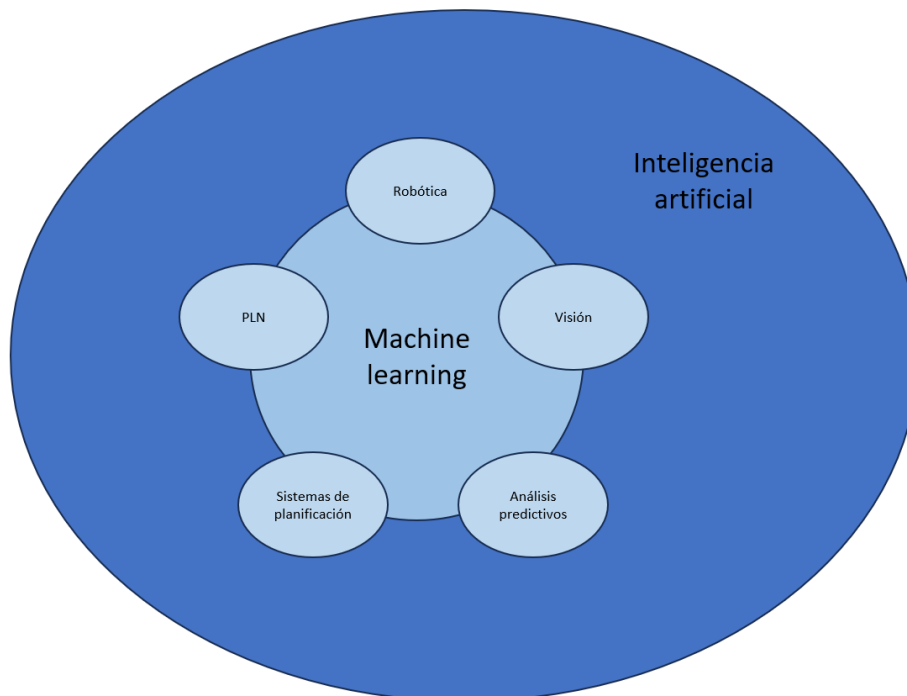


Figura 7. Técnicas aplicadas en la inteligencia artificial. Se exponen las diferentes técnicas que pueden ser utilizados dentro de la inteligencia artificial y, más específicamente, en los procesos de machine learning que se desarrollan en el ámbito sanitario. PLN: Procesado de lenguaje natural.

Todas estas características han provocado un creciente interés para su uso en el ámbito de la medicina. En esta área existe un gran potencial por el gran número de datos generados, sin embargo, no es habitual contar con sistemas de gestión conjunta de estos datos. Registros nacionales o internacionales en el que se incluyan grandes cantidades de pacientes pueden ser

1. Introducción

útiles para implementar diversas técnicas de ML. Estas herramientas son idóneas para optimizar el rendimiento de la atención médica a través de estas grandes bases.

No sólo se puede utilizar para la ayuda en el diagnóstico, sino también para el apoyo en la toma de decisiones en relación con el tratamiento más indicado. Incluso, permite optimizar los recursos, en ocasiones limitados, para el tratamiento de diversas enfermedades. Todas estas particularidades podrían permitir realizar una medicina más personalizada, y un consumo de recurso lo más eficiente posible. Por ello, en los últimos años se ha producido un gran aumento de trabajos en medicina publicados relacionados con la IA. Si se realiza hoy en día una búsqueda en Pubmed con los términos *“artificial intelligence”* o *“machine learning”* se puede ver el crecimiento exponencial que existe desde la última década.

Todas estas características son muy interesantes en el campo de la medicina, especialmente en aquellas enfermedades raras en las que la adquisición de experiencia por parte de los sanitarios es más lenta y complicada.

1.3.3. Tipos de aprendizaje automático

El ML, como se mencionaba previamente, consiste en que las máquinas adopten la capacidad de aprender a partir de observar datos. Para ello se crea un modelo basado en los datos suministrados, que posteriormente se valida con nuevas observaciones. De esta forma se crean los modelos de predicción de forma similar a la estadística inferencial, aunque con una potencia de análisis mucho mayor. Existen diferentes tipos de aprendizaje automático, entre los que se encuentran el aprendizaje supervisado, el no supervisado, el semisupervisado y el aprendizaje por refuerzo.

1.3.3.1. *Aprendizaje supervisado*

Es el más comúnmente usado y desarrollado. En este modelo en primer lugar es necesario asignar a un conjunto de datos una etiqueta. En base a esta etiqueta generada, el algoritmo es capaz de crear un modelo de predicción con estos datos. Este nos proporcionará la probabilidad

1. Introducción

de que cada uno de los nuevos datos que se incluyan pertenezca a la etiqueta previamente seleccionada (Figura 8). Este tipo de aprendizaje está muy extendido y ha sido uno de los principales artífices de la expansión del uso de la IA. Como desventaja, se requiere la intervención de los seres humanos. Para llevar a cabo estos modelos es necesario que se le indique al software cuáles son las etiquetas de cada dato analizado, para poder así crear inferencias y desarrollar el modelo. Por ejemplo, si queremos que el programa consiga analizar diferentes imágenes y distinguir en ellas entre dos frutas, deberemos primero decirle qué etiquetas corresponde a cada dato o imagen, para que así consiga entrenar el algoritmo. Además, en muchas ocasiones requieren de cantidades muy grandes de ejemplos, por lo que el trabajo puede ser muy laborioso.

Estos modelos ya se han utilizado en diferentes ámbitos de la medicina, como es la cardiología intervencionista. Un ejemplo de ello es el artículo de Zweck et al⁵⁶, en el que intentaron predecir el riesgo de mortalidad al año tras la reparación mitral percutánea transcatóter. Para ello utilizaron un conjunto de datos que dividen en diferentes subgrupos para realizar el entrenamiento y la validación externa. En este análisis se utilizó un modelo supervisado, en el que en los datos se incluía la mortalidad al año. En este caso esta era la etiqueta que se aportaba al modelo, y es la probabilidad que se deseaba predecir.

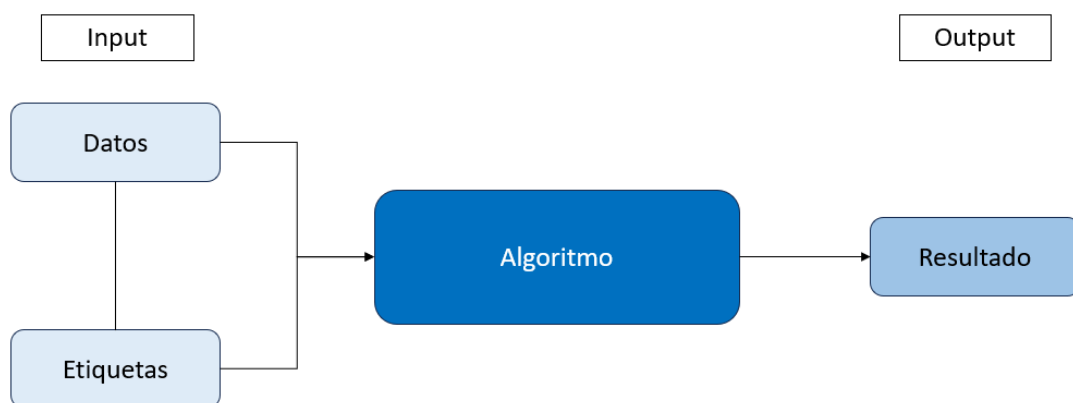


Figura 8. Esquema aprendizaje supervisado.

1. Introducción

1.3.3.2. *Aprendizaje no supervisado*

En este caso, es el propio algoritmo el que por sí mismo es capaz de identificar diferentes patrones, que se repiten en el conjunto de datos, dividiéndolos en diferentes subconjuntos con patrones similares. De esta forma, sin necesidad de asignarle etiquetas, el algoritmo consigue proporcionar la probabilidad de que cada dato pertenezca a un subconjunto determinado (Figura 9). Este tipo puede asimilarse al aprendizaje que realiza un ser humano sin la enseñanza de ninguna otra persona. En este caso, no se le da la información de las etiquetas a las que corresponden los datos, sino que el propio algoritmo, los organizará en diferentes grupos. Este tipo de técnicas son capaces entonces de identificar un grupo con características similares como, por ejemplo, un mismo diagnóstico, sin conocer previamente este dato. En su estudio, Chao et al⁵⁷, utilizan un análisis no supervisado y, mediante su algoritmo, consiguen identificar, con nueve variables ecocardiográficas diferentes, tres clústeres distintos de disfunción diastólica (normal, alteración de la relajación y aumento de presiones de llenado). Esta clasificación, se asoció con un pronóstico determinado en función del grupo al que pertenecían. Este es un claro ejemplo de modelo no supervisado, en el que se identifican diferentes grupos dentro de la base de datos en base a los patrones que detecta el algoritmo, y estos se asocian a un mejor o peor pronóstico en función de ello.

Sin embargo, este tipo de aprendizaje también tiene sus desventajas, ya que puede clasificar los datos de acuerdo con otras características diferentes a las deseadas por el investigador. En algunos casos, podremos al menos definir el número de grupos en los que queremos dividir el conjunto de datos. Por el contrario, tienen las ventajas de ser menos costoso y no precisar del trabajo de un ser humano para etiquetar todos los datos. Esto es muy atractivo para explotar grandes bases de datos que podrían surgir gracias al *big data*.

1. Introducción



Figura 9. Esquema aprendizaje no supervisado.

1.3.3.3. Aprendizaje semisupervisado

Consisten en modelos mixtos que reúnen características de los dos anteriores. En este caso, se tratan de algoritmos en los que no todos los datos presentan etiquetas. Por este motivo, se crea un modelo no supervisado, que utiliza las etiquetas recibidas para retroalimentarse (Figura 10). El número de etiquetas siempre será menor que el número de datos.

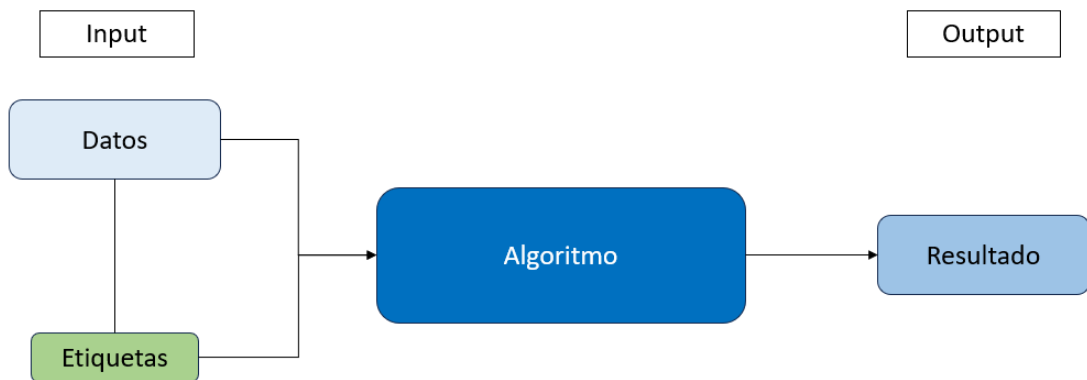


Figura 10. Esquema aprendizaje semisupervisado.

1.3.3.4. Aprendizaje por refuerzo

Este tipo de aprendizaje se caracteriza porque es el propio algoritmo el que, tras obtener resultados con los datos aportados, evalúa su propia respuesta y reajusta el modelo en función de esta (Figura 11). En otras palabras, el propio modelo va mejorando mediante el método de ensayo y error, aprendiendo de los aciertos y los errores que comete. Gracias a estas

1. Introducción

particularidades, son modelos capaces de funcionar con un menor número de datos, lo cual puede ser interesante en ciertas ocasiones.

Liu et al⁵⁸, en su revisión sobre el uso de aplicaciones de apoyo a la decisión en las unidades de cuidado críticos, muestran diferentes estudios en los que se utilizaba este tipo de aprendizaje por refuerzo. Se incluyeron trabajos en los que se empleaba esta tecnología, fundamentalmente para la toma de decisiones sobre las dosis terapéuticas a utilizar o sobre la necesidad de una intervención en paciente críticos .

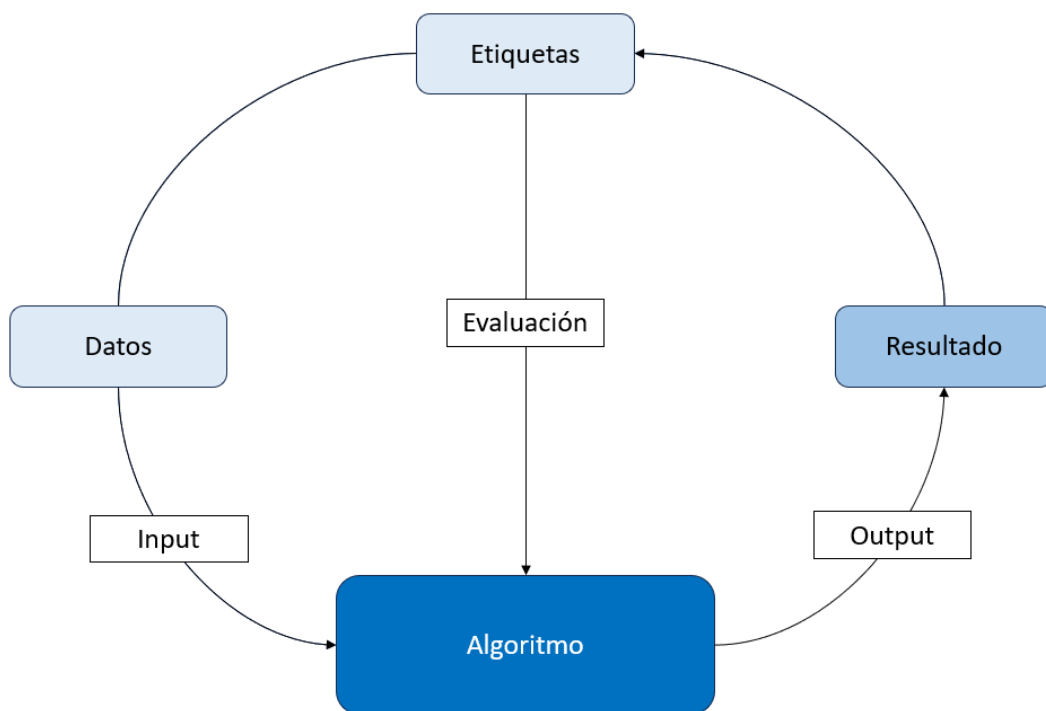


Figura 11. Esquema aprendizaje por refuerzo.

Todas estas características hacen de esta tecnología una herramienta muy interesante en el ámbito sanitario. Esto se debe principalmente a la gran cantidad de datos que se generan en este entorno. Por lo tanto, estos tipos de algoritmos pueden facilitar no solamente la capacidad asistencial sobre los pacientes, sino también contribuir a mejorar o facilitar la investigación médica. Este tipo de tecnología adquiere un gran potencial a la hora de llevar a cabo estudios de investigación, tanto en enfermedades raras, como en eventos raros de patologías más frecuentes que conllevan una repercusión clínica relevante. La evidencia creada sobre estos temas suele ser

1. Introducción

escasa, probablemente por lo compleja y poco rentable que es. Esto se debe principalmente a la dificultad para obtener casos y datos clínicos y, para conseguir modelos de predicción con una potencia estadística lo suficientemente razonable.

Para ello será necesario crear bases de datos fiables y lo más amplias posibles, consiguiendo así sacar el mayor partido de ellas. En el ámbito médico existen numerosas fuentes de datos, sin embargo, estos no están debidamente recogidos y las bases de datos no siempre están fácilmente disponibles. Desde la historia clínica, las imágenes médicas, las pruebas de laboratorio, datos de ensayos clínicos, otros datos aportados por los pacientes o hasta sensores ambientales, todos, son fuentes de datos que pueden ser utilizados para realizar los análisis que deseemos mediante las herramientas del aprendizaje automático.

En este escenario, emerge la historia clínica electrónica como una gran oportunidad para facilitar, tanto la recogida, como el almacenamiento, de todos estos datos que se desprenden de la atención sanitaria de los pacientes, o de las medidas de prevención que se aplican sobre la población. Asimismo, sería de vital importancia homogenizar e integrar los datos para así poder realizar mejores y más potentes análisis. Esto podría ayudar a aportar mayor evidencia sobre diferentes patologías o pruebas, apoyando la decisión médica de los sanitarios consiguiendo una medicina más personalizada y segura.

1.3.4. Tipos de algoritmos

A la hora de realizar un análisis de una serie de datos con ML, debemos seleccionar al inicio el tipo de algoritmo que deseamos utilizar. Cada algoritmo encaja mejor para un tipo u otro de análisis, y no existe ninguno que se adapte perfectamente a un modelo. En función de los datos que se quieran analizar, y de los resultados que se quieran obtener escogeremos unos u otros. Estos se dividen principalmente en: clasificación, regresión, *clustering* o reducción dimensional (Figura 12).

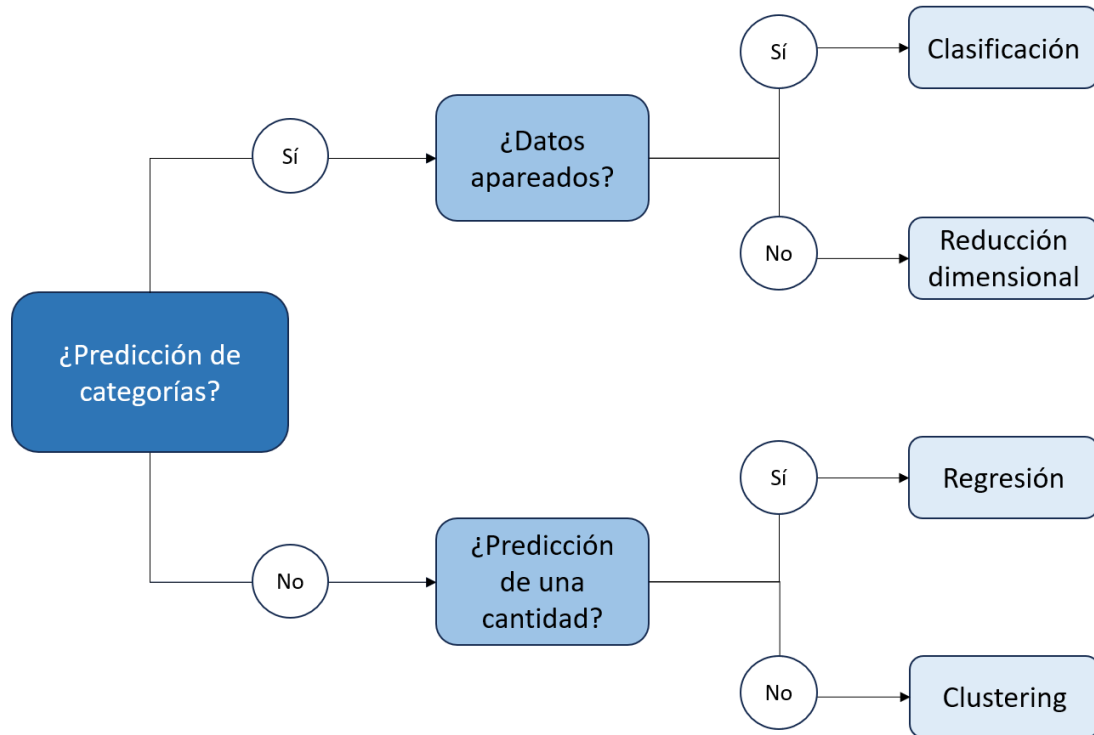


Figura 12. Tipos de algoritmos. Para utilizar las técnicas de machine learning en un proyecto, deberemos seleccionar el algoritmo que mejor se adapte a nuestros objetivos. En este esquema se representan las decisiones que han de tomarse a la hora de esta elección. Si se tratan de variable categóricas deberemos valorar además si son datos apareados. En el caso contrario, deberemos valorar si se trata de cantidades o no.

1.3.4.1. Clasificación

Los algoritmos de clasificación son los que deben utilizarse en el caso de que el resultado que se desee obtener se trate de una variable categórica. El modelo, en función de los datos analizados, nos ofrecerá una probabilidad de pertenecer a esta categoría. Además de ser aquel con el que se tiene más experiencia, hoy en día es el que obtiene los mejores resultados. Existen diferentes subtipos dentro de los algoritmos de clasificación que se describen a continuación. En la RL, el resultado se muestra como una regresión lineal acotada entre los valores 0 y 1, traduciéndose entonces en la probabilidad de pertenecer a dicha categoría. Los árboles de decisión utilizan ramas, nodos y hojas para el análisis, y presentan, al contrario que la RL, un resultado no lineal. Las máquinas de vectores de soporte consiguen la mayor distancia entre diferentes grupos de datos en el hiperplano. En el aprendizaje de conjunto o *bagging* se extraen un número de muestras del conjunto de datos, y con el clasificador obtenido de cada subgrupo se extrae un

1. Introducción

promedio. El *boosting* realiza una clasificación similar, aunque al contrario que el *bagging*, en este caso los análisis se van realizando de forma iterativa, consiguiendo así centrarse en los errores previos y mejorando en cada paso. En definitiva, se dispone de diferentes métodos de clasificación que se deben escoger según su adaptación a los datos que se tengan y a cada modelo específico.

1.3.4.2. Regresión

Por otro lado, los algoritmos de regresión tratan de predecir una variable cuantitativa a partir de las variables del conjunto de datos. El más utilizado es el de regresión lineal, aunque existen diferentes modelos.

1.3.4.3. Clustering

En este caso, el modelo trata de relacionar los datos en subconjuntos. Estos estarán formados por aquellos datos más similares entre sí. Deben seleccionarse cuando la variable a predecir no es ni categórica ni una cantidad. Debido a estas características, son muy utilizados en los análisis no supervisados, ya que tienen la capacidad de distinguir grupos de datos sin conocer las etiquetas previamente. Dentro de estos algoritmos existen diferentes subtipos dependiendo del método para confeccionar los subgrupos lo más homogéneos posibles.

1.3.4.4. Reducción dimensional

Las técnicas de reducción dimensional tienen como objetivo reducir la dimensión de los datos sin perder información relevante. Una de las características más reseñables es que se combinan con los otros tipos de algoritmos previamente descritos. Esta reducción dimensional puede ser lineal o no lineal. En ambos casos, se trata de agrupar los datos en un nuevo sistema de coordenadas. La primera lo hace de forma lineal, y la segunda es más compleja y requiere una potencia computacional mayor, pero ha podido resolver problemas que no podía la reducción dimensional lineal. Entre los más utilizados, se encuentran el t-SNE (*t-distributed stochastic neighbor embedding*) y el UMAP (*Uniform Manifold Approximation and Projection*).

1. Introducción

1.3.5. Construcción de modelos de *machine learning*

A la hora de construir los modelos de ML es necesario llevar a cabo una serie de pasos. En primer lugar, es fundamental establecer cuál es el problema que se requiere abordar, y dejar claro cuál es el objetivo final del modelo. Si no se establecen de forma clara estas bases, será complicado seleccionar de forma eficaz el tipo de algoritmo que se debe utilizar. En segundo lugar, es necesario evaluar los datos que tenemos. El número de casos disponibles, el tipo de variables recogidas, la incidencia del evento a estudio en la muestra, o la calidad de los datos recogidos serán claves a la hora de iniciar el análisis. Por ejemplo, existen diferencias si los datos se han recogido de forma manual, siendo probable que haya errores durante la confección de la base de datos, o si por el contrario se han extraído de forma automática de un registro oficial.

Posteriormente, se procederá a realizar una limpieza de los datos. Comprobar y modificar aquellos datos erróneos puede ser tedioso, pero es imprescindible para obtener un buen resultado. Para empezar, es recomendable realizar un estudio descriptivo del conjunto de variables, para así detectar aquellos valores extraños o alejados del resto, que pudieran comprometer el resultado del algoritmo. Por otro lado, en este tipo de análisis los valores ausentes han de ser rellenados también. Para ello se pueden utilizar diferentes técnicas de imputación o simplemente utilizar la media o la mediana.

Una vez que la base de datos esté correctamente revisada y transformada, se procederá a realizar la construcción del modelo. Sin embargo, antes de comenzar es conveniente valorar si disponemos de una muestra de datos externa que permita realizar una validación fiable del modelo entrenado. En caso contrario, se dispone de distintos métodos para conseguir un buen ajuste del modelo, como por ejemplo dividir el conjunto de datos en varias partes. Una será utilizada para el entrenamiento del algoritmo, otra para llevar a cabo la validación y otra para el test final. La primera servirá para ajustar el modelo, la segunda permitiría mejorar los resultados del modelo, y con el tercer grupo de datos se valorará la calidad del ajuste previo. Esto puede

1. Introducción

ser factible en caso de grandes bases de datos, sin embargo, para otras más pequeñas será necesario realizar alguna técnica de remuestreo. Una de las más conocidas y utilizadas es la validación cruzada o *k-fold*. La muestra se divide en k partes iguales, una de ellas se utiliza para realizar el test, y los grupos restantes sirven para realizar el entrenamiento y la validación. Esto se repite hasta que todos los subgrupos hayan sido utilizados como test. Finalmente, con el resultado de cada algoritmo, se realiza una media que será el resultado final del modelo. Por el contrario, la variabilidad de los resultados se podrá obtener con el análisis de la varianza.

A pesar de realizar todos los pasos de forma correcta, cabe resaltar que se pueden cometer errores durante la creación de los modelos. Estos deben evitarse y detectarse antes de finalizar el análisis. Estos errores, que se pueden producir durante el entrenamiento, son el sobreajuste, o una muy poca reproducibilidad a la hora de realizar una validación externa en otro subgrupo de datos. Además, la selección de los hiperparámetros de cada algoritmo será clave y puede ser necesario probar diferentes combinaciones hasta encontrar la idónea. Esta selección, nunca será perfecta y dependerá del conjunto de datos y de las variables que se vayan a analizar. Esto obligará a optimizar cada modelo de forma específica.

En caso de intentar predecir eventos raros, pero con gran implicación clínica, se deberá priorizar detectar los verdaderos positivos. Esto se consigue primando la exhaustividad. Por el contrario, si la predicción de un falso positivo conlleva unas consecuencias insatisfactorias, se favorecerá un modelo que detecte mejor los verdaderos positivos, anteponiendo una mayor precisión a la exhaustividad. Finalmente, en el caso de que se busque un modelo equilibrado, se utilizará el *F-score* como medida idónea para optimizar el poder predictivo del algoritmo.

En conclusión, es conveniente seguir siempre los pasos imprescindibles, para no cometer errores, y poder realizar un modelo lo más optimizado posible. Esto será específico para cada análisis, y el objetivo será siempre escoger el mejor tipo para cada caso.

1. Introducción

1.4. Educación en aprendizaje automático

1.4.1. Situación de la formación en inteligencia artificial y aprendizaje automático en los trabajadores en el ámbito sanitario

El nivel de conocimiento sobre métodos estadísticos clásicos no es excesivamente profundo en la población general, concretamente en los profesionales del ámbito sanitario^{59,60}. Siendo los conocimientos en IA y ML son todavía menores^{61,62}.

Debido a todas sus potenciales aplicaciones, esta nueva tecnología comienza a ser cada vez más habitual en el entorno sanitario⁴⁹⁻⁵¹. De hecho, en el año 2018 la OMS (Organización Mundial de la Salud) ya predijo que la IA será crucial para poder conseguir sus objetivos⁶³. Además, en los últimos años, las publicaciones relacionadas con plataformas de ML se incrementan progresivamente de forma exponencial^{64,65}.

Todos estos datos apuntan a que el uso de la IA será cada vez más extendido, por lo que será necesario que todos los sanitarios tengan ciertos conocimientos sobre este tema. Incluso, puede ser sorprendente que ya se utilizan diversos modelos en la práctica diaria, a pesar de que aquellos sujetos que los utilizan probablemente no comprendan su funcionamiento. Así como son necesarios amplios conocimientos sobre cálculo, matemáticas o programación para diseñar ciertos protocolos o máquinas, no es necesario un conocimiento tan exhaustivo para poder sacarles todo el rendimiento por parte de los profesionales médicos⁶¹. Cada vez se desarrollan más aplicaciones de la IA que facilitan la atención médica, tanto para ayudar al diagnóstico, tratamiento, estratificación de riesgos o como soporte en el tratamiento.

Estos motivos son los que llevan a plantearse la necesidad de una mayor formación tanto en estadística clásica, como en técnicas de IA como el ML. Sin embargo, uno de los mayores problemas radica en la evidente falta de profesores formados en IA en la educación secundaria o incluso en las facultades de medicina⁶¹. Para poder progresar en el futuro, será indispensable formar a las próximas generaciones desde el inicio. Además, no solo a estos, si no que la rápida

1. Introducción

adopción de estas metodologías, hará que aquellos profesionales que ya se encuentren desempeñando su trabajo precisen formarse también de forma acelerada⁶⁵.

Cuando hablamos de ML tenemos que hablar de forma inherente de grandes conjuntos de datos. El análisis del *big data* se ha adentrado en el entorno sanitario debido al elevado potencial que presenta en algunas ramas como la oncología o la cardiología, ya que son capaces de generar un importante volumen de datos⁶⁶. Esto es una de las claves para conseguir buenos resultados. En la medicina, se da la situación de que existe la posibilidad de recoger múltiples datos sobre cada paciente, siendo un potencial nicho para la IA. Estos datos pueden provenir de la instrumentación médica en la que se incluyen datos como las constantes vitales, la imagen médica de cualquier prueba diagnóstica que se realice, pruebas de laboratorio que pueden aportar infinidad de variables, datos de los ensayos clínicos publicados, aquellos aportados por los pacientes como los síntomas, o de los sensores ambientales que pueden vigilar la polución en cada determinado momento. En este contexto, la historia clínica será siempre una fuente de datos prácticamente inagotable que, eso sí, será imprescindible recoger o registrar para poder analizarla cuando sea necesario.

Esta cantidad de datos, sumado a los modelos ya desarrollados empiezan a aportar información relevante a la práctica clínica diaria. Desde llevar a cabo la monitorización de la salud de los pacientes, gestionar la medicación, prevenir enfermedades, llevar a cabo programas de rehabilitación con robots, detección temprana y diagnóstico más preciso, cirugías rutinarias con mejores resultados, hasta aumentar la seguridad del paciente. Todas estas características que puede aportar la IA nos dirigen a una medicina de precisión y personalizada en la que los resultados pueden ser mucho mejores, con mayores tasas de éxito y menos complicaciones.

Este desarrollo de la potencia computacional que se ha producido en la última década ha provocado una expansión de estas herramientas en todos los sectores, siendo incluso llamada como la cuarta revolución industrial. Es la situación global actual en la que nos encontramos, en

1. Introducción

la que la medicina no debe quedarse atrás. Es el momento de adentrarnos en esta nueva tecnología, cuyo potencial es enorme, sin olvidar también que presenta desventajas no desdeñables que han de tenerse en cuenta a la hora de adoptarla.

Para ello es vital una formación temprana y completa para así poder avanzar en este siglo XXI al mismo ritmo que otros sectores. Primero será necesario realizar una evaluación concienzuda del papel actual de la IA o del ML en el ámbito sanitario. Deberá valorarse cuáles son las infraestructuras disponibles y necesarias para llevar a cabo su implementación. Además, tendrá que llevarse a cabo una formación adecuada de los sanitarios no cualificados. Para saber desde qué punto se parte, será necesario conocer cuál es el nivel actual de conocimientos de esta tecnología. A partir de ahí, se deberán aunar esfuerzos para incrementar el aprendizaje, no solo de los actuales profesionales, sino más importante, de aquellos que todavía están por formarse y constituirán el sector médico del futuro.

Se están empezando ya a diseñar programas para enseñar IA a los alumnos de secundaria y bachillerato, para que en el futuro sean capaces de utilizarla de forma rutinaria y sacarle el máximo rendimiento⁶⁷⁻⁶⁹. Unos jóvenes cualificados encontrarán mucho más sencillo la aplicación de esta tecnología en la práctica clínica. Por otro lado, aquellos que decidan cursar la formación universitaria, deberán ser instruidos en las técnicas de ML aplicadas a las ciencias de la salud. Para ello, los planes de estudio de cualquier facultad deberán adoptar asignaturas sobre estos temas.

Por otro lado, aquellos sectores con más experiencia en el aprendizaje automático, han comenzado a realizar proyectos y estudios para crear plataformas que ayuden a usuarios no experimentados a utilizar o crear modelos a través de ML^{70,71}. No sólo la formación es necesaria, sino en una sociedad extremadamente especializada, facilitar la adopción de estas nuevas herramientas será clave para su expansión.

1. Introducción

2. Justificación

2. Justificación

La IA, y más específicamente el ML, surgen como herramientas que pueden ser muy útiles en el ámbito sanitario. Es posible utilizar esta nueva tecnología para realizar estudios que no son posibles efectuar, o son muy costosos, mediante los métodos estadísticos clásicos. De hecho, en cardiología, las guías de práctica clínica basan muchas de sus recomendaciones en las opiniones de expertos, sin poder sustentarse en ningún estudio. Existe entonces un hueco en el que el ML puede aportar soluciones, ante la falta de evidencia en determinados temas. Muchas veces esto se debe a que se tratan de patologías raras de baja prevalencia, o eventos infrecuentes con alto impacto clínico dentro de patologías más prevalentes.

En cuanto al coste, al principio este puede ser más elevado que en otros estudios, hasta que se consiga dotar de las infraestructuras necesarias para ello. Sin embargo, se podrán realizar análisis retrospectivos y prospectivos, con elevados niveles de computación. Aunque para ello, deberá mejorarse y protocolizarse la recogida y codificación de los datos clínicos y médicos para así disponer de grandes bases de datos con las que trabajar.

La FA es una arritmia muy frecuente en la población mayor en los países occidentales. Sin embargo, el cierre de orejuela es un procedimiento que se lleva cabo a día de hoy en solamente un pequeño porcentaje de los pacientes con este diagnóstico. Además, un grupo muy reducido de paciente a los que se les practica el cierre percutáneo de la orejuela izquierda presentan durante el seguimiento una trombosis del dispositivo implantado. Esta complicación implica una repercusión clínica muy importante, ya que aumenta el riesgo de eventos tromboembólicos, y suele frecuentemente precisar reintroducir la anticoagulación, que muchas veces está contraindicada en estos pacientes. Es por estos motivos por los que es necesario la creación de un modelo de predicción potente, y en la actualidad los modelos estadísticos clásicos no han conseguido hallar predictores consistentes. Por lo tanto, el ML puede ser una herramienta clave a la hora de buscar nuevos modelos que consigan ayudar a predecir aquellos pacientes con mayor riesgo de presentar trombosis sobre el dispositivo de cierre de orejuela. Este evento es poco frecuente, pero tiene una implicación clínica tal que seleccionar aquellos con mayores

2. Justificación

probabilidades podría permitir llevar a cabo un seguimiento más cercano o un tratamiento más agresivo. Este potencial uso del ML está en expansión en los últimos años, y puede ser clave para predecir eventos raros y con ello conseguir crear una medicina más personalizada.

Por estos motivos, surge la posibilidad de aplicar el aprendizaje automático en la trombosis del dispositivo de cierre de orejuela, ya que se trata de un evento muy poco prevalente, pero con un impacto clínico muy relevante. Además, los estudios publicados hasta la fecha son escasos y con limitada evidencia o potencia estadística.

La aplicación de métodos de ML no está todavía sistematizada por parte de los profesionales sanitarios. La infrautilización de la IA dentro de la medicina puede perjudicar a la calidad de la atención sanitaria. Uno de los principales motivos puede ser la escasa formación del sector médico en estos nuevos modelos. Por ello surge la necesidad de crear planes formativos ya desde la educación secundaria o en las escuelas de medicina, siendo necesario en primer lugar evaluar cuál es el nivel de conocimientos de este sector. Por otro lado, parece interesante crear herramientas que faciliten el uso de la IA por parte de personas poco experimentadas. Esto podría ayudar a explotar de forma más eficiente el potencial del ML dentro de este campo. Un uso generalizado ayudaría a mejorar muchas facetas de la atención médica, como planes preventivos, ayudas en el diagnóstico o en la toma de decisiones. Incluso la robótica podría ayudar a mejorar los resultados quirúrgicos respecto incluso a personas muy experimentadas.

3. Hipótesis

3. Hipótesis

Los modelos creados a través de la tecnología de ML aportan más potencia predictiva que los métodos estadísticos convencionales para predecir aquellos pacientes con mayor riesgo de presentar trombosis del dispositivo tras el cierre percutáneo de orejuela izquierda.

La creación y desarrollo de plataformas para facilitar el uso de los modelos de ML por parte de los sanitarios son necesarias. Una mayor formación, unido a aplicaciones o herramientas que faciliten su uso conllevarán un aumento de la implementación de estas nuevas tecnologías con un alto potencial.

4. Objetivos

4. Objetivos

- Aplicación de algoritmos de aprendizaje automático para identificar potenciales predictores de trombosis de dispositivo de orejuela.
- Comparación de la potencia de predicción de trombosis sobre dispositivo de cierre de orejuela entre los métodos estadísticos clásicos y las técnicas de ML.
- Desarrollo de una plataforma para facilitar el uso de modelos de IA en el ámbito sanitario. (Proyecto KOOPA ML).
- Evaluación del conocimiento sobre IA y ML de los profesionales de la salud, y comparar la capacidad de ellos para crear modelos de predicción con herramientas de ML mediante la plataforma KOOPA ML entre dos grupos de participantes no expertos en IA, recibiendo o no recomendaciones durante el proceso.

5. Resultados

5. Resultados

5.1. Aplicación del aprendizaje automático en la medicina. Comparación de la potencia de predicción de trombosis sobre dispositivo de cierre de orejuela entre los métodos estadísticos clásicos y las técnicas de *machine learning*

En primer lugar, se llevó a cabo un experimento en el que se trató de aplicar diferentes modelos de IA a través de ML para conseguir predecir la trombosis sobre el dispositivo de cierre de orejuela. Dada la escasa evidencia publicada previamente, con estudios que no consiguieron encontrar predictores con una potencia predictiva aceptable, surge la posibilidad de intentar conseguir una mejor potencia predictiva con la IA.

5.1.1. Objetivos

El objetivo primario de este estudio era evaluar si existían diferencias en el poder predictivo de trombosis sobre dispositivo de cierre de orejuela entre los modelos de ML y aquellos métodos clásicos. En este caso, queríamos comparar la capacidad predictiva de métodos estadísticos como el análisis multivariable, con el análisis mediante modelos de aprendizaje automático, utilizando algoritmos de clasificación. Estos últimos, tienden a mejorar su eficiencia cuantos más datos se dispongan. Al tratarse de una muestra con gran cantidad de datos, queríamos valorar si en este caso las técnicas de ML podrían aportar algo diferente para la predicción de un evento poco frecuente, pero clínicamente muy relevante, como la trombosis sobre el dispositivo de cierre de orejuela. Además, buscamos las posibles variables predictoras mediante ambos modelos y la analizamos, para posteriormente buscar diferencias entre ellas. Dentro de las variables recogidas, se encontraban características basales de los pacientes y características técnicas del procedimiento, que se consideraban que podrían tener influencia en el desarrollo posterior de dicho evento. Esto presenta gran relevancia, ya que detectar estas variables con poder predictivo sería de gran ayuda a la hora del manejo clínico de estos pacientes. Por último, al no contar con una muestra para la validación externa de los modelos, decidimos revisar

5. Resultados

detalladamente ambas metodologías, aplicando o no técnicas de remuestreo buscando diferencias entre ellas para encontrar posibles predictores. A diferencia de otros estudios previos, en nuestro caso decidimos aplicar la bondad de ajuste en nuestro análisis.

5.1.2. Métodos

Se llevó a cabo un análisis retrospectivo de los datos de una base multicéntrica que incluía 1150 pacientes en los que se había llevado a cabo el cierre percutáneo de la orejuela izquierda. Estos procedían de cinco centros de diferentes países.

Se incluyeron pacientes en los que se procedió al cierre percutáneo de la orejuela izquierda, con un seguimiento ambulatorio posterior. La técnica de cierre de orejuela es muy reproducible, sin embargo, pueden existir diferencias en la misma entre diferentes centros. Normalmente es necesario también realizar un estudio de imagen previo al procedimiento para descartar la presencia de trombo dentro de la orejuela, y para evaluar cómo es la anatomía de esta. Esto se realiza mediante ETE o mediante TC cardíaco. En cada centro se realizó según su protocolo habitual. En cuanto al seguimiento tras el procedimiento también dependía del centro en el que se realizara dicha intervención. En ocasiones los pacientes eran dados de alta el mismo día, o ingresaban en el hospital durante al menos un día. El seguimiento posterior se realizó, en cualquier caso, de forma ambulatoria, con nuevos estudios de imagen de control, bien ETE o bien TC cardíaco. Este se realizó en al menos en una ocasión durante los primeros tres meses tras el procedimiento, para poder identificar cierres incompletos por fugas peridispositivo significativas, o para descartar trombosis sobre el dispositivo. La frecuencia de estos controles, tanto clínicos como con pruebas de imagen, durante el seguimiento posterior, variaba según el hospital.

5.1.2.1. Definiciones

La definición de trombosis sobre dispositivo que se utilizó es la propuesta por Aminian et al en su artículo³⁶. Esta se definió como una imagen de mayor densidad que la sangre, que no se

5. Resultados

explique por ningún artefacto o endotelización normal del dispositivo. Esta debía verse en diferentes planos, bien en el ETE o en el TC, y encontrarse en contacto con el dispositivo. A pesar de establecer estos criterios diagnósticos de trombosis del dispositivo, cabe mencionar que en cada centro las pruebas de imagen eran realizadas y evaluadas por diferentes operadores, con el posible sesgo que ello conlleva.

5.1.2.2. Recogida de datos

Por otro lado, se recogieron de forma retrospectiva las diferentes variables, tanto basales como del procedimiento y del seguimiento. La selección de estas variables se realizó según su relevancia clínica y estadística para la predicción de trombosis del dispositivo tras el cierre percutáneo de la orejuela. Se incluyeron todos los pacientes de cada centro en los que se había realizado previamente el cierre de orejuela, independientemente de si presentaron durante el seguimiento trombosis del dispositivo. Esta recogida de datos se llevó a cabo de forma independiente en cada centro a partir de sus bases de datos correspondientes. Finalmente se conformó una base mayor conjunta que reunía todos estos datos y es la que se utilizó para el análisis. Todo el proceso de recogida de datos se realizó cumpliendo con los criterios de la Declaración de Helsinki de 1975.

5.1.2.3. Diseño

Se realizaron cuatro experimentos diferentes divididos en método clásico o ML, y en aquellos que se aplicaron las técnicas de remuestreo o no. Así, se establecieron cuatro experimentos diferentes resumidos en la Figura 13.

5. Resultados

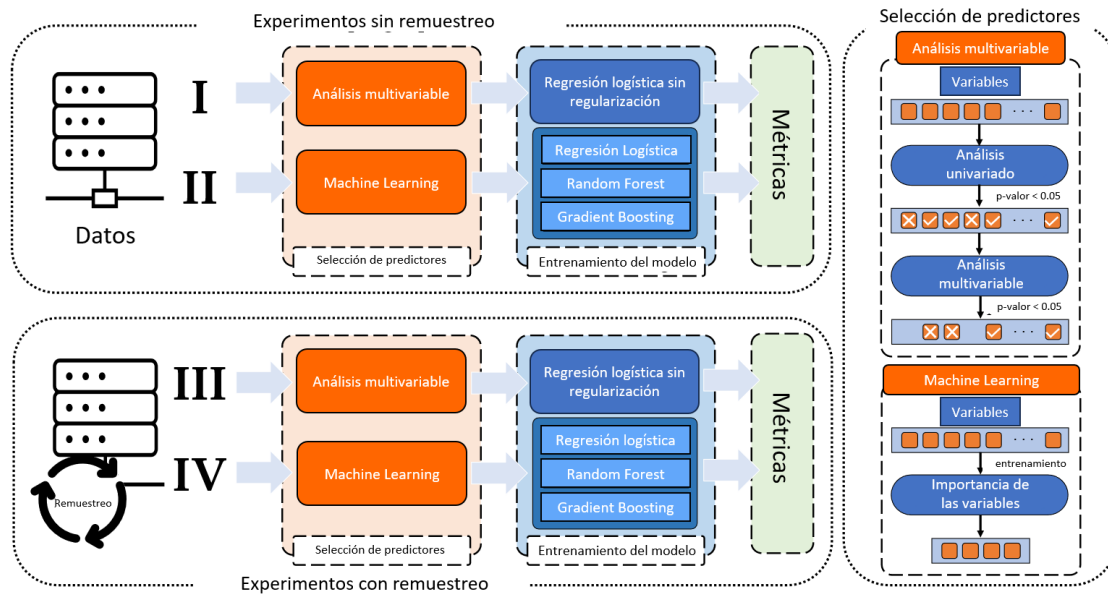


Figura 13. Diseño y flujo de trabajo de los experimentos.

Se realizó el análisis multivariable incluyendo toda la muestra en el experimento I, sin llevar a cabo técnicas de remuestreo. Por el contrario, en el experimento III, sí que se realizaron técnicas de remuestreo como el *k-fold*, con $k=2$ y 5 repeticiones, creándose así hasta 10 modelos diferentes. En cuanto a los experimentos II y IV se realizaron clasificaciones como la RL, el RF (*random forest*) o el GB (*gradient boosting*), utilizándose en el experimento IV las mismas técnicas de remuestreo mencionadas previamente.

5.1.2.4. Análisis multivariable (experimentos I y III)

Para la creación de modelos de predicción mediante el método estadístico clásico se escogió el análisis multivariable. En este caso es el método seleccionado ya que el objetivo es predecir únicamente la trombosis del dispositivo tras el cierre de orejuela.

Para el análisis multivariable hemos realizado primero un análisis univariable de cada una de las variables seleccionadas. Dependiendo del tipo de variable se realizó su correspondiente prueba. Dado que se trataba de una muestra grande, para las variables continuas se realizó bien el t-test o el ANOVA, en función de si seguían una distribución normal o no respectivamente. Para variables categóricas se utilizó la prueba de chi cuadrado. Por último, para las variables ordinales

5. Resultados

se utilizó el test U de Mann-Whitney. Una vez realizado el análisis univariable, se seleccionaron únicamente aquellas variables que resultaron significativas (p -valor $< 0,05$) a la hora de predecir la trombosis sobre el dispositivo. Con este paso se consigue reducir de forma considerable el número de variables que se incorporarán posteriormente al análisis multivariable. Esto es de vital importancia en caso de bases de datos con un número elevado de variables, ya que simplifica posteriormente el análisis. El siguiente paso fue confeccionar el modelo de regresión multivariable, para así evaluar si las variables previamente seleccionadas se comportaban verdaderamente como predictores de trombosis de dispositivo de cierre de orejuela. Se estableció de nuevo un p -valor $< 0,05$ como criterio de referencia para la significación.

5.1.2.5. Método de machine learning seleccionado (experimentos II y IV)

Dentro del ML existen muchas metodologías. En este caso, para los experimentos II y IV, se decidió elegir un método ampliamente utilizado para la selección de variables predictoras. Este se basa en el peso de cada variable dentro del modelo, conocido como clasificación basada en el peso.

En primer lugar, se realizó un análisis y evaluación de los datos disponibles. Posteriormente se procedió a la limpieza de estos, y a rellenar los valores faltantes. Una vez se completó la revisión de la base de datos, se procedió a realizar la clasificación con diferentes modelos como la RL con regularización, el RF y el GB. Una vez obtenidos los resultados, se reclasificaron según el método de clasificación basado en el peso.

Este método se basa en realizar ajustes en función del peso de cada variable. Una vez el modelo se ha definido y ajustado, la selección de las variables predictoras dependerá del peso de cada una de estas en el modelo. Se establece entonces un punto de corte, por debajo del cual aquellas variables con un peso inferior serán descartadas y se retirarán del modelo.

En segundo lugar, cabe destacar que en los modelos de ML, al igual que en cualquier tipo de análisis estadístico, es de vital importancia la calidad de los datos utilizados, no solamente la

5. Resultados

calidad del modelo o análisis realizado^{72,73}. El uso de datos mal balanceados puede llegar a producir errores de predicción, o el fenómeno previamente mencionado de sobreajuste⁷⁴. Este proceso precisa de un remuestreo o bien de una validación externa. Disponer de un subconjunto de datos externo e independiente es lo ideal, ya que utilizar esta muestra permite que nuestro modelo sea más generalizable. Sin embargo, esto muchas veces no es posible en los estudios médicos, ya que la falta de datos o bases con un número limitado de casos no permite este modelo.

Las técnicas de remuestreo fueron creadas principalmente para evitar los sesgos que se objetivaban en las muestras y que arrastraban errores hasta el final de los experimentos, alterando así gravemente los resultados. Asimismo, dividir la muestra en dos subconjuntos, uno para el entrenamiento y otro para el test, puede conllevar a ciertos sesgos azarosos, especialmente en muestras pequeñas.

Por estos motivos, y por el hecho de que la trombosis sobre el dispositivo de cierre de orejuela es un evento raro, se llevaron a cabo experimentos tanto incluyendo técnicas de remuestreo como sin ellas. Para solucionar estos problemas se aplicaron técnicas de remuestreo de validación cruzada. Entre ellas se eligieron el *k-folds*⁷⁵ con repeticiones y *shuffle split*⁷⁶.

La técnica de *k-folds* trata de dividir la muestra en *k* diferentes subconjuntos del mismo tamaño siempre que sea posible. Posteriormente, se utilizan *k-1* subconjuntos para el entrenamiento y creación del modelo de predicción y, con el subgrupo restante se realiza el test. Este proceso se repite *n* veces aplicando el remuestreo cada vez.

El método de *shuffle split* se encarga de generar un número de subgrupos previamente definido, que serán utilizados posteriormente para el entrenamiento y para el test respectivamente. Primero se aleatoriza la muestra y seguidamente se divide en diferentes subgrupos.

Las variables predictoras fueron escogidas en base a un criterio de referencia común, en este caso mediante el análisis del área bajo la curva ROC.

5. Resultados

5.1.2.6. Experimentos realizados

Dado que el objetivo primario era comparar los métodos estadísticos clásicos con los modelos de IA, se establecieron dos líneas diferentes de pruebas. Una de ellas se enfocaba el análisis multivariable y la otra los métodos de ML.

Los métodos de ML seleccionados fueron la RL con regularización^{77,78}, el RF⁷⁹, GB⁸⁰. Por el contrario, para realizar el análisis multivariable se utilizó el método clásico de RL, pero sin regularización.

Todos los predictores fueron reportados con su p-valor correspondiente. Además, para mostrar la potencia predictora, se escogieron parámetros como el área bajo la curva ROC, la especificidad y la sensibilidad. El punto de corte utilizado en cada experimento se determinó dependiendo de la optimización de cada curva ROC.

Sin técnicas de remuestreo

En estos casos, se realizó en análisis en la muestra de datos completa, creándose a partir de esta cada modelo.

Con técnicas de remuestreo

Dada la baja incidencia de la trombosis sobre el dispositivo de orejuela durante el seguimiento de estos pacientes, se realizaron también modelos con estas técnicas de remuestreo. Para ello se utilizó la técnica de validación cruzada o *k-fold*, estableciendo en cada caso una división aleatoria en dos subconjuntos, reservando un subconjunto de un 40% del tamaño de la muestra para la realización del test. Este procedimiento se repitió cinco veces en cada caso. Con este método, se asegura que se incluyan en cada subgrupo un número aceptable de paciente con el evento a estudio.

5.1.2.7. Software utilizado

Se utilizó Python como lenguaje de programación, y también se utilizaron las librerías *scikit-learn*⁷⁶, *statsmodels*⁸¹, and *Scipy*⁸².

5. Resultados

El código desarrollado es de libre acceso y puede consultarse en el siguiente enlace:

https://github.com/IACardiologia-husa/LAAC_Thrombus_detection_MLvsClassical⁸³.

5.1.3. Resultados

En primer lugar, cabe describir brevemente las características de los datos utilizados para este estudio. Del total de 1150 pacientes incluidos en nuestra base de datos, solamente se incluyeron en el análisis 813 pacientes finalmente. Esto se debe a que se excluyeron aquellos con demasiados valores omitidos, y aquellos que no completaron un seguimiento adecuado porque era imposible establecer si presentaron el evento a estudio o no. Las características basales de estos pacientes se incluyen en la Tabla 5. La trombosis sobre el dispositivo de cierre durante el seguimiento se observó en 35 casos, que corresponde al 4.31%.

Tabla 5. Características basales. Valores mostrados como media±SD o % (n/N).

Variables	Resultado (N = 813)
Edad (años)	75,62±8,43 (813)
Sexo (masculino)	61,01 (496/813)
Peso (Kg)	74,96±15,55 (806)
Altura (cm)	164,73±9,41 (804)
IMC (kg/m ²)	27,55±4,94 (804)
SC (m ²)	1,81±0,21 (804)
HTA (Sí)	87,33 (710/813)
Diabetes (Sí)	33,7 (274/813)
Fumador (Sí)	14,37 (127/696)
Ictus isquémico previo (Sí)	35,36 (279/789)
Ictus hemorrágico previo (Sí)	24,8 (185/746)
Embolismo sistémico previo (Sí)	9,85 (78/792)
Enfermedad vascular periférica (Sí)	20,54 (167/813)
Enfermedad coronaria previa (Sí)	31,89 (258/809)
Infarto de miocardio previo (Sí)	20,08 (154/767)
Cirugía valvular mitral previa	2,70 (22/813)
CHADS2	2,96±1,24 (768)
CHA2DS2VASc	4,44±1,54 (813)
HAS-BLED	3,71±1,05 (811)
Episodios de sangrado previo	1,57±1,15 (768)
INR lábil (Sí)	11,07 (90/813)
Tratamiento previo con AAS (Sí)	30,26 (246/813)
Tratamiento previo con inhibidores ADP (Sí)	11,07 (90/813)
Tratamiento previo con antiVitK (Sí)	23,37 (190/813)
Insuficiencia mitral grave (Sí)	15,23 (122/801)
Estenosis mitral significativa (Sí)	0,26 (2/765)
Dispositivo reposicionado (si disponible)	0,59±1 (570)

5. Resultados

Contraste durante el procedimiento (mL)	134,03±88,09 (695)
Derrame pericárdico periprocedimiento (Sí)	5.9 (48/813)
Inhibidor ADP al alta (Sí)	62,64 (508/811)
AAS: ácido acetilsalicílico, ADP: receptor adenosina difosfato, HTA: hipertensión arterial, IMC: índice de masa corporal, INR: ratio internacional normalizado, SC: superficie corporal	

Como se mencionaba previamente sobre los estudios publicados, no solo es importante establecer qué factores eran predictores independientes de la trombosis del dispositivo. Además de esto, es necesario averiguar cuál es el poder predictivo de los modelos creados, ya que entonces la relevancia clínica de los hallazgos variará considerablemente. Los resultados, se presentarán en dos diferentes apartados. En primer lugar, se expondrá el poder predictivo de los experimentos realizados y, posteriormente, se esclarecerán cuáles han sido las variables predictoras.

5.1.3.1. Poder predictivo

Para valorar el resultado es necesario evaluar y exponer el poder predictivo de cada modelo de los experimentos realizados. Para ello, se utilizaron las métricas del área bajo la curva de las curvas ROC, la sensibilidad y la especificidad.

En la tabla 6 se exponen los resultados de los cuatro experimentos realizados, ya descritos previamente.

Tabla 6. Resultados de los cuatro experimentos. *El ajuste de las medidas se realizó mediante la optimización.*

Experimento	Remuestreo	Selección de la herramienta	Modelo	Área bajo la curva ROC	Sensibilidad	Especificidad
I	-	Análisis multivariable	RL*	0.5456	0.8857	0.2198
II	-	Según el modelo	RL	0.7974	0.6857	0.7776
			RF	1.0	1.0	1.0
			GB	1.0	1.0	1.0
III	<i>Shuffle split</i>	Análisis multivariable	RL*	0.4387 ± 0.0904	0.1811 ± 0.1143	0.6544 ± 0.2263
	<i>k-fold</i>	Análisis multivariable	RL*	0.5174 ± 0.0531	0.4318 ± 0.3255	0.5504 ± 0.3173
IV	<i>Shuffle split</i>	Según el	RL	0.4838 ±	0.1697 ±	0.6960 ±

5. Resultados

		modelo		0.0118	0.0724	0.0109
			RF	0.3989 ± 0.0240	0.0 ± 0.0	0.9014 ± 0.0639
			GB	0.4614 ± 0.0437	0.0444 ± 0.0459	0.9217 ± 0.0073
	k-fold	Según el modelo	RL	0.5325 ± 0.0349	0.2804 ± 0.0771	0.7634 ± 0.0518
			RF	0.4250 ± 0.354	0.0418 ± 0.0462	0.7049 ± 0.0193
			GB	0.4893 ± 0.0668	0.0754 ± 0.0501	0.9005 ± 0.0176
GB: <i>Gradient Boosting</i> , RL: Regresión logística, RL*: regresión logística sin regularización, RF: <i>Random forest</i>						

Los resultados sugieren que la capacidad predictiva en todos los experimentos es limitada. En el experimento I se objetiva un claro desequilibrio entre la sensibilidad y la especificidad. Una buena sensibilidad permitiría en este caso detectar casi todos los casos con trombosis sobre el dispositivo, pero a costa de una alta cantidad de falsos positivos. Sin embargo, el área bajo la curva ROC ronda el azar al ser próximo a 0.5. Contrariamente, el experimento II presenta una predicción perfecta (área bajo la curva ROC 1.0) en los algoritmos no lineales como el RF y el GB. Este resultado ha de ser evaluado con recelo, ya que los modelos están claramente sobreajustados, por lo que carecerán de validez externa alguna.

En los experimentos III y IV se realizaron las técnicas de remuestreo mencionadas previamente, tanto el *k-fold* con $k=2$ y con cinco repeticiones y el *shuffle split* con dos divisiones y con un subconjunto para el test del 40%. Además, al contrario de los experimentos I y II, se acumularon los resultados de las repeticiones anteriores. En el experimento III, el mejor resultado se obtuvo mediante el remuestreo con *k-fold*. En el experimento IV, el área bajo la curva mayor fue también con remuestreo con *k-fold* y utilizando el modelo de RL. Sin embargo, ambos resultados distan mucho de un modelo con gran poder predictivo.

De acuerdo con estos resultados, el remuestreo de *k-fold* y el modelo de RL fueron los métodos con mejor resultado en los experimentos III y IV, y también el más coherente en el experimento

5. Resultados

II, ya que los otros dos modelos no lineales presentaban un claro sobreajuste. Por estos motivos, estas técnicas fueron las utilizadas para los siguientes subanálisis.

En la Figura 14 se muestran los resultados de cada repetición en los experimentos III y IV. No se encontraron diferencias significativas entre los valores medios de ambos experimentos. Cabe destacar que en el experimento III los resultados son más dispares y variables en las distintas repeticiones, aunque en el IV la sensibilidad obtenida en todas las repeticiones es repetidamente baja.

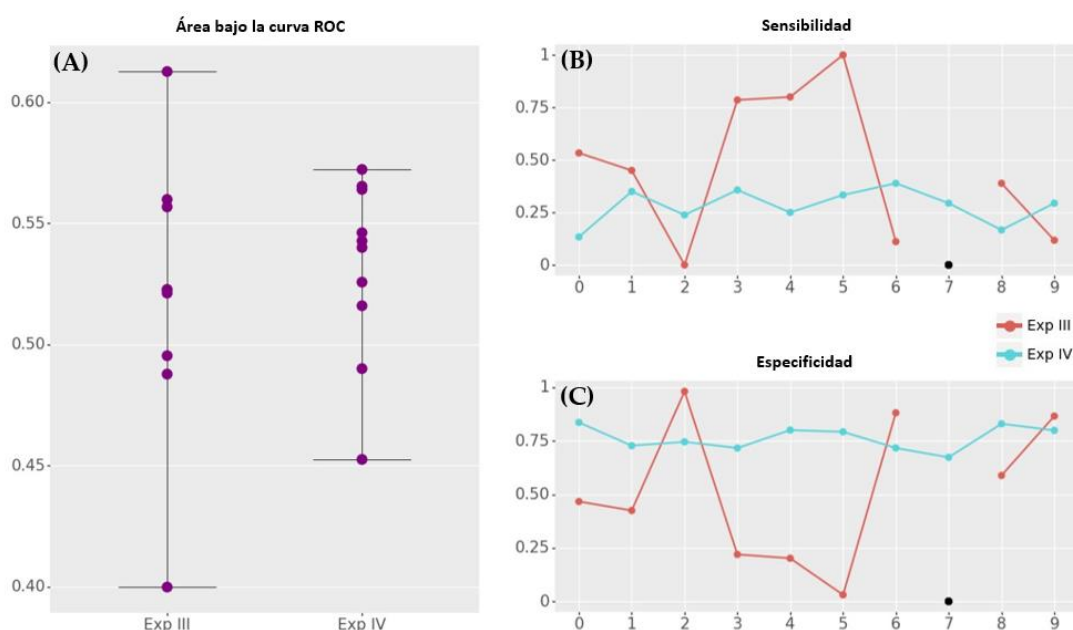


Figura 14. Resultados del modelo de regresión logística. Área bajo la curva ROC (A), sensibilidad (B), y especificidad (C) del modelo de regresión logística con remuestreo con el método k-fold (k=2, 5 iteraciones)

5.1.3.2. Variables predictoras

Las variables que resultaron como predictoras independientes de trombosis de dispositivo de cierre de orejuela de forma significativa tanto en el análisis univariable como el en multivariable (Experimento I) se exponen en la tabla 7. Del total de 76 variables analizadas, únicamente 3 resultaron significantes en el análisis univariable. De este modo, estas tres variables son las que se incluyeron en el análisis multivariable, resultando únicamente el puntaje HAS-BLED como predictor independiente.

5. Resultados

Tabla 7. Selección de variables predictoras en el experimento I. *Resultados del análisis univariable y multivariable*

Variable	Univariable p-valor	Multivariable p-valor
CHADS2	< 0.01	0.852
CHA2DS2-VASC	< 0.01	0.414
HAS-BLED	< 0.01	< 0.01

En la figura 15 se representan las variables que fueron seleccionada por el método de ML de RL durante el experimento II. En ella se pueden apreciar el peso, tanto positivo como negativo, que presenta cada variable en el modelo para la predicción del evento en estudio.

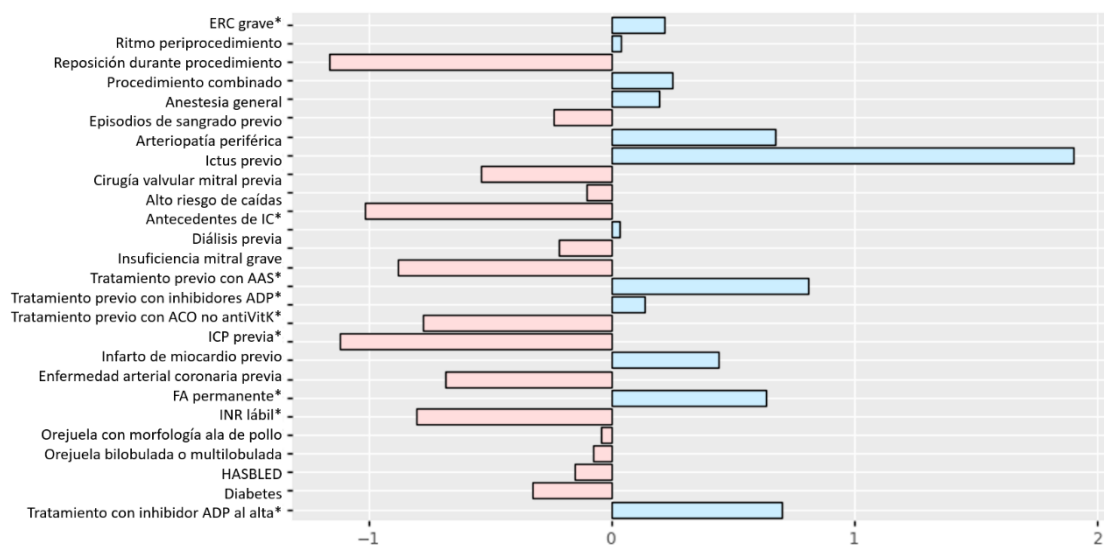


Figura 15. Peso de cada variable en el experimento II al utilizar la regresión logística. AAS: ácido acetilsalicílico, ACO: anticoagulación oral, ADP: receptor adenosina difosfato, ERC: enfermedad renal crónica, IC: insuficiencia cardíaca, ICP: intervención coronaria percutánea, INR: ratio internacional normalizado, FA: fibrilación auricular.

En la tabla 8 se representa el número de veces que las variables coinciden como predictores de trombosis sobre el dispositivo en todos los análisis realizados durante los experimentos III y IV. En total se tratan de hasta 10 análisis, ya que como se comentó, se utilizó un remuestreo *k-fold* con $k=2$, con cinco repeticiones por cada uno. Por este motivo el máximo es 10 y el mínimo posible es 0. En particular, en el experimento III se especifican ambos resultados, tanto en el análisis univariable como en el multivariable.

5. Resultados

Tabla 8. Coincidencia como predictores de trombosis sobre dispositivo de cierre de orejuela en cada método. Resultados para un remuestreo con la técnica de k-fold ($k=2$, 5 iteraciones).

Variable	Experimento III		Experimento IV
	Univariable	Multivariable	
Enfermedad coronaria previa	-	-	10
HASBLED	10	8	10
Tratamiento previo con antiVitK	-	-	10
CHADS2	10	-	6
CHA2DS2VASc	10	-	6
Orejuela con morfología ala de pollo	-	-	9
Insuficiencia mitral grave	1	1	9
Episodios de sangrado previo	-	-	9
INR lábil	-	-	9
Inhibidor ADP al alta	-	-	8
Reposición del dispositivo	-	-	8
ERC grave	-	-	8
Derrame pericárdico periprocedimiento	1	-	8
Diabetes	-	-	7
Ictus isquémico previo	-	-	7
Tratamiento previo con inhibidor ADP	-	-	7
Tratamiento previo con AAS	-	-	7
Embolismo sistémico previo	-	-	7
Arteriopatía periférica	-	-	6
Infarto de miocardio previo	-	-	6
Ictus hemorrágico previo	-	-	6
Volumen de contraste durante procedimiento	1	1	-
Cirugía valvular mitral previa	1	1	5

AAS: ácido acetilsalicílico, ADP: receptor adenosina difosfato, ERC: enfermedad renal crónica, INR: ratio internacional normalizado.

En el experimento III, se objetivan resultados similares a los hallados en el experimento I. Las escalas CHADS2, CHA2DS2VASc y HASBLED aparecen de forma consistente en todo los análisis univariables. Incluso, cuando posteriormente se aplica el multivariable, únicamente se mantiene el puntaje HASBLED como predictor independiente, como en el experimento previo. Por otro lado, en el experimento IV realizado con algoritmos de ML se detectan solamente tres variables de forma consistente como predictores. Estas son los antecedentes de enfermedad coronaria, la escala HASBLED de nuevo, y el tratamiento previo con fármacos antiVitK. El resto de las variables

5. Resultados

aparecen de forma inconsistente como predictoras, alguna incluso solamente en la mitad de los modelos.

Para analizar la fluctuación con el remuestreo tanto del p-valor de los métodos clásicos, como del peso en el modelo de los métodos de ML, es recomendable estimar un predictor de la variación. Este se ha definido en cada caso por la siguiente fórmula:

$$\text{ratio de variación del predictor} = \frac{V_i - V_{ME}}{\sigma} \quad (i = \text{iteración}, ME = \text{muestra entera})$$

V equivale a los p-valores en los métodos estadísticos clásicos, y a los pesos en los algoritmos de ML.

El numerador de la fórmula se trata de la diferencia entre el valor en cada iteración y el valor obtenido en toda la muestra de datos durante el experimento. El denominador es la desviación estándar de los valores obtenidos para la variable predictora en cada una de las diferentes iteraciones.

En la figura 16 se muestran los resultados de la ratio de variación de cada variable seleccionada. Para valorar si existen diferencias significativas entre las dos metodologías de los experimentos III y IV, se utilizó la prueba de U de Mann-Whitney, precedida por el test de Levene. Este último fue significativo (p-valor = 0.039), mientras que el primero no lo fue (p-valor = 0.937).

5. Resultados

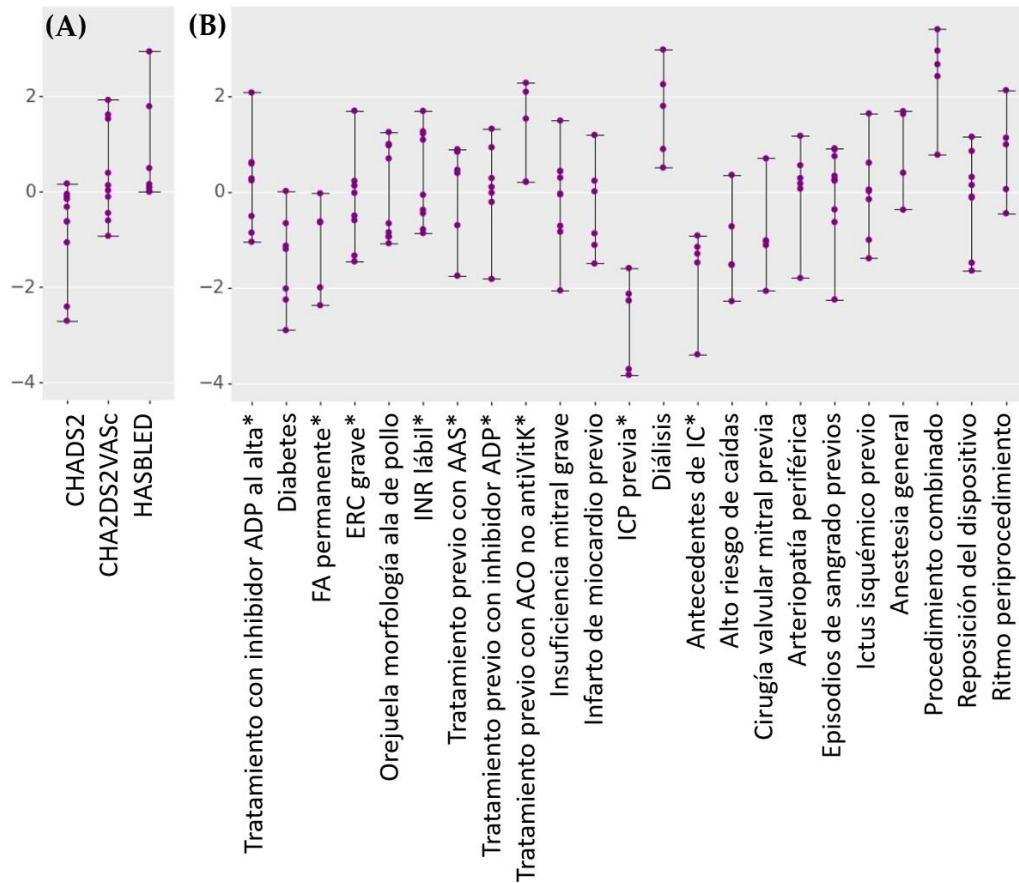


Figura 16. Ratio de variación del predictor en cada remuestreo en el experimento III (A) y en el IV (B) utilizando la regresión logística. Ambos experimentos realizados utilizando el remuestreo con el método de k -fold ($k=2$, 5 iteraciones). AAS: ácido acetilsalicílico, ACO: anticoagulación oral, ADP: receptor adenosina difosfato, ERC: enfermedad renal crónica, IC: insuficiencia cardíaca, ICP: intervención coronaria percutánea, INR: ratio internacional normalizado, FA: fibrilación auricular.

5. Resultados

5.2. Evaluación del conocimiento de los profesionales de la salud y desarrollo de una plataforma para facilitar el uso de modelos de inteligencia artificial en el ámbito sanitario. Proyecto KOOPA ML

En los últimos años se han desarrollado diferentes aplicaciones en las que los métodos de IA son la base para su uso como ChatGPT (<https://openai.com/blog/chatgpt/>) o DALL-E (<https://openai.com/dall-e-2/>). Estas aplicaciones de uso fácil e intuitivo han conseguido expandir más el uso y el conocimiento sobre los modelos de aprendizaje automático. No obstante, no todo son aspectos positivos, ya que también surgen nuevos dilemas éticos sobre su uso en la investigación biomédica⁸⁴. Es imposible destacar estas nuevas herramientas sin reseñar el papel clave de internet para que esto se produjese. Internet ha conseguido que la accesibilidad a estos proyectos sea muy sencilla y poco costosa, por lo que prácticamente toda la población, al menos de los países desarrollados, tiene las mismas opciones para su uso. La expansión de herramientas con IA hace que sea de vital importancia una correcta implementación de estos métodos, para lo cual es necesario una formación adecuada que, a ser posible, debería impartirse a todos los niveles y en todos los ámbitos, con mayor énfasis en las personas más jóvenes. Las ventajas de su introducción y utilización se comparten en diferentes sectores como pueden ser el de la educación o el sanitario. En este último, como se ha mencionado previamente, esta tecnología podría ayudar no solamente a establecer riesgos y pronósticos de los pacientes, sino a ayudar en el diagnóstico y el tratamiento de los mismos. Esta ayuda en la toma de decisiones podría conseguir reducir el número de errores o negligencias que se producen y mejorar los resultados tanto de calidad como de esperanza de vida.

Por estos motivos, consideramos que la formación en IA y aprendizaje automático es muy importante para los profesionales del ámbito sanitario, ya que actualmente el conocimiento de esta tecnología es limitado. Conocer y saber implementar estas herramientas podrá ayudarles a

5. Resultados

conseguir un mayor aprovechamiento de estas y así poder mejorar la atención a los pacientes como fin último^{85,86}.

En los últimos años, el número de plataformas destinadas a mejorar o facilitar el uso de la IA y, concretamente, el ML se han incrementado de forma exponencial^{64,87}. Estas aplicaciones abarcan un amplio rango de potenciales usuarios, desde aquellas que proporcionan librerías para poder desarrollar tus propios algoritmos para personas con conocimientos más avanzados, hasta otra destinadas a usuarios poco entrenados o formados que precisarán incluso la información más básica para poder iniciarse en este sector.

En este contexto surge la plataforma KoopaML, desarrollada por el grupo de trabajo de Vázquez-Ingelmo y García-Peñalvo desde la Universidad de Salamanca⁸⁸. Consiste en una aplicación que trata de ayudar a los sanitarios a formarse en el ámbito de IA, para la creación y aplicación de modelos de ML⁸⁹. Se trata de una interfaz sencilla no sólo para formar sobre las herramientas del ML a sanitarios no expertos en IA, sino también para ayudar a estos en la creación de nuevos algoritmos propios que pueden ser de gran utilidad en la práctica clínica diaria. KoopaML tiene fundamentalmente dos objetivos:

- Ofrecer una plataforma con una interfaz fácil, sencilla e intuitiva para los profesionales que decidan iniciarse en el mundo de la IA
- Ofrecer formación en forma de experiencia que les permita entender el funcionamiento de estas técnicas, y así como las potenciales aplicaciones que presenta.

Esta plataforma se encuentra todavía en desarrollo con el objetivo de que los usuarios sin experiencia con la IA o el ML consigan sacarle el mayor rendimiento posible. Debido a la falta de formación y familiaridad con esta nueva tecnología, se realizó un estudio en el que se incluyeron diferentes profesionales del Servicio de Cardiología del Hospital Universitario de Salamanca con los objetivos de:

5. Resultados

- 1- Evaluar una de las principales características del KoopaML mediante la realización de varios ejercicios para aplicar métodos de ML a una base de datos sencilla.
- 2- Entender las posibles dificultades a las que se pueden enfrentar las personas sin conocimientos profundos en IA a la hora de intentar realizar un modelo de predicción con aprendizaje automático.
- 3- Evaluar si pequeñas recomendaciones en formato de texto pueden facilitar la realización y creación de estos modelos.

Es necesario que potenciales usuarios prueben este tipo de aplicaciones para poder desarrollarlas. Esto ayuda a crear las primeras versiones beta, encontrar errores y mejorarlas. En este caso, es fundamental que personal sanitario sin conocimiento pruebe la plataforma para evaluar si esta es intuitiva y fácil de usar o, por el contrario, su uso supone un reto para el usuario inexperimentado.

5.2.1. KoopaML

- Descripción: KoopaML se trata de una aplicación en línea que ayuda a crear modelos de predicción a usuarios no experimentados en IA, mediante una interfaz sencilla e intuitiva. Dispone de una serie de nodos que están asociados a una tarea específica, como la subida del conjunto de datos, completar los valores faltantes, los modelos de clasificación o regresión elegidos y las métricas seleccionadas para la evaluación del algoritmo (Figura 17). Una vez finalizado el modelo con todas sus tareas, se puede ejecutar obteniendo los resultados del algoritmo.
- Recomendaciones y manejo ensayo-error: una de las principales y más reseñables características es el manejo y aprendizaje mediante el método de ensayo-error, ya que los usuarios aprenden mientras crean nuevos modelos con ML. Cada paso que se realice producirá una actualización del proceso de creación del algoritmo, y de las recomendaciones que aparecerán para poder continuar con el trabajo propuesto.

5. Resultados

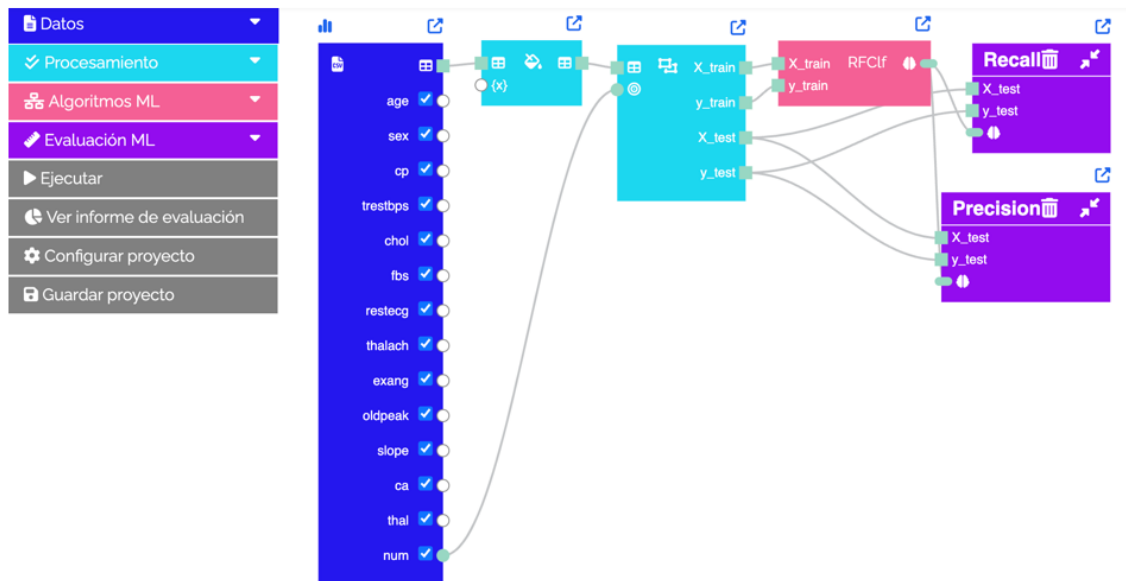


Figura 27. Ejemplo de un flujo de trabajo con KoopaML. En la parte izquierda se observa un desplegable con todas las tareas que se pueden realizar. En la parte central y derecha se observa el espacio de trabajo, donde se exponen los nodos y se conectan entre sí para confeccionar el algoritmo.

5.2.2. Métodos

Para poder evaluar la usabilidad de la plataforma hemos empleado el método *think-aloud*^{90,91}, ya que nos permite perfectamente identificar dónde se encuentran los problemas los participantes durante las pruebas. Esto nos ayudará a poder mejorar la plataforma, incidiendo más específicamente en aquellas tareas dónde, mediante este método, encontremos más dificultades. Para llevar a cabo la recopilación de esta información se procedió a grabar en vídeo a los participantes para que posteriormente fueran evaluados de forma concienzuda. Por último, esta forma de evaluación nos permitía capturar la interacción del usuario con la plataforma durante todos los ejercicios. Así nos mostraba un mapa de calor de toda la pantalla para saber en qué zonas se detenía más el participante. Esto ayudaría también a mejorar próximas versiones de la plataforma. Para poder llevar a cabo este análisis de los mapas de calor de los usuarios durante los ejercicios hemos utilizado una aplicación denominada Hotjar.js (<https://www.hotjar.com/>).

5. Resultados

5.2.2.1. Participantes

Se incluyeron un total de 8 profesionales sanitarios que trabajaban en el Servicio de Cardiología del Hospital Universitario de Salamanca. Creemos que este es un número correcto para aplicar el método de *think-aloud*, ya que la n recomendada por Nielsen⁹² para este método era de 4±1 participantes. De esta forma se consigue extraer de forma exitosa toda la información obtenida durante la grabación, los problemas a los que se enfrentaron cada uno de ellos y cómo consiguieron o no resolverlos.

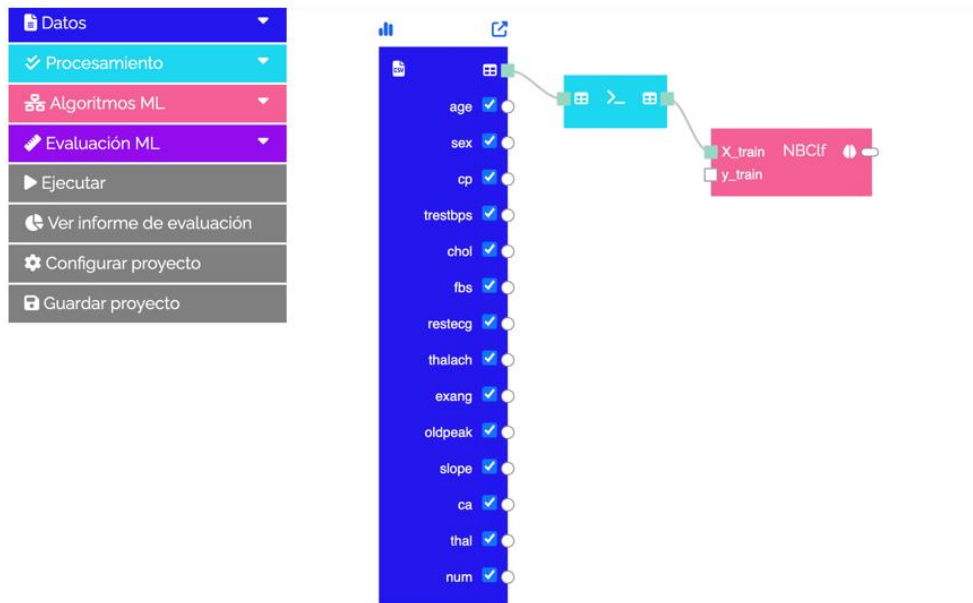
En la tabla 9 se resumen las características de los participantes, incluido el nivel de conocimientos sobre ML previo a su participación y el grupo en el que se incluyó (con o sin recomendaciones adicionales). El nivel de conocimientos previos se dividió en cuatro grupos: el alto, con conocimientos y experiencia en ML; el medio, con pocos conocimientos (sabe qué es la IA pero sin experiencia); el bajo, para aquellos en los que el conocimiento del ML es mínimo; y el grupo con aquellos sin conocimientos algunos sobre aprendizaje automático. De una forma equilibrada, se procedió a repartir a los participantes según el nivel de conocimientos en los grupos a comparar. El primero utilizaría la plataforma básica, y el segundo utilizaría la misma plataforma pero que con recomendaciones para facilitar los pasos correctos que hubieran de seguir (Figura 18). Este reparto ayudaba a mejorar la comparación entre los dos grupos a estudio, con o sin recomendaciones.

Tabla 9. Sanitarios participantes del estudio y sus conocimientos sobre inteligencia artificial.

Participante	Área de trabajo	Nivel de conocimientos en ML	Recomendaciones adicionales
P1	Médico residente	Alto	No
P2	Médico adjunto	Medio	Sí
P3	Enfermera de investigación	Bajo	Sí
P4	Médico residente	Bajo	No

5. Resultados

P5	Médico residente	Bajo	No
P6	Médico adjunto	Ninguno	Sí
P7	Médico residente	Ninguno	No
P8	Médico residente	Ninguno	Sí




💡 For a machine learning algorithm, it is important that there are no missing values in the training set. Insert the 'Fill Missing Values' node and connect it to the dataset you want to process through the connections containing the icon .

Figura 18. Recomendaciones adicionales. En la parte inferior de la imagen se observa un ejemplo de las recomendaciones que recibía un grupo de los participantes durante los ejercicios.

5.2.2.2. Protocolo del estudio

Se realizó una entrevista en línea mediante videoconferencia de forma programada con cada uno de los participantes, llevadas a cabo por la misma persona para evitar sesgos. Una vez efectuadas todas las entrevistas, todos los componentes del equipo efectuaron una evaluación de cada uno profesionales sanitarios poniéndose en común posteriormente. Se realizaron entonces los siguientes pasos:

5. Resultados

- 1- Se presentó al paciente el estudio y se les instó a compartir su pantalla durante la videoconferencia para que pudiera ser grabada y posteriormente analizada.
- 2- A todos los participantes se les enseñó un vídeo en el que se explicaban las principales características de la plataforma, así como los conceptos esenciales sobre ML. Aquí se incluyeron nociones sobre el entrenamiento y la validación posterior de estos modelos, las diferencias entre clasificación y regresión y cuáles eran los pasos necesarios para llevar a cabo la creación del algoritmo deseado.
- 3- Apartado de preguntas y respuestas sobre los dos primeros epígrafes.
- 4- Los participantes accedieron a la plataforma y se registraron.
- 5- Se compartió a través del chat de la videoconferencia el archivo con los datos (formato CSV) y se explicaron las tareas que debería hacer cada uno de los candidatos.

Para realizar el análisis se escogió una base de datos abierta y fácilmente manejable. Este conjunto de datos incluye más de 300 pacientes, que incluyen 14 diferentes variables sobre el poder predictivo de las técnicas no invasivas para predecir enfermedad coronaria. Se escogió esta base de datos por lo sencilla que es, y porque se trata de un tema familiar para los sanitarios participantes en el estudio, todos con un trabajo relacionado con la cardiología. El objetivo principal del estudio era validar el uso de recomendaciones en la plataforma, por lo que esto permitiría facilitar la comprensión de los datos, en participantes con escaso conocimientos en ML.

Se les pidió realizar dos ejercicios diferentes que se resumen en la tabla 10.

Tabla 10. Ejercicios y tareas propuestos a los participantes.

Ejercicio 1 (Task 1: T1)	
T 1.1	Importar el conjunto de datos
T 1.2.	Visualizar el conjunto de datos
T 1.3	Dividir el conjunto de datos en dos partes. Usar el 80% para el

5. Resultados

	entrenamiento y el 20% para el test del modelo
T 1.4	Entrenar un modelo de Regresión Logística
T 1.5	Obtener el resultado del área bajo la curva ROC del modelo
Ejercicio 2 (Task 2: T2)	
T 2.1	Importar el conjunto de datos
T 2.2	Visualizar el conjunto de datos
T 2.3	Conectar
T 2.4	Dividir el conjunto de datos en dos partes. Usar el 80% para el entrenamiento y el 20% para el test del modelo
T 2.5	Entrenar un modelo de <i>Random Forest</i>
T 2.6	Obtener el resultado del <i>recall</i> y la precisión del modelo

A pesar de que en el vídeo inicial se explicaron los diferentes tipos de algoritmos, se decidió especificar qué tipo de ejercicios deberían utilizar en cada caso. Para facilitar la tarea se escogieron dos algoritmos fáciles de usar y encontrar en la plataforma como RL y RF. Para calcular el poder predictivo de los modelos en cada ejercicio se escogieron métricas conocidas utilizadas en el pasado. Se emplearon métricas como el área bajo la curva ROC, *recall* o precisión, que se han utilizado también de forma clásica para los análisis de supervivencia.

Finalmente, los participantes fueron grabados en vídeo durante su participación. De esta forma, se les instaba a que fuera comentando las dificultades que se iban encontrando durante los ejercicios. Estos vídeos fueron posteriormente analizados por el equipo investigador para extraer toda la información de la interacción de los sujetos con la plataforma, y así exponer los resultados y sacar conclusiones.

5. Resultados

5.2.3. Resultados

Para exponer los resultados que nos hemos encontrado durante el estudio analizaremos los hallazgos en cada tarea y subtarea de forma individual para detectar los problemas que se encuentran los participantes. Estos se muestran a continuación.

5.2.3.1. Subida y visualización de los datos

La primera parte de cada tarea consistía en subir el conjunto de datos a la plataforma y poder visualizarlos. Para ello era necesario, desde el desplegable que se encuentra a la izquierda del escritorio, proyectar el nodo correspondiente (Figura 19). Este nodo permite importar la base de datos desde el sistema en el que se trabaje como archivos CSV. Estos datos, corresponderán a los *inputs* que el modelo tendrá para realizar el análisis. En la parte derecha se proyectan las columnas del archivo que corresponden a las variables a estudio. Estas se pueden conectar con los diferentes nodos de forma individual a través de los círculos o de forma global a través del cuadrado superior. Por último, los datos se pueden visualizar en el icono superior de la izquierda.

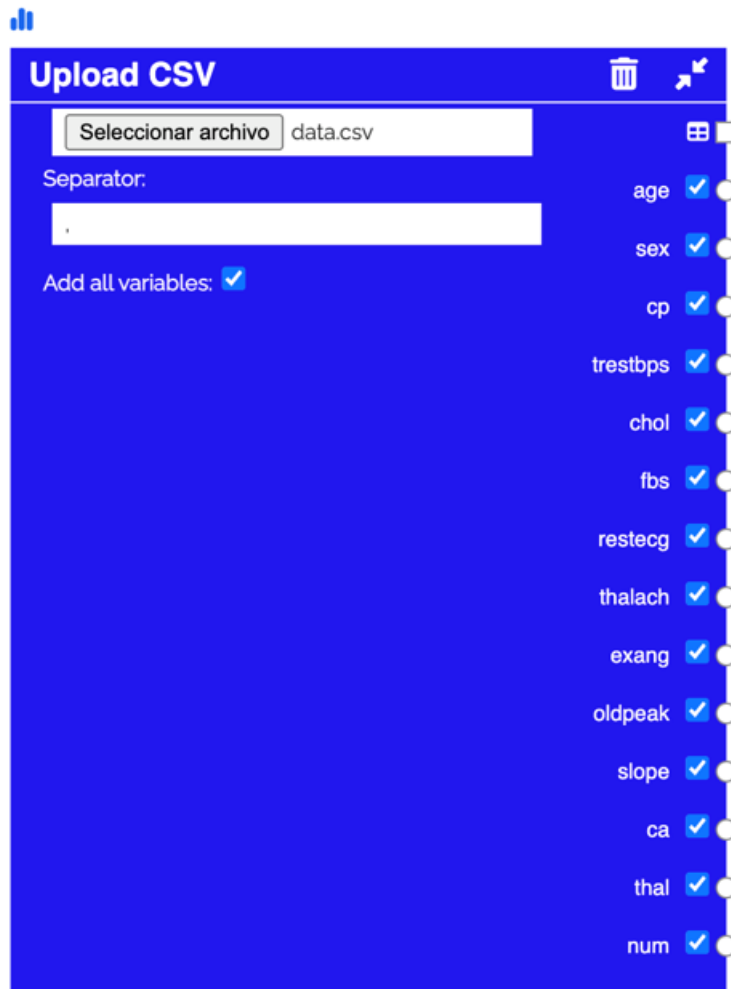


Figura 19. Nodo correspondiente a la subida de datos.

Tanto en el primer ejercicio como en el segundo todos los participantes fueron capaces de importar el conjunto de datos (T1.1 y T2.1). Sin embargo, cabe destacar que algunos de ellos necesitaron bastante tiempo para poder encontrar dónde y cómo proyectar el nodo correspondiente en el espacio de trabajo. Por el contrario, la mayoría no fueron capaces de visualizar los datos (solamente 3 de 8 lo consiguieron), tanto en el primer ejercicio como en el segundo. Además, a pesar de poder visualizarlos, la interpretación de estos resultó costosa y difícil.

5.2.3.2. División de los datos para el entrenamiento y el test del modelo

La siguiente tarea consistió en dividir el conjunto de datos en dos subgrupos que se utilizarían de forma respectiva para el entrenamiento y el test posterior del modelo. El nodo utilizado para

5. Resultados

esto posee dos conexiones como *inputs* en la parte izquierda (Figura 20). La parte superior cuadrada corresponde a la base de datos y la inferior circular a la variable diana del estudio. Además, en el apartado “% test” se puede configurar que parte del conjunto de datos se destinará al test. En lado derecho se pueden observar los outputs del nodo que serán conectados a los nodos restantes, y que se dividen en los del entrenamiento y los del test, tanto para los datos (X_{train} y X_{test}) como para las etiquetas de estos (Y_{train} y Y_{test}).

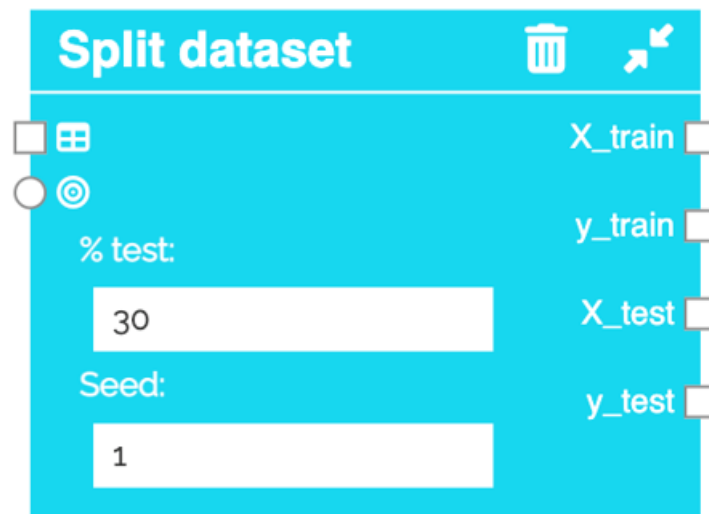


Figura 20. Nodo para la división de los datos en el subconjunto para entrenamiento del modelo y otro para el test posterior.

Durante el ejercicio 1 en la mayoría de los casos (5 de 8 participantes) la división en subconjuntos (T 1.3) se realizó de forma exitosa. Por otro lado, uno de ellos se saltó directamente el paso y prosiguió a la siguiente tarea. A pesar de ello, sí que se encontraron ciertas dificultades durante este proceso:

- Dos participantes consiguieron conectar las variables de forma correcta, pero lo hicieron tras muchos intentos.
- Uno de los participantes no conseguía encontrar el nodo. Sí lo encontraba en el desplegable pero no lo localizaba en el escritorio. Una vez lo encontró, completó la tarea.

5. Resultados

- Dos de los participantes en un primer instante no incluyeron este nodo pasando directamente al de clasificación, pero posteriormente se dieron cuenta y pudieron completarlo sin mayor problema.

En el ejercicio 2, los participantes consiguieron mejorar los resultados del primero (T 2.4).

- Siete participantes consiguieron incluir el nodo.
- Tres de ellos consiguieron completar la tarea de forma exitosa.
- Dos consiguieron realizar la conexión, pero quitaron la selección de las variables del nodo anterior.

5.2.3.3. Algoritmo

Todos los nodos de los algoritmos de ML tienen una estructura similar (Figura 21). En este caso, los inputs del modelo aparecen en la izquierda, con conexiones tanto para los datos del entrenamiento del modelo, como de las etiquetas de estos. Los resultados del modelo se pueden descargar en forma de archivo (formato *Pickle*) y se puede conectar a los nodos de evaluación.

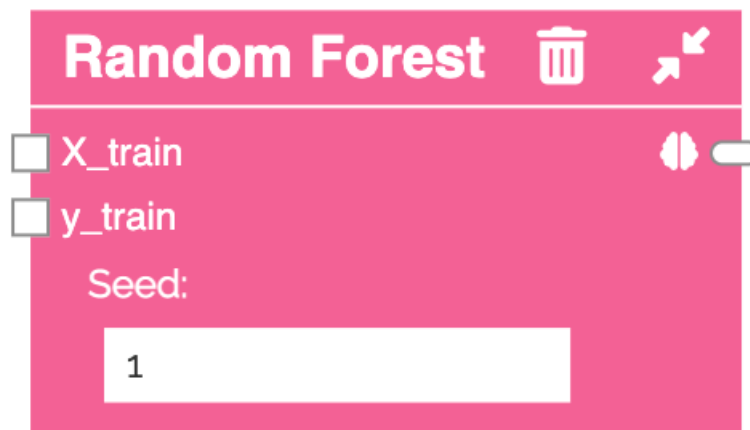


Figura 21. Nodo del algoritmo seleccionado para confeccionar el modelo, en este caso el Random Forest.

En este caso, en el ejercicio 1 se les indicó llevar a cabo una RL como método de clasificación. A pesar de haberse explicado que se debería realizar una clasificación, la mayoría de los participantes (6 de 8) tuvieron problemas para poder llevarla a cabo. Incluso muchos de ellos (5 de 6) buscaban inicialmente el nodo en el apartado desplegable de regresiones. Esto llama la

5. Resultados

atención, y es probable que se deba a un error de concepto por falta de formación en aprendizaje automático. En todos estos casos el término regresión conllevó a una mala interpretación de lo que se pretendía que analizaran. Por el contrario, en el ejercicio 2, en general no obtuvieron apenas problemas para encontrar el nodo requerido de RF.

A pesar de finalmente encontrar los nodos en cada una de las tareas (T 1.4 y T 2.5), los participantes no fueron capaces de entrenar el modelo. El principal problema que referían durante el ejercicio era que no sabían cómo utilizarlo o entrenarlo, incluso aquellos a los que se le habían incluido recomendaciones.

5.2.3.4. Validación

Finalmente, los nodos de validación también presentan una estructura parecida entre ellos, con leves diferencias en función de la métrica seleccionada para ello (Figura 22). En la parte izquierda, al igual que el resto de los nodos, se encuentran las conexiones con los *inputs*. En este caso dispondremos de aquellos para conectar al subgrupo reservado para el test (tanto los datos como las etiquetas) y para la conexión con el modelo más abajo (mismo símbolo del cerebro que en el nodo del modelo).

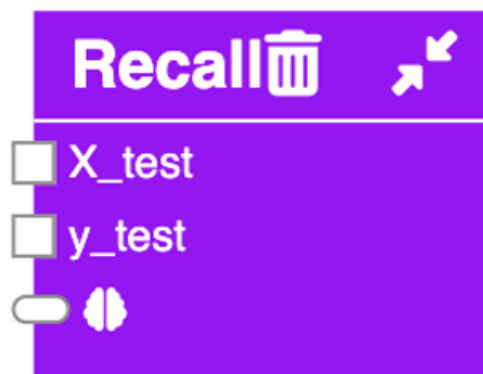


Figura 22. Nodo para la validación del modelo.

La correcta validación de ambos modelos en los ejercicios 1 y 2 (T 1.5 y T 2.6) solamente fue completada por uno de los participantes. El resto se encontraron con diferentes problemas que no les permitieron finalizar esta tarea.

5. Resultados

En la mayoría de los casos (5 de 8) los errores cometidos previamente conllevaron que no fuera posible completar esta última tarea. El resto presentaron problemas a la hora de encontrar los nodos de validación (2 de 8) o de conectarlos con los nodos anteriores (1 de 8). Analizando todos estos errores, se debían fundamentalmente a la falta de conocimiento sobre diferentes conceptos en relación con ML, lo cual producía que acumulasen y arrastrasen errores durante todo el ejercicio.

Finalmente, hay que señalar que no se encontraron diferencias entre ambos grupos con y sin recomendaciones a la hora de completar esta tarea.

5.2.3.5. Comentarios finales

Por último, conviene también mostrar otros detalles sobre la ejecución de los ejercicios. En primer lugar, cabe reseñar que en 3 de 8 participantes los mensajes de error que aparecían cuando cometían errores les ayudaron a completarla. Por el contrario, a 4 de 8 de los participantes, los mensajes de errores en la ejecución no les ayudaron a completar estos ejercicios. No obstante, el moderador intervino en algún caso para ayudarles, pero los sujetos no consiguieron finalizar las tareas por no comprender bien qué se les pedía, como así expresaron durante el ejercicio y posteriormente al finalizarlo.

Además, en el segundo ejercicio se pidió llevar a cabo una tarea extra, rellenar los valores faltantes en el conjunto de datos (T 2.3). Ningún participante llevó a cabo este apartado. Probablemente, esto se deba a que no se entregó una base de datos con valores ausentes, por lo que entonces la plataforma no emitió ningún mensaje de error que hiciera recapacitar a los participantes y proceder a rellenarlos.

Respecto a la usabilidad de la plataforma, cabe destacar que los participantes se encontraron dificultades a la hora de encontrar diversos nodos en el listado desplegable de la izquierda, o en el momento de conectar los elementos de la derecha de los nodos.

5. Resultados

Por otro lado, ninguno de los participantes llegó a consultar al final los resultados del modelo. Este detalle parece reseñable, aunque cabe decir que esa tarea no se solicitó en ningún momento dentro de los ejercicios.

Por último, no se han observado diferencias notables entre los dos grupos que recibieron recomendaciones o que realizaron los ejercicios sin ellas. Sin embargo, los que sí se objetivó es que los que recibían los consejos o cualquiera de ellos cuando recibían los mensajes de errores, conseguían terminar las tareas mejor que los otros. Esto se reflejó de manera más clara en el segundo ejercicio, debido a una mayor familiarización con la plataforma tras el primer ejercicio.

En la tabla 11 se muestran los resultados de cada participante ordenados de mayor a menor conocimiento previo sobre ML. Como se puede observar, las recomendaciones aportadas no parecen influir en el resultado. Al contrario, lo que sí se relaciona con un mayor éxito en la realización de los ejercicios es el nivel de conocimientos previos sobre IA.

Tabla 11. Resultados de las tareas completadas por los participantes en función del grupo.

Participante	Grupo	Ejercicio 1 (tareas completadas)	Ejercicio 2 (tareas completadas)	Problemas
P1	NR	5/5	6/6	-
P2	R	1/5	5/6	T1.3
P3	R	4/5	6/6	T1.3
P4	NR	2/5	6/6	T1.3
P5	NR	1/5	2/6	T1.3-2.4
P6	R	1/5	1/6	T1.3-2.4
P7	NR	1/5	1/6	T1.3-2.4
P8	R	1/5	5/6	T1.3-2.4

Participantes ordenados en función de sus conocimientos en IA de mayor a menor. Grupos: NR, sin recomendaciones; R, con recomendaciones. Problemas: tareas en las que los participantes se atascaron durante la realización de los ejercicios.

6. Discusión

6. Discusión

6.1. Aplicación del aprendizaje automático en la medicina. Comparación de la potencia de predicción de trombosis sobre dispositivo de cierre de orejuela entre los métodos estadísticos clásicos y las técnicas de *machine learning*

La trombosis sobre dispositivo de cierre de orejuela continúa siendo objetivo de estudio ya que se trata de uno de los retos más desafiantes en este campo. El proceso de endotelización del dispositivo tras el implante no se conoce profundamente todavía. Actualmente, se utilizan diferentes protocolos que varían desde la doble antiagregación, la anticoagulación con fármacos antivitaminas K, anticoagulantes de acción directa o incluso alguno de estos dos últimos en combinación con la antiagregación simple. Por este motivo, se ha analizado de forma prospectiva en un estudio no aleatorizado el tratamiento idóneo los primeros meses tras la intervención⁹³. Así mismo, están en marcha diferentes ensayos clínicos aleatorizados en los que se evalúa el tratamiento óptimo para favorecer la endotelización y disminuir así el riesgo de trombosis sobre el dispositivo. Algunos de ellos como el estudio ADALA⁹⁴ con número EudraCT 2018-001013-32 (estudio randomizado que compara dos estrategias de tratamiento antitrombótico: doble antiagregación vs apixaban) o el estudio ANDES con código NCT03568890 (*'Short-Term Anticoagulation Versus Antiplatelet Therapy for Preventing Device Thrombosis Following Left Atrial Appendage Closure'*, que compara también la anticoagulación con fármacos de acción directa vs tratamiento con doble antiagregación), podrán aportar más evidencia al respecto. Esto es importante no solo para evitar la trombosis del dispositivo, sino también porque la mayoría de los pacientes presentan un riesgo de sangrado muy elevado, por lo que es prioritario reducir el tiempo de tratamiento lo máximo posible. En consonancia con un tratamiento antitrombótico más adecuado, también los nuevos dispositivos y materiales ayudan a mejorar la endotelización temprana y disminuir el riesgo de trombosis. Prevenir esta complicación, o por lo menos conseguir seleccionar aquellos pacientes con mayor riesgo de presentarla es crucial, ya que

6. Discusión

permitirá realizar un tratamiento más personalizado. Evitar tratamientos antitrombóticos prolongados que aumenten el riesgo de sangrado es de vital importancia, ya que con pacientes cada vez más añosos, cualquier episodio de sangrado comporta un importante impacto pronóstico.

Diversos estudios han tratado de detectar predictores de trombosis del dispositivo tras el procedimiento, con resultados dispares. Hasta ahora todos realizaron estos análisis mediante los métodos estadísticos clásicos. Simard et al³⁸ propusieron tanto predictores mayores como menores. Entre los mayores se encontraban el derrame pericárdico periprocedimiento y un estado de hipercoagulabilidad diagnosticado previamente. Como factores predictivos menores propusieron la profundidad del implante (más de 10 mm desde el borde del ligamento del Marshall), la enfermedad renal crónica o la presencia de FA persistente o permanente. Establecieron que la presencia de un factor mayor o dos menores duplicaba el riesgo de presentar trombosis del dispositivo. Dukkupati et al⁴⁰ encontraron, sin embargo, otros predictores como el antecedente de ictus isquémico o AIT, arteriopatía periférica y el diámetro de la orejuela. Aunque la presencia de FA permanente se registró en ambos estudios como un predictor independiente, el poder de predicción no se mostró en ninguno de los dos artículos. En cuanto al derrame pericárdico y la enfermedad renal crónica, ambos fueron incluidos también en nuestro análisis sin resultar como predictores de forma significativa. Contrariamente, la escala de riesgo tromboembólico CHA₂DS₂VASc había sido ya definida como predictor de trombosis sobre dispositivo de cierre de orejuela en el registro de Watchman 2.5³⁷.

Para nuestro estudio, hemos utilizado un registro multicéntrico e internacional de casos con cierre percutáneo de orejuela izquierda en centros con un alto volumen y elevada experiencia en este procedimiento. La incidencia de trombosis sobre el dispositivo se registró durante algún momento del seguimiento en el 4.31% de los pacientes. Este porcentaje es similar a registros previamente publicados³⁸, por lo que podemos establecer que se trataba de una muestra

6. Discusión

representativa de la situación actual. Por este motivo, al tratarse de un evento adverso poco frecuente pero muy relevante, a la hora de realizar modelos de ML es importante procesar correctamente los datos. Aplicar técnicas como el remuestreo ayudará a no controlar posibles sesgos en los resultados.

En nuestros modelos, no hemos encontrado diferencias significativas en la capacidad de predicción de la trombosis entre los métodos clásicos y las técnicas de ML. En ambos casos el poder de predicción es muy limitado, en consonancia con resultados anteriores. De esta forma, encontrar la forma de predecir con consistencia estos eventos continúa siendo complicado, y es posible que existan otros factores latentes que pueden ser los potenciales predictores, o que estos pasen desapercibidos.

Además, no solo destaca esta baja potencia de predicción, sino que estos predictores que se hallaron en cada análisis no son consistentes. Si bien en los modelos creados mediante el análisis multivariable con la metodología clásica, se encontraron variables independientes de forma repetida en todos los modelos, estos son diferentes a los previamente publicados. De las escalas de riesgo tromboembólico y de sangrado reportadas, solamente el CHA₂DS₂VASc se ha publicado anteriormente como predictor independiente. En contraposición, en los modelos de ML, múltiples variables aparecen como predictores de trombosis del dispositivo, pero de forma muy inconsistente. Esto va en consonancia con los hallazgos diversos de cada registro publicados hasta la fecha.

Llama la atención también, que tanto el tratamiento al alta, como el tratamiento previo al procedimiento, no surjan en los modelos como factores predictores. Desde el punto de vista fisiopatológico, podríamos creer que esto es una de las claves, y por ello gran parte de los esfuerzos están enfocados en encontrar cuáles es el tratamiento y su duración óptimos. Sin embargo, estos hallazgos son consistentes con aquellos mostrados por Simard et al³⁸ en su estudio.

6. Discusión

Los resultados obtenidos presentan diferencias en función de haber realizado o no las técnicas de remuestreo. En los dos primeros experimentos, en los que no se realizó encontramos diferencias importantes. En primer lugar, la potencia predictiva de los modelos de ML (con áreas bajo la curva ROC para RL de 0.7974; RF de 1.0; GB de 1.0) es claramente superior a los del método convencional (área bajo la curva ROC de 0.5446). En segundo lugar, también se detecta un mayor equilibrio entre la sensibilidad y la especificidad en estos modelos creados con ML, con menos variabilidad. Sin embargo, esto debe analizarse con recelo, ya que tanto la creación del modelo como su evaluación y prueba, se han realizado en los dos casos utilizando el conjunto de datos completo. Esto explica que estos modelos están probablemente sobreajustados⁹⁵. Este es el problema de no disponer de una muestra externa para poder realizar el test y la validación.

Fue necesario realizar un remuestreo y repetir la comparación entre los dos métodos para poder realizar un análisis más fiable. Entonces, no se encuentran diferencias entre el análisis clásico (área bajo la curva 0.5174) y el de ML (área bajo la curva ROC: RL 0.5325, RF 0.4250, GB 0.4893). Incluso, estos resultados, próximos a 0.5, prácticamente no difieren con respecto a un modelo de predicción al azar. Esta ausencia de diferencias tras el uso del remuestreo, corrobora nuestra hipótesis de que en el experimento II existía un sobreajuste claro⁹⁵.

Cabe destacar también que, en estos modelos, la sensibilidad es constantemente baja, lo cual indica la baja probabilidad de detectar trombos. La trombosis del dispositivo de cierre de orejuela es un evento raro, pero con alta repercusión clínica al aumentar el riesgo de eventos tromboembólicos, y que insta a reintroducir el tratamiento anticoagulante en pacientes ya de por sí con alto riesgo hemorrágico. Por otro lado, la trombosis sobre el dispositivo de cierre de orejuela no siempre aparece en las primeras semanas tras la intervención, sino que en algunos casos aparece más allá de los primeros seis meses⁴⁰. Es por este motivo por el que las pruebas de imagen necesarias y su frecuencia no estén perfectamente consensuadas y los protocolos difieran entre los centros. Una alta sensibilidad, a costa de una baja especificidad podría ser

6. Discusión

parcialmente útil. Esto se explica porque si fuera así, el modelo podría seleccionar pacientes a los que deberíamos realizar un seguimiento más exhaustivo y prolongado durante más meses con estudios repetido de imagen como el ETE o el TC cardíaco. Y poder dejar un seguimiento más laxo a aquellos que el modelo no detectó como de alto riesgo. Sin embargo, en este evento poco frecuente, una baja sensibilidad no permite que estos modelos tengan un valor relevante en la práctica clínica habitual.

A pesar de que no se encontraron diferencias en el poder de predicción de los modelos, sí que se encontró alguna diferencia en cuanto a las variables de predicción identificadas en los dos métodos. En el método clásico, se encontraron resultados más robustos y consistentes en este aspecto. Tanto sin como con remuestreo, la escala HASBLED de riesgo hemorrágico mayor al año surgió como un factor independiente de predicción de trombosis del dispositivo. En contrapartida, los métodos de ML utilizaron un espectro de variables mucho mayor, pero solamente tres variables se repitieron de forma constante en todas las series. Estas fueron la enfermedad coronaria previa, el tratamiento previo con fármacos antivitaminas K y, de nuevo, la escala HASBLED.

Por lo tanto, con estos hallazgos nuestro estudio cuestiona en primer lugar las posibilidades de predicción de los métodos de ML frente a aquellos métodos estadísticos clásico⁹⁶. En segundo lugar, se demuestra que sigue existiendo una discrepancia importante en las variables predictoras entre los diferentes estudios^{36-40,42,97}. Además, en otros estudios anteriores, se habían encontrado otros factores como predictores como la edad⁴², el antecedente de ictus^{40,42}, FA permanente, arteriopatía periférica, el diámetro de la orejuela⁴⁰, ser fumador, el sexo³⁹ o el tamaño del ostium de la orejuela³⁶.

Incluso, a pesar de que algún estudio prospectivo haya demostrado menor incidencia de trombosis sobre el dispositivo con la mitad de dosis de anticoagulante de acción directa⁹³, ninguno ha demostrado su papel predictivo. Es llamativo también, que solamente uno de los

estudios había realizado *bootstrapping*⁹⁷, y que en ninguno de ellos se reporta el poder estadístico de la predicción.

6.1.1. Limitaciones

Como principal limitación del estudio cabe señalar que nuestro estudio solamente ha utilizado tres tipos de modelos de ML y dos técnicas de remuestreo. De todas formas, el objetivo principal del estudio era comparar modelos de predicción de ML con el análisis multivariable estadístico clásico, lo cual sí se ha hecho de forma apropiada.

Otra de las limitaciones reseñables es la falta de acuerdo en el tratamiento y seguimiento tras el procedimiento. En cuanto al tratamiento, actualmente existen diferentes protocolos sin un consenso. Estudios en marcha como el ADALA (NCT05632445)⁹⁴ o el ANDES (NCT03568890) podrán aportar más evidencia próximamente y es posible que consigan establecer un protocolo de tratamiento más estandarizado. Además, la indicación de cierre de orejuela es muy heterogénea, con perfiles de pacientes muy distintos, por lo que muchas veces el tratamiento ha de individualizarse en base al riesgo tromboembólico y hemorrágico. El seguimiento con alguna de las técnicas de imagen disponibles también varía en función del centro, por lo que la detección de la trombosis sobre el dispositivo de cierre de orejuela podría variar en función del centro.

El carácter retrospectivo del estudio es también una limitación en sí misma, y como tal, únicamente se pueden generar hipótesis en base a estos resultados.

De todas formas, cabe destacar que se han realizado experimentos muy exhaustivos, y se ha añadido la ratio de variación de los predictores para poder contrastar rigurosamente ambos métodos.

Por último, la base de datos multicéntrica utilizada para este estudio es similar a las utilizadas en publicaciones recientes, con variables sobre todo clínicas. A pesar de presentar una incidencia

6. Discusión

de trombosis de dispositivo similar a registros previos, existe un desequilibrio claro entre pacientes con y sin evento durante el seguimiento. Es posible que nuevos conjuntos con datos más equilibrados, o que incluyan nuevas variables más allá de aquellas clínicas clásicas, como pueden ser pruebas de imagen o datos hematológicos, consigan un poder predictivo mayor. Es posible que, dado que el ML no es capaz de encontrar buenos predictores, y que hay una heterogeneidad importante en los hallazgos entre todos los estudios, existan características fisiopatológicas que desconocemos todavía. Sin embargo, conjuntos de datos mayores, o incluir variables anatómicas, antecedentes de rechazo de otros dispositivos como la reestenosis intrastent, incidencias durante el procedimiento, o alteraciones de la coagulación, puedan aportar más poder de predicción tanto para modelos clásicos como de ML.

En todo caso, parece necesario bien realizar una validación externa o bien una validación cruzada a la hora de crear modelos de predicción y así poder establecer la potencia predictiva de estos. De esta forma podremos evaluar de una manera consistente los valores predictivos que encontremos.

6.1.2. Resumen de los hallazgos

La trombosis sobre el dispositivo de cierre de orejuela es un evento raro pero que se asocia a un riesgo incrementado de eventos tromboembólicos. Encontrar variables predictoras es muy importante de cara a mejorar el tratamiento y el seguimiento tras la intervención. Hemos analizado un registro multicéntrico, y no se han encontrado diferencias en el poder predictivo entre los modelos creados mediante el análisis estadístico clásico y aquellos con ML. En ambos casos, el poder predictivo de los modelos es muy bajo, próximo a los resultados de modelos azarosos. Estos cuestionan los resultados previamente publicados, en los que se establecen diferentes predictores independientes, con resultados muy heterogéneos y sin reportar en ninguno de ellos el poder predictivo de esos modelos. Es posible que los algoritmos de ML consigan mejores resultados con registros más extensos.

6.2. Problemática de los automatismos matemáticos

Con estos hallazgos surge la duda de si estamos aplicando bien los modelos de ML para predecir eventos raros como la trombosis sobre el dispositivo de cierre de orejuela. Revisando de forma minuciosa los algoritmos creados, no parece que haya un problema en el proceso de creación y entrenamiento de los algoritmos. Sin embargo, la potencia estadística de los resultados es limitada. Ante estos hallazgos consideramos varias reflexiones necesarias.

Tanto en los métodos clásicos como en los de ML, se necesitan conjuntos de datos muy grandes para conseguir crear modelos con una potencia de predicción elevada. Si bien, debemos valorar si ante eventos de baja frecuencia como el que hemos estudiado, el conjunto de datos utilizado para aplicar estos modelos ha de ser mucho mayor. Incluso, puede ser posible que, para conseguir un mejor resultado, sea necesario obtener una muestra mucho más balanceada, con muchos más casos con el evento índice.

Otra de las reflexiones que podríamos extraer de este estudio es que es posible que los datos utilizados contengan errores o, que no sean significativos. Esto supondría que el análisis con ML esté perfectamente confeccionado, y que esta tecnología sí puede ser útil para identificar modelos de predicción, incluso para eventos raros. Es posible que todavía no conozcamos de forma correcta cuáles son los factores clave para que algunos pacientes desarrollen la trombosis sobre el dispositivo, y que por lo tanto no dispongamos de estas variables en las bases de datos. Diferentes grupos han intentado dilucidar cuáles son estos factores clave, con resultados muy dispares y poco consistentes. Incluir otro tipo de variables podría ser la solución para conseguir mejores resultados.

Finalmente, también debemos plantearnos si, a pesar de que los modelos estén bien confeccionados, es posible automatizar todo tipo de análisis. En el caso de los estudios sobre fenómenos biológicos, la complejidad de estos puede comprometer tanto los resultados de los modelos matemáticos clásicos, como los del aprendizaje automático.

6. Discusión

En su artículo, Gyllingberg et al⁹⁸, ponen de manifiesto las dificultades a las que se enfrentan estas dos técnicas en los próximos años. Se definen los fenómenos biológicos, a diferencia de aquellos fenómenos físicos, como eventos complejos, que pueden ser abordados desde infinitos puntos de vista. Esto puede deberse a la difícil extrapolación de los datos desde los niveles moleculares hasta los colectivos como especie. Además, es probable que el carácter social inherente a los fenómenos biológicos infiera una dificultad extra a la hora de generar predicciones mediante modelos. Los autores critican en este trabajo que se traten de enfocar estos problemas mediante la unificación, intentando universalizar los modelos matemáticos que expliquen los sistemas biológicos. Contrariamente, exponen que debe realizarse un enfoque más plural, en el que se promueva la creación de nuevos modelos más creativos y diferentes a los ya establecidos. De esta forma, se pueden ir probando estos nuevos modelos y desechando aquellos que no sirvan, en lugar de simplemente analizar y modificar levemente los modelos ya conocidos como abogan los seguidores de los modelos universales. La idea que se propone es la de utilizar múltiples modelos a diferentes niveles, que incluso pueden llegar a ser contradictorios, para explicar estos fenómenos biológicos.

La automatización del análisis en los eventos biológicos puede suponer entonces una estrategia contraproducente. No solo la universalización, sino también se critica que el uso de ML como técnica de predicción puede ayudar a detectar ciertos patrones, pero sin llegar a conseguir explicar estos fenómenos, ajustándose simplemente a los datos utilizados. En este punto sería necesario añadir un punto de vista conceptual del problema a resolver, y debería valorarse entonces la causalidad de las variables en relación con el evento a estudio.

De esta forma, puede ser necesario la creación de nuevos modelos que, tanto a través de técnicas estadísticas tradicionales como de aprendizaje automático que aporten formas novedosas de entender el comportamiento los fenómenos biológicos. En este caso los esfuerzos se deberían centrar en intentar ser más creativos con nuevos modelos, para tener una variedad

6. Discusión

mucho mayor en nuestro arsenal, y probablemente poder utilizar combinaciones de diferentes estrategias que nos aporten una visión más global y fidedigna del problema a analizar.

Esta estrategia se asimila también a la teoría desarrollada por Nassim Nicholas Taleb en su obra. En ella, el autor expone sus teorías del “Cisne negro” o la “Antifragilidad”^{99,100}. En la primera define que eventos poco probables, pueden tener un impacto muy relevante en cualquier ciencia a la que se aplique. En la segunda, este matemático expone que la división de los sistemas provocará un beneficio último común. Esto se puede extrapolar a lo previamente comentado, la creación de múltiples nuevos modelos para su uso en el ML. Probar múltiples modelos nuevos, podrá conseguir que algunos de ellos sobresalgan de manera significativa frente al resto, y que acaben por aportar nuevas soluciones a los problemas que han ido surgiendo a lo largo de los años.

En resumen, obtener nuevas metodologías o mayor número y mejor calidad de los datos, podrían ser las claves para lograr que el aprendizaje automático o ML consiga mejores predicciones sobre eventos futuros complejos, como los son generalmente en la investigación en el ámbito sanitario.

Todos estos aspectos, hacen replantearse la necesidad de mejorar la formación del personal sanitario y, más específicamente, de los médicos en temas estadísticos o sobre ML. Como se comentó previamente en algunos de los estudios, los métodos estadísticos son mejorables para poder aportar mayor evidencia a los resultados que se obtienen. Además, aunque su uso está claramente en aumento, el aprendizaje automático continúa siendo una herramienta poco utilizada en la investigación médica.

Esto puede tener una explicación en la poca formación, tan específica y diferente a la medicina, que presenta el sector sanitario sobre este tema. Dado que en la carrera no existe por el momento un plan formativo que incluya estas nuevas tecnologías, es posible que sea necesario empezar a implementar una enseñanza reglada, no solo a los estudiantes que comiencen sus

estudios a partir de ahora, sino también a los sanitarios que ya realizan una labor asistencial^{61,65}. Las ventajas que puede aportar la IA pueden facilitar la labor médica, ayudando en el diagnóstico y manejo de los pacientes, repercutirá de forma positiva en el paciente, pudiendo mejorar su pronóstico y calidad de vida⁶².

De esta forma, es necesario antes de iniciar estos programas, evaluar cuál es el nivel de los profesionales sanitarios. Además, dada la complejidad de algunos temas relacionados, es posible que la creación de plataformas de ayuda a la implementación y uso del ML puedan facilitar su expansión e implementación en esta área.

6.3. Evaluación del conocimiento de los profesionales de la salud y desarrollo de una plataforma para facilitar el uso de modelos de inteligencia artificial en el ámbito sanitario. Proyecto KOOPA ML

6.3.1. Discusión general

El objetivo principal de este estudio era evaluar las posibles dificultades a las que se pueden enfrentar profesionales del ámbito sanitario a la hora de utilizar herramientas de ML para crear modelos de predicción a través de una plataforma sencilla. Además, también se quería evaluar el impacto de las recomendaciones incluidas en la plataforma para facilitar su uso en participantes con escasa formación en este tema.

En los hallazgos encontrados destaca que el nivel de conocimientos previos sobre IA se relaciona más claramente que los mensajes con recomendaciones durante los ejercicios, con el éxito para llevar a cabo modelos de ML. Esto confirma que intentar crear modelos sin una formación básica es tremendamente complejo, a pesar de que para intentar suplirlo se añadan mensajes con recomendaciones durante el proceso. Aunque la mayoría de los participantes ha conseguido realizar varias tareas, la mayoría de estos no consigue completar los ejercicios propuestos.

6. Discusión

Aun cuando en la mitad de los casos se incluían mensajes de texto con recomendaciones que aportaban toda la información necesaria para realizar las tareas, no se encontraron diferencias en los resultados entre los dos grupos. Incluso, los participantes que sí disponían de ellas, una vez las leían las cerraban y no volvían a prestarle atención. De esta forma no podían saber si en el cuadro de diálogo se mostraba más información acerca de cómo avanzar en los ejercicios. Los participantes reflejaron en sus opiniones finales que, como no se les informó previamente que podían recibir ayuda mediante estos consejos, les restaron importancia.

Por el contrario, los participantes hacían más caso a los mensajes que aparecían tras ejecutar los algoritmos informando de los errores en los que se había incurrido. Probablemente el hecho de no haber sido capaces de completar el modelo los alertaba más y se seguían más estos consejos, consiguiendo en varias ocasiones completar más tareas previamente no completadas de forma satisfactoria.

Se realizó también un análisis de los mapas de calor durante los ejercicios (Figura 23A). En este caso se objetiva como un participante interacciona muy poco con el cuadro de diálogo que tiene las recomendaciones en la parte inferior. En este caso en particular, se clicaba en esta zona para poder cerrar el cuadro más que para leer estos consejos. Esta falta de interés por las recomendaciones puede ser debido también a la falta de conocimientos sobre el tema y que se ha utilizado un lenguaje demasiado técnico, como confirmaron posteriormente los participantes al terminar los ejercicios.

6. Discusión

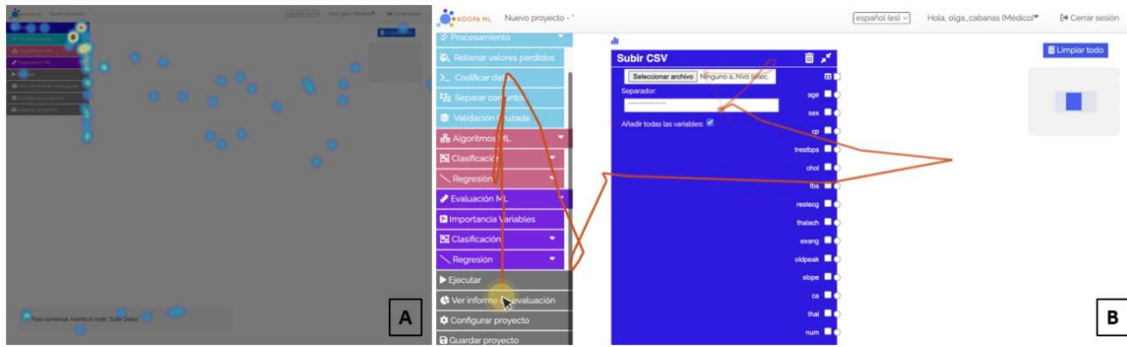


Figura 23. Mapas de utilización de la aplicación KooptaML. A: *mapa de calor de uno de los participantes durante el uso de la aplicación. Se observa escasa interacción con la parte inferior donde se encontraban las recomendaciones adicionales.* B: *mapa donde visualizan todos los desplazamientos del ratón que realizó un usuario para poder encontrar el nodo deseado para una de las tareas solicitadas.*

Con estos resultados de las interacciones de los participantes con la plataforma podemos concluir que estas recomendaciones no son útiles para mejorar la usabilidad de la aplicación.

Cuando se les interroga acerca del uso que habían hecho de las recomendaciones, referían que estas les abrumaban y que ese era el motivo por el que las ignoraron. Esta sensación, patente en todos los casos que recibieron ayuda, permite justificar que no se hayan encontrado diferencias entre ambos grupos, ya que el grupo que sí las tenía les prestó poca atención. Por estos motivos se revisarán y modificarán las recomendaciones para hacerlas más adecuadas al nivel de formación en el aprendizaje automático.

Otra de las lecciones aprendidas del estudio es la dificultad que obtuvieron en casi todos los casos para entender cuáles eran los nodos correctos necesarios para cada tarea. Los participantes perdían buena parte del tiempo buscando en el listado desplegable el nodo indicado.

Para poder mejorar la plataforma se nos plantean dos opciones. La primera de ellas sería esconder o minimizar el desplegable una vez se ha elegido el nodo que se va a utilizar. La segunda sería disminuir el número de opciones ya que, en personas con poca formación sobre este tema, ofrecer más posibilidades podría enlentecer o abrumar al usuario entorpeciendo su labor.

6. Discusión

En la figura 23B se pueden observar todos los movimientos que realiza un participante con el ratón del ordenador en una de las tareas. Solamente para encontrar el nodo requerido necesita buscar a lo largo de todo el desplegable ya que no es capaz de escoger directamente el apartado necesario. Esto puede estar también relacionado con la escasa formación en la creación de algoritmos de ML, lo cual dificulta la búsqueda de los nodos que están clasificados por diferentes apartados.

Otra de las dificultades que se hallaron, es que los nodos eran demasiado grandes para el espacio de trabajo disponible. Esto provocó que se perdieran durante el proceso de creación del modelo, lo que dificultaba la conexión de los diferentes nodos con el ratón del ordenador.

A pesar de que casi todos los participantes no fueron capaces de completar correctamente las tareas requeridas, todos exponen que probablemente muchos de los errores cometidos se debían a la falta de formación sobre ML. De hecho, remarcaban la facilidad para utilizar la plataforma, lo intuitiva que era y cómo en un corto espacio de tiempo, con el paso de las tareas y de los ejercicios se sentían cada vez más cómodos. Todas estas dificultades que han aparecido durante este estudio serán evaluadas y servirán para mejorar la próxima versión de la plataforma.

6.3.2. Discusión desde el punto de vista médico

En este estudio hemos tratado de evaluar cómo plataformas como el KoopaML, de apoyo para la utilización de ML para la creación de modelos de predicción en personas con poca formación sobre IA, ayudan a los profesionales sanitarios.

La IA surge como una nueva herramienta que aporta soluciones y ventajas para la investigación en el ámbito sanitario^{49,50,101}. Aunque ya se han llevado a cabo numerosos estudios y publicaciones en los que se utiliza esta tecnología en diversos sectores, como en la medicina, en la investigación médica todavía existe cierto desconocimiento por parte de los profesionales sanitarios. Los modelos creados mediante ML son capaces de mejorar muchas partes de la

6. Discusión

práctica médica diaria, bien sea mejorando y ayudando en el diagnóstico, como en la estratificación de subgrupos de pacientes por riesgos que podrán beneficiarse de un seguimiento más estrecho, o para realizar tratamientos específicos para cada paciente^{53,56,57,75,102}. En este contexto se crea la plataforma KoopaML, para poder dar soporte y ayudar a los profesionales sanitarios con menor formación en este ámbito, y así poder extender su uso y mejorar los resultados en el cuidado de los pacientes.

La creación de modelos de predicción mediante IA requiere de una mínima formación, no solamente en aspectos relacionados con el ML, sino con métodos estadísticos, cálculo o programación^{103,51}. En el caso de los profesionales de la salud, la formación sobre estos temas es limitada^{61,65}. Esto explica los hallazgos de este estudio, sin observarse diferencias entre los dos grupos que recibieron o no las recomendaciones. Sin embargo, cabe destacar que aquellos que sí las recibieron completaron las subtarefas de forma más rápida y fácil.

A pesar de que el poco conocimiento de estas herramientas ha limitado los resultados de los ejercicios realizados, todos los participantes remarcan lo simple y fácil que es el uso de esta aplicación. Esto se confirma al analizar el tiempo que invierten en la realización del primer ejercicio respecto al segundo. Una pequeña experiencia con la aplicación provoca que los resultados sean mejores, como se puede objetivar en la tabla 11.

Todos los hallazgos apuntan a que la falta de formación en estas nuevas herramientas dentro del sector sanitario es clave y fundamental para poder expandir esta tecnología en la investigación médica. Este es el factor más limitante hoy en día, por lo que se deberá hacer mucho hincapié en creación de programas de formación para médicos y sanitarios¹⁰¹. Estos deben ir dirigidos no solamente a aquellos que ya realizan labores de investigación o práctica clínica, sino incidir también en los futuros profesionales. Para ellos será necesario enseñar no sólo el funcionamiento, la metodología o los conceptos necesarios para crear los modelos, sino también las potenciales aplicaciones que tiene, para así poder ser utilizadas cada vez más en el día a día.

6. Discusión

Es crucial también entender los usos que la IA puede aportar. El apoyo en la toma de decisiones, tanto para el diagnóstico como para el tratamiento o para la estratificación de los pacientes, es una de las aplicaciones más importantes. Incluso, ya se empiezan a desarrollar máquinas capaces de realizar diagnósticos sin necesidad de supervisión médica⁵⁷. También se han desarrollado robots quirúrgicos que podrían empezar a realizar intervenciones, ya que han demostrado ser superiores a los seres humanos en los experimentos^{55,104}. Sin embargo, para que todo esto evolucione más, será necesario abordar nuevos temas y problemas, tanto éticos como legales. A medida que se siga investigando, seguirán produciéndose errores, sobre todo en aquellos avances en calidad o esperanza de vida de los pacientes, por ello será de vital importancia analizar y evaluar los errores, ya que pueden acarrear ciertas responsabilidades.

6.3.3. Perspectivas futuras

La formación que se imparte en la educación obligatoria o durante el bachillerato acerca de los métodos estadísticos es muy básica⁶¹. Además, en la formación universitaria de estas carreras tampoco se incide de forma importante en rellenar el déficit que presentan los alumnos⁶⁵. Si uno lo piensa detenidamente, parece lógico que en carreras de ciencias como son estas, en las que la investigación adquiere un papel muy importante para conseguir los avances que se han ido produciendo a lo largo de la historia, se haga hincapié en una enseñanza más intensiva en cálculo y estadística. Sin embargo, la corriente actual aboga por planes de estudios que dejan de lado estas capacidades^{49,62}. Esto conlleva a que la formación en esta rama científica quede al amparo de las ganas y el interés individual de cada sanitario.

Esto se observa en la calidad de los estudios realizados, en los que los métodos estadísticos utilizados en algunos casos pueden ser cuestionables³⁹⁻⁴¹. El rigor científico es necesario y crucial para continuar con el desarrollo del ser humano.

En cuanto a una nueva tecnología como es la IA, con sus ramas correspondientes como el aprendizaje automático o ML, la situación es claramente peor.^{62,101} Estas nuevas herramientas

6. Discusión

tienen una vida mucho más corta en comparación con la estadística clásica. Puede ser este uno de los motivos por el que el conocimiento que tiene el ciudadano medio sea mucho menor que con la estadística.

Será necesario implementar nuevos planes de formación en IA a todos los niveles de la sociedad, incluyendo las generaciones futuras y aquellos que se encuentran ya en el mercado laboral dentro del ámbito sanitario. Además, será preciso contar con profesores formados, que puedan enseñar al resto esta nueva tecnología. Una vez se estandarice la formación, y la IA se expanda y sea adoptada de forma mayoritaria, será posible sacarle el máximo rendimiento.

6.3.4. Resumen de los hallazgos

En este estudio se evalúa la plataforma KoopaML que pretende dar soporte a usuarios no experimentados para la creación de modelos mediante ML.

Los principales objetivos del estudio son evaluar las facilidades o dificultades a los que se pueden enfrentar los usuarios de esta plataforma, y valorar si las recomendaciones incluidas para facilitar su uso mejoran los resultados. Para ello se utilizaron unos ejercicios prácticos.

Los hallazgos encontrados confirman que estas recomendaciones no mejoran los resultados obtenidos comparado con el grupo que no recibía estas ayudas durante los ejercicios. Esto parece estar relacionado con la falta de formación y la escasa experiencia de los sanitarios en la IA.

Sin embargo, se extraen conclusiones relevantes que pueden ayudar a mejorar el desarrollo de la plataforma. La mayoría de las dificultades que se han encontrado durante el estudio se deben al uso de lenguaje técnico, tanto en los listados desplegados de la aplicación como en las recomendaciones añadidas. Esto sumado al escaso nivel formativo en este campo indica que las nuevas versiones de la aplicación deben pulir estas características para disminuir las posibles dificultades que puedan surgir con su uso.

6. Discusión

Por lo tanto, se llevarán a cabo nuevas actualizaciones de KoopaML para mejorar su usabilidad y su adopción por usuarios inexpertos. Además, se pondrán en marcha nuevos estudios para aumentar el conocimiento sobre la plataforma y poder entender cómo los sanitarios pueden sacar el mayor rendimiento a plataformas como la nuestra. De esta forma se podrá conseguir una mayor adopción del aprendizaje automático en el ámbito sanitario.

Los hallazgos de este estudio confirman lo que previamente sospechábamos y el motivo por el que se inició el proyecto. La formación sobre ML, así como en métodos estadísticos y de cálculo en el sector sanitario es limitado. Por este motivo debemos avanzar hacia la enseñanza de este tipo de nuevas herramientas desde edades tempranas, ya que tienen un potencial muy interesante. Plantear su estudio durante la formación obligatoria o, al menos, durante la formación universitaria puede ser clave para un gran desarrollo tanto de la IA, como de la medicina en los próximos años. Sin embargo, no todos son ventajas, presenta una parte negativa y se trata de una tecnología que todavía está en crecimiento, por lo que se plantean unos retos todavía por resolver.

6.4. Ventajas e inconvenientes del uso del machine learning

Debido a la expansión de la IA, con una presencia cada vez mayor en la vida de la población, especialmente de los países desarrollados, se han empezado a distinguir tanto ventajas como posibles desafíos y amenazas sobre el aprendizaje automático. Conocer todos estos aspectos será necesario a la hora de aplicar estas nuevas herramientas en el ámbito sanitario. Estos deberán ser abordados y definidos cuanto antes para poder llevar a cabo programas de aplicación de la IA en la medicina, y así poder optimizar y evitar problemas con su uso.

6.4.1. Ventajas o potenciales usos

Dentro de las posibles ventajas podemos encontrar:

- Beneficios para la ciudadanía: esta nueva tecnología puede ayudar a mejorar la calidad de vida de la población, gracias a una mejora en la atención sanitaria o en los medios de

6. Discusión

transporte, que podrán ser más eficientes y seguros. En este caso, se podría conseguir un acceso a la sanidad mucho más sencillo y cercano, con sistemas de asistencia a personas con movilidad reducida o dependientes.

- Oportunidades empresariales: la IA puede ayudar a desarrollar nuevos productos y servicios creando oportunidades de negocio no desarrolladas previamente. Además, puede facilitar la logística de las empresas en diferentes sectores como la agricultura, el transporte o la salud. Para ello, la optimización de rutas de transporte, obtener una mayor eficiencia energética, o mejorar la calidad o el precio de los productos, son aplicaciones clave para el crecimiento de los sistemas de salud, que pueden beneficiarse de esta nueva competencia.
- Servicios públicos: la optimización de los recursos públicos conforme la población crece es fundamental en una sociedad desarrollada. Es aquí donde la IA puede ayudar a mejorar y reducir los costes del transporte público, asegurar una correcta eficiencia energética y conseguir por ejemplo una gestión de los residuos óptima.
- Seguridad y protección: el análisis masivo de los datos podrá ayudar no solo a la aplicación del sistema penal, sino también a la prevención y predicción de posibles delitos. Por ejemplo, el reconocimiento facial, puede ayudar en la identificación de posibles sospechosos de crímenes gracias a las imágenes de diferentes cámaras.

6.4.2. Desventajas o retos pendientes de clarificar

- Uso adecuado de la IA: la infrautilización o el uso excesivo sin regulación de los procesos de aprendizaje automático pueden ocasionar un problema en la sociedad. Un uso menor de lo debido, debido al miedo o la reticencia en cuanto a la implantación de nuevas tecnologías puede suponer la pérdida de una ventaja competitiva frente a otros sectores o regiones. Por ello es fundamental promover el uso de la IA desde los gobiernos, mediante subvenciones o las de estrategias para la implementación de una correcta infraestructura que permita a los sanitarios implementar y desarrollar nuevos y mejores

6. Discusión

sistemas de salud. En el lado contrario, un uso excesivo no regulado, puede provocar un uso indebido de los datos de los pacientes, datos que son muchas veces íntimos o de carácter sensible⁶⁴. El uso indebido por parte de grandes empresas privadas u organizaciones públicas (gobiernos o la administración) puede conllevar a un perjuicio para toda la sociedad.

- Amenazas a los derechos fundamentales: existe la posibilidad de que de forma intencionada o no, se realicen análisis con sesgos o errores que pueden inferir conclusiones erróneas¹⁰⁵. Incluso, el uso de algoritmos de decisión con ML puede provocar que se tomen decisiones en los que exista una desigualdad entre los ciudadanos dependiendo de su etnia, religión, sexo u orientación política, entre otros. Esto puede producir un sesgo mayor entre los diferentes grupos poblacionales.
- Impacto en el empleo: el incremento de la adopción de la IA en diferentes sectores podría llegar a reducir la oferta de empleo, incluido el ámbito sanitario, ya que serán necesarias menos personas para llevar a cabo los mismos procesos. El uso de máquinas con métodos de IA podría ayudar a abaratar costes y optimizar los recursos siendo más eficientes, lo cual propiciaría una reducción en el número de empleados. Sin embargo, esto puede no ser del todo cierto, ya que la aparición y expansión de la IA se puede interpretar como una oportunidad para la creación de nuevos puestos de trabajo que no existían previamente, como ya sucedió en épocas anteriores de la historia, como durante la primera revolución industrial en el siglo XVIII.
- Competencia y desigualdad: la implementación y adopción de la IA puede no ser igual en todos los sectores o países, dependiendo del grado de infraestructuras y poder de inversión que presenten. Es ahí cuando la IA puede provocar que desaparezca la competencia, debido a que aquellos países o ciudadanos con mayor capacidad para acumular información y con mayor poder de computación para procesarla, obtendrán una ventaja competitiva demasiado grande que conllevará a una desigualdad no

6. Discusión

desdeñable. Esto puede provocar diferencias importantes en la calidad de la atención sanitaria entre diferentes países o, incluso, entre diferentes grupos de la sociedad dentro de la misma región.

- Riesgos en la seguridad y protección: dada la relación cercana entre el ser humano y las máquinas que pueden usar la IA, surge la necesidad de proteger a la población frente a posibles afrentas que pueden surgir. Problemas en la seguridad, piratería o un mal uso de armas que utilicen algoritmos de aprendizaje automático pueden suponer un peligro. En el caso del ámbito sanitario, se trata de información muy sensible, por lo que debe priorizarse su protección frente a potenciales piratas informáticos.
- Transparencia y privacidad: el uso del reconocimiento facial a través de cámaras o el seguimiento de la población durante su vida o sus navegaciones en internet pueden llegar a crear perfiles de las personas y podría llegar a utilizarse en su contra sin consentimiento de estas. El trato de estos datos ha de llevarse a cabo mediante la mayor legalidad posible para evitar fraudes o un uso inadecuado de esta información. Dentro del ámbito sanitario, en el caso de los seguros o de las clínicas privadas adquiere mayor importancia, ya que podrían ejercer una práctica abusiva si estos datos se explotan de forma inadecuada.

6.5. Retos éticos y aspectos legales de la IA

La adopción de robots o máquinas cada vez más autónomas, así como la automatización de diferentes procesos conlleva a plantearse diferentes situaciones que podrán darse cada vez con mayor frecuencia. Conforme esta tecnología se pone en práctica más a menudo y en más tipos de situaciones, comienzan a surgir problemas o debates éticos y legales respecto a su uso.

Desde el punto de vista ético, el manejo de los datos obtenidos y procesados mediante estos sistemas de IA debe ser muy cuidadoso. En primer lugar, existe una preocupación importante por parte de la mayoría de la población de Reino Unido respecto al manejo de sus datos, incluso

6. Discusión

aunque sea para estudiar mejoras de la asistencia sanitaria¹⁰⁶. En segundo lugar, han de respetarse las leyes vigentes que defienden los intereses de la población en relación con la protección de datos. En el caso de España, está vigente la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales¹⁰⁷.

Por otro lado, debe quedar claro que, a pesar de la evolución obtenida durante las últimas décadas, no existe ningún sistema operativo que esté 100% libre de errores. Por este motivo, es necesario esclarecer quiénes serán los responsables o los culpables, en caso de que se comentan errores con el uso de estos aparatos y, más importante si cabe, en el caso de aquellos que se utilicen en el ámbito sanitario, en el que la salud de las personas puede verse vulnerada^{104,108}.

Para establecer los límites en los que el ser humano debe ser el responsable o culpable de los errores que se cometan, podrían establecerse en la medicina unos niveles similares a los establecidos en la automoción. En este caso se establecen niveles en función del grado de autonomía o dependencia de los vehículos, siendo los seres humanos meros supervisores en los grados más bajos, de forma que las personas continúan teniendo una responsabilidad elevada. De hecho, en Europa, hasta el momento, no es legal la conducción autónoma más allá de estos niveles. En cuanto a la medicina, ya sea en modelos de predicción, algoritmos de diagnóstico por imagen o los robots quirúrgicos, estamos lejos de utilizar estos sistemas de forma completamente autónoma. Sin embargo, cada vez estamos más cerca, por ello es necesario establecer claramente como se gestionarán esos potenciales errores.

Algunos autores como O'Sullivan et al¹⁰⁴, sugieren la posibilidad de desarrollar tanto la tecnología, como los aspectos legales y éticos en el ámbito militar. Sin embargo, existen clara controversias a este respecto ya que en la guerra las situaciones son límites y en ocasiones se intenta terminar con la vida de otras personas. Por el contrario, en el ámbito sanitario se trata de mejorar la calidad de vida y aumentar la esperanza de vida de los pacientes, por lo que los problemas éticos o legales pueden diferir de forma relativamente grande.

En otro orden de cosas, la seguridad de los datos y la ciberseguridad son elementos fundamentales en este proceso. Cuidar los datos y el acceso remoto a nuestras máquinas será de vital importancia para que no caigan en manos de personas con intereses controvertidos. Robots autónomos o algoritmos automáticos pueden llegar a ser hackeados y reprogramados para ir en contra de los intereses de los pacientes por algún fin último no deseado. Por este motivo se ha de ser muy cuidadoso a la hora de añadir los cortafuegos necesarios a nuestros ordenadores o bases de datos para evitar así el acceso por parte de rivales u otros delincuentes. Para ello, una de las soluciones es implantar un sistema en el que no se trate con programas completamente autónomos y que estos precisen ser supervisados por un humano siempre, aunque sea de forma somera.

6.5.1. Responsabilidad

Desde el nacimiento del aprendizaje automático, y tras la gran expansión que ha sufrido en los últimos años, surgen dudas al respecto de quién posee la responsabilidad de los potenciales daños ocasionados por el uso de la IA. Una manera de definir la responsabilidad que se ejerce con un acto es la capacidad de dar una explicación racional a las acciones llevadas a cabo¹⁰⁴. En el caso de errores provocados durante el uso de estas máquinas o algoritmos de IA, surgirá siempre la discusión de quién es el responsable de estos errores. Hoy en día las máquinas o robots no son considerados como posibles responsables de errores cometidos. Es en este contexto, en el que se deberá de valorar quién es el encargado del correcto funcionamiento de estas nuevas herramientas. Los potenciales responsables podrían ser el creador del sistema, el operador que lleva a cabo su uso, o el que se dedica a realizar un mantenimiento óptimo. Esclarecer y dejar pautas de quiénes deben ser los responsables será fundamental a corto plazo con la continua expansión de esta tecnología.

6.5.2. Obligaciones

Una de las capacidades de los modelos de ML es la posibilidad de sugerir decisiones frente a inputs, obteniendo y compartiendo unos outputs que nacen de la experiencia del propio algoritmo. Debido a esta forma de funcionamiento, los sistemas pueden sacar conclusiones erróneas, secundarias a sesgos o una conducta o mantenimiento inadecuados. Es así como surge la idea de que, de forma similar a los seres humanos, se exija a las máquinas o modelos tener un registro de todas las acciones realizadas durante la ejecución de un algoritmo. En este registro deberían quedar anotadas todas las decisiones realizadas consecuentemente a los *inputs* recibidos, recopilando así la información que equivaldría a una caja negra como la de los aviones. Esto no serviría únicamente para valorar si alguien no ha hecho bien las cosas, sino también para reevaluar continuamente el correcto funcionamiento de los modelos hasta entonces desarrollados.

6.5.3. Aspectos legales y culpabilidad

Actualmente nos encontramos en un escenario en el que no existe una completa autonomía sin la participación, aunque mínima, de las personas. Por lo tanto, las actuales normas y leyes otorgan responsabilidad todavía al ser humano. Sin embargo, será crucial comenzar a establecer cuáles serán las leyes que rijan el futuro de estas máquinas, ya que en un corto plazo de tiempo será necesario tenerlas desarrolladas y aprobadas¹⁰⁸. La culpabilidad conlleva de forma inherente un castigo, que el caso de las máquinas no es posible aplicar. Así, no se podrá culpabilizar a un algoritmo o modelo, y el culpable último será siempre un ser humano. Tanto Estados Unidos como Europa se comienza a establecer el control sobre el tema. En el primero, se ha optado por iniciar el diseño ético y de educación con el que gestionar este tipo de herramientas, posiblemente desde edades tempranas. Por el contrario, en Europa se intenta hacer prevalecer los derechos humanos y proponer organismos y planes regulatorios para su correcto funcionamiento. Además, será necesario saber en qué situación el operador es el

6. Discusión

responsable, o bien lo es el fabricante o el encargado del mantenimiento. Para ellos será necesario recoger la información registrada durante su funcionamiento de forma verosímil.

6.5.4. Retos éticos

Respecto a los retos a los que se enfrenta la IA, unos de los principales son los éticos, que surgen del uso de procesos automáticos en humanos¹⁰⁸. La ética es una parte de la filosofía inherente al ser humano, de la que no nos podemos desprender, incluso aunque actuemos mediante máquinas. Siempre ha sido necesario, para cualquier tipo de prueba o procedimiento, informar a los pacientes de los posibles riesgos y beneficios de estos y, bajo su propia autonomía obtener el consentimiento informado para llevarse a cabo. En cuanto al uso de modelos de IA, este aspecto cobra mayor importancia, ya que será complicado posteriormente culpabilizar o responsabilizar a las máquinas de los potenciales errores que se pueden producir. La aceptación por parte del paciente de los riesgos de cada intervención deberá ser necesario, y su ausencia excluyente. Esto puede ser entendido como novedoso o, por el contrario, algunos autores como E. Datteri¹⁰⁹ asumen que son las mismas cuestiones éticas que se han observado en otras situaciones a lo largo de la historia reciente. Por estos motivos, tanto en Estados Unidos como en Europa, se debe obtener el consentimiento informado recogido por el médico que lleva a cabo la intervención o la supervisión¹⁰⁴.

En contrapartida, surge otro dilema ético, ya que esta tecnología es todavía muy costosa, por lo que en la actualidad solamente está disponible países desarrollados. Existen diferentes aplicaciones, máquinas o robots que usan IA que han demostrado simplificar los procesos y mejorar los resultados en el ámbito sanitario^{110,111}. De ser así, poder aplicar estas nuevas herramientas solamente en los países más ricos podría suponer un problema ético. Incluso, dentro de estos países, no toda la población tendrá acceso a las mismas, por lo que se produce todavía más desigualdad. Los gobiernos de los diferentes países tendrán que afrontar esta disyuntiva que podrá hacerse más evidente a medida que pasen los años, con una brecha que

6. Discusión

puede ser cada vez mayor debido al crecimiento de la desigualdad que existe en nuestra sociedad.

Una de las más aplicaciones conocidas y más de moda sobre IA es el ChatGPT⁸⁴. Esta herramienta aporta, de forma fácil y muy intuitiva la posibilidad de que cualquier persona, incluso no cualificada tenga acceso a una plataforma sencilla que se comunica a través del lenguaje normal. Esta es capaz de facilitar en análisis de datos y de estudios de investigación ofreciendo un resumen rápido claro y conciso que hubiera llevado mucho más tiempo de haberlo hecho de forma tradicional. En los artículos de revisión podría ser de gran ayuda facilitando el trabajo a las personas que lo utilicen. Sin embargo, también posee desventajas o potenciales riesgos. En el hipotético caso de estar sustentado en diferentes sesgos, aportará una información equivocada sin que nos demos cuenta. Además, provocará una falta o descenso de la capacidad crítica de las personas, lo cual podrá implicar fuertes consecuencias en otros sectores o situaciones de la vida. Otro aspecto a tener en cuenta es el plagio que se produce al utilizar estos métodos para la realización de trabajos académicos, siendo en ocasiones muy complicado diferenciar si se trata de un trabajo copiado o auténtico.

Por último, no solamente en la medicina, pero en el caso de que estas máquinas lleguen a ser completamente autónomas, continuará siendo imposible que sean capaces de realizar razonamientos éticos. Sin embargo, en la problemática ética existe mucha discrepancia dentro de los diferentes sectores de la sociedad, por lo que parece inverosímil pretender que los algoritmos de aprendizaje automático tengan juicio ético, cuando ni siquiera las personas nos ponemos de acuerdo.

6.5.5. Plagio o derechos de autor

Con el desarrollo de aplicaciones como ChatGPT, emerge el riesgo potencial de su uso a la hora de generar artículos científicos. Incluso se podría debatir si con el material utilizado durante su entrenamiento se produjeron problemas de copyright. Se ha demostrado que esta aplicación es

6. Discusión

capaz de crear textos que esquiven los sistemas actuales para detectar fraudes por plagio, consiguiendo una estimación del incluso el 100% de originalidad⁸⁴. Esto supondrá un reto también en los próximos años ya que dividirá a la comunidad científica sobre su uso.

7. Conclusiones

7. Conclusiones

- En un registro europeo sobre trombosis de dispositivo de cierre de orejuela, los métodos de ML no identificaron variables clínicas, analíticas o ecocardiográficas predictoras de nuevos eventos de trombosis del dispositivo con una potencia predictora al menos moderada. Los modelos de predicción realizados presentan áreas bajo la curva de ROC inferiores al 0,6.
- En nuestro experimento, los modelos de aprendizaje automático no mejoran aquellos modelos realizados a través de métodos estadísticos clásicos para la predicción de la trombosis sobre el dispositivo de cierre de orejuela.
- La formación en temas sobre inteligencia artificial y *machine learning* en los profesionales sanitarios es escasa, a pesar del incremento de su uso en este ámbito en los últimos años. Este escaso conocimiento del tema implica una mayor dificultad para crear modelos de predicción incluso a través de plataformas sencillas, a pesar de que se realicen recomendaciones técnicas y/o teóricas durante el proceso.
- Aplicaciones como KoopaML que facilitan el uso del *machine learning* por parte de gente escasamente experimentada pueden ayudar a los sanitarios a mejorar sus conocimientos y a expandir su uso entre ellos. Incluso sin una formación en IA avanzada y con escasa práctica con herramientas de ML, los participantes mejoraron su rendimiento durante los ejercicios para crear modelos de predicción conforme aumentaba el tiempo de uso de la plataforma.

8. Bibliografía

8. Bibliografia

1. January CT, Wann LS, Alpert JS, et al. 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 Practice Guidelines and the Heart Rhythm Society. *Circulation*. 2014;130(23). doi:10.1161/CIR.000000000000041
2. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *European Heart Journal*. 2021;42(5):373-498. doi:10.1093/eurheartj/ehaa612
3. Benjamin EJ, Muntner P, Alonso A, et al. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. *Circulation*. 2019;139(10). doi:10.1161/CIR.0000000000000659
4. Rivera-Caravaca JM, Roldan V, Esteve-Pastor MA, et al. Long-Term Stroke Risk Prediction in Patients With Atrial Fibrillation: Comparison of the ABC-Stroke and CHA₂DS₂-VAS_c Scores. *Journal of the American heart association*. 2017;6(7):e006490.
5. Andrew NE, Thrift AG, Cadilhac DA. The Prevalence, Impact and Economic Implications of Atrial Fibrillation in Stroke: What Progress Has Been Made? *Neuroepidemiology*. 2013;40(4):227-239. doi:10.1159/000343667
6. Kotecha D, Lam CSP, Van Veldhuisen DJ, Van Gelder IC, Voors AA, Rienstra M. Heart Failure With Preserved Ejection Fraction and Atrial Fibrillation. *Journal of the American College of Cardiology*. 2016;68(20):2217-2228. doi:10.1016/j.jacc.2016.08.048
7. Kalantarian S, Stern TA, Mansour M, Ruskin JN. Cognitive Impairment Associated With Atrial Fibrillation: A Meta-analysis. *Ann Intern Med*. 2013;158(5_Part_1):338. doi:10.7326/0003-4819-158-5-201303050-00007
8. Blum S, Muff C, Aeschbacher S, et al. Prospective Assessment of Sex-Related Differences in Symptom Status and Health Perception Among Patients With Atrial Fibrillation. *JAHA*. 2017;6(7):e005401. doi:10.1161/JAHA.116.005401
9. Gaita F, Ferraris F, Anselmino M, Calò L. Atrial fibrillation fundamentals: from physiopathology to transcatheter ablation. *European Heart Journal Supplements*. 2023;25(Supplement_C):C7-C11. doi:10.1093/eurheartjsupp/suad003
10. Wolf PA, Dawber TR, Thomas HE, Kannel WB. Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: The fiamingham Study. *Neurology*. 1978;28(10):973-973. doi:10.1212/WNL.28.10.973
11. Lip GY, Lim HS. Atrial fibrillation and stroke prevention. *The Lancet Neurology*. 2007;6(11):981-993. doi:10.1016/S1474-4422(07)70264-8
12. Glikson M, Wolff R, Hindricks G, et al. EHRA/EAPCI expert consensus statement on catheter-based left atrial appendage occlusion – an update. *EP Europace*. 2020;22(2):184-184. doi:10.1093/europace/euz258
13. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of Clinical Classification Schemes for Predicting Stroke: Results From the National Registry of Atrial Fibrillation. *JAMA*. 2001;285(22):2864. doi:10.1001/jama.285.22.2864

8. Bibliografía

14. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach. *Chest*. 2010;137(2):263-272. doi:10.1378/chest.09-1584
15. Zhu W, He W, Guo L, Wang X, Hong K. The HAS-BLED Score for Predicting Major Bleeding Risk in Anticoagulated Patients With Atrial Fibrillation: A Systematic Review and Meta-analysis: HAS-BLED for predicting bleeding risk in AF. *Clin Cardiol*. 2015;38(9):555-561. doi:10.1002/clc.22435
16. January CT, Wann LS, Calkins H, 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. *Circulation*. 2019;140(2). doi:10.1161/CIR.0000000000000665
17. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: Antithrombotic Therapy to Prevent Stroke in Patients Who Have Nonvalvular Atrial Fibrillation. *Ann Intern Med*. 2007;146(12):857. doi:10.7326/0003-4819-146-12-200706190-00007
18. Granger CB, Alexander JH, McMurray JJV, et al. Apixaban versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med*. 2011;365(11):981-992. doi:10.1056/NEJMoa1107039
19. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation. *N Engl J Med*. 2011;365(10):883-891. doi:10.1056/NEJMoa1009638
20. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med*. 2009;361(12):1139-1151. doi:10.1056/NEJMoa0905561
21. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med*. 2013;369(22):2093-2104. doi:10.1056/NEJMoa1310907
22. Marzec LN, Wang J, Shah ND, et al. Influence of Direct Oral Anticoagulants on Rates of Oral Anticoagulation for Atrial Fibrillation. *Journal of the American College of Cardiology*. 2017;69(20):2475-2484. doi:10.1016/j.jacc.2017.03.540
23. O'Brien EC, Holmes DN, Ansell JE, et al. Physician practices regarding contraindications to oral anticoagulation in atrial fibrillation: Findings from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF) registry. *American Heart Journal*. 2014;167(4):601-609.e1. doi:10.1016/j.ahj.2013.12.014
24. Steinberg BA, Greiner MA, Hammill BG, et al. Contraindications to Anticoagulation Therapy and Eligibility for Novel Anticoagulants in Older Patients with Atrial Fibrillation. *Cardiovasc Ther*. 2015;33(4):177-183. doi:10.1111/1755-5922.12129
25. Takada T, Yasaka M, Nagatsuka K, Minematsu K, Yamaguchi T. Blood Flow in the Left Atrial Appendage and Embolic Stroke in Nonvalvular Atrial Fibrillation. *Eur Neurol*. 2001;46(3):148-152. doi:10.1159/000050788
26. Holmes DR, Reddy VY, Turi ZG, et al. Percutaneous closure of the left atrial appendage versus warfarin therapy for prevention of stroke in patients with atrial fibrillation: a randomised non-inferiority trial. *Lancet*. 2009;374(9689):534-542. doi:10.1016/S0140-6736(09)61343-X

8. Bibliografía

27. Reddy VY, Doshi SK, Sievert H, et al. Percutaneous left atrial appendage closure for stroke prophylaxis in patients with atrial fibrillation: 2.3-Year Follow-up of the PROTECT AF (Watchman Left Atrial Appendage System for Embolic Protection in Patients with Atrial Fibrillation) Trial. *Circulation*. 2013;127(6):720-729. doi:10.1161/CIRCULATIONAHA.112.114389
28. Holmes DR, Kar S, Price MJ, et al. Prospective randomized evaluation of the Watchman Left Atrial Appendage Closure device in patients with atrial fibrillation versus long-term warfarin therapy: the PREVAIL trial. *J Am Coll Cardiol*. 2014;64(1):1-12. doi:10.1016/j.jacc.2014.04.029
29. Reddy VY, Doshi SK, Kar S, et al. 5-Year Outcomes After Left Atrial Appendage Closure. *Journal of the American College of Cardiology*. 2017;70(24):2964-2975. doi:10.1016/j.jacc.2017.10.021
30. Osmancik P, Herman D, Neuzil P, et al. 4-Year Outcomes After Left Atrial Appendage Closure Versus Nonwarfarin Oral Anticoagulation for Atrial Fibrillation. *Journal of the American College of Cardiology*. 2022;79(1):1-14. doi:10.1016/j.jacc.2021.10.023
31. Turagam MK, Osmancik P, Neuzil P, Dukkipati SR, Reddy VY. Left Atrial Appendage Closure Versus Oral Anticoagulants in Atrial Fibrillation. *Journal of the American College of Cardiology*. 2020;76(23):2795-2797. doi:10.1016/j.jacc.2020.08.089
32. Chan KE, Giugliano RP, Patel MR, et al. Nonvitamin K Anticoagulant Agents in Patients With Advanced Chronic Kidney Disease or on Dialysis With AF. *Journal of the American College of Cardiology*. 2016;67(24):2888-2899. doi:10.1016/j.jacc.2016.02.082
33. Boersma LV, Ince H, Kische S, et al. Efficacy and safety of left atrial appendage closure with WATCHMAN in patients with or without contraindication to oral anticoagulation: 1-Year follow-up outcome data of the EWOLUTION trial. *Heart Rhythm*. 2017;14(9):1302-1308. doi:10.1016/j.hrthm.2017.05.038
34. Boersma LV, Ince H, Kische S, et al. Evaluating Real-World Clinical Outcomes in Atrial Fibrillation Patients Receiving the WATCHMAN Left Atrial Appendage Closure Technology: Final 2-Year Outcome Data of the EWOLUTION Trial Focusing on History of Stroke and Hemorrhage. *Circ: Arrhythmia and Electrophysiology*. 2019;12(4):e006841. doi:10.1161/CIRCEP.118.006841
35. Cruz-González I, Trejo-Velasco B. Percutaneous left atrial appendage occlusion in the current practice. *Kardiologia Polska (Polish Heart Journal)*. 2021;79(3):255-268.
36. Aminian A, Schmidt B, Mazzone P, et al. Incidence, Characterization, and Clinical Impact of Device-Related Thrombus Following Left Atrial Appendage Occlusion in the Prospective Global AMPLATZER Amulet Observational Study. *JACC Cardiovasc Interv*. 2019;12(11):1003-1014. doi:10.1016/j.jcin.2019.02.003
37. Kaneko H, Neuss M, Weissenborn J, Butter C. Predictors of thrombus formation after percutaneous left atrial appendage closure using the WATCHMAN device. *Heart Vessels*. 2017;32(9):1137-1143. doi:10.1007/s00380-017-0971-x
38. Simard T, Jung RG, Lehenbauer K, et al. Predictors of Device-Related Thrombus Following Percutaneous Left Atrial Appendage Occlusion. *J Am Coll Cardiol*. 2021;78(4):297-313. doi:10.1016/j.jacc.2021.04.098

8. Bibliografía

39. Saw J, Tzikas A, Shakir S, et al. Incidence and Clinical Impact of Device-Associated Thrombus and Peri-Device Leak Following Left Atrial Appendage Closure With the Amplatzer Cardiac Plug. *JACC Cardiovasc Interv.* 2017;10(4):391-399. doi:10.1016/j.jcin.2016.11.029
40. Dukkipati SR, Kar S, Holmes DR, et al. Device-Related Thrombus After Left Atrial Appendage Closure: Incidence, Predictors, and Outcomes. *Circulation.* 2018;138(9):874-885. doi:10.1161/CIRCULATIONAHA.118.035090
41. Vij V, Piayda K, Nelles D, et al. Clinical and echocardiographic risk factors for device-related thrombus after left atrial appendage closure: an analysis from the multicenter EUROCC-DRT registry. *Clin Res Cardiol.* Published online July 18, 2022. doi:10.1007/s00392-022-02065-4
42. Fauchier L, Cinaud A, Brigadeau F, et al. Device-Related Thrombosis After Percutaneous Left Atrial Appendage Occlusion for Atrial Fibrillation. *J Am Coll Cardiol.* 2018;71(14):1528-1536. doi:10.1016/j.jacc.2018.01.076
43. Sedaghat A, Vij V, Al-Kassou B, et al. Device-Related Thrombus After Left Atrial Appendage Closure. *Circulation: Cardiovascular Interventions.* 2021;14:544-553. doi:10.1161/CIRCINTERVENTIONS.120.010195
44. Fountain RB, Holmes DR, Chandrasekaran K, et al. The PROTECT AF (WATCHMAN Left Atrial Appendage System for Embolic PROTECTION in Patients with Atrial Fibrillation) Trial. *American Heart Journal.* 2006;151(5):956-961. doi:10.1016/j.ahj.2006.02.005
45. Cepas-Guillén P, Flores-Umanzor E, Leduc N, et al. Impact of Device Implant Depth After Left Atrial Appendage Occlusion. *JACC: Cardiovascular Interventions.* 2023;16(17):2139-2149. doi:10.1016/j.jcin.2023.05.045
46. Japanese Cross-ministerial Strategic Innovation Promotion Program “Innovative AI Hospital System”; How Will the 4th Industrial Revolution Affect Our Health and Medical Care System? *JMA J.* 2022;5(1):1-8. doi:10.31662/jmaj.2021-0133
47. Kim H, Shon T. Industrial network-based behavioral anomaly detection in AI-enabled smart manufacturing. *J Supercomput.* 2022;78(11):13554-13563. doi:10.1007/s11227-022-04408-4
48. Kim B, Moon JY, Shin JY, Jeon HR, Oh SY, Kim SY. Effect of smartphone app-based health care intervention for health management of high-risk mothers: a study protocol for a randomized controlled trial. *Trials.* 2022;23(1):486. doi:10.1186/s13063-022-06425-3
49. Chen M, Decary M. Artificial intelligence in healthcare: An essential guide for health leaders. *Healthc Manage Forum.* 2020;33(1):10-18. doi:10.1177/0840470419873123
50. Gupta MD, Kunal S, Girish MP, Gupta A, Yadav R. Artificial intelligence in cardiology: The past, present and future. *Indian Heart Journal.* 2022;74(4):265-269. doi:10.1016/j.ihj.2022.07.004
51. Watson X, D’Souza J, Cooper D, Markham R. Artificial intelligence in cardiology: fundamentals and applications. *Internal Medicine Journal.* 2022;52(6):912-920. doi:10.1111/imj.15562
52. Hernandez Medrano I, Tello Guijarro J, Belda C, et al. Savana: re-using electronic health records with artificial intelligence. Published online 2018.

8. Bibliografía

53. Shetty MK, Kunal S, Girish MP, et al. Machine learning based model for risk prediction after ST-Elevation myocardial infarction: Insights from the North India ST elevation myocardial infarction (NORIN-STEMI) registry. *International Journal of Cardiology*. 2022;362:6-13. doi:10.1016/j.ijcard.2022.05.023
54. Sermesant M, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardiovascular imaging. *Nat Rev Cardiol*. 2021;18(8):600-609. doi:10.1038/s41569-021-00527-2
55. Crawford M. *Top 6 Robotic Applications in Medicine, the American Society of Mechanical Engineers*. 2016 (9).; 2019.
56. Zweck E, Spieker M, Horn P, et al. Machine Learning Identifies Clinical Parameters to Predict Mortality in Patients Undergoing Transcatheter Mitral Valve Repair. *JACC: Cardiovascular Interventions*. 2021;14(18):2027-2036. doi:10.1016/j.jcin.2021.06.039
57. Chao CJ, Kato N, Scott CG, et al. Unsupervised Machine Learning for Assessment of Left Ventricular Diastolic Function and Risk Stratification. *Journal of the American Society of Echocardiography*. 2022;35(12):1214-1225.e8. doi:10.1016/j.echo.2022.06.013
58. Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. *J Med Internet Res*. 2020;22(7):e18477. doi:10.2196/18477
59. Bartlett G, Gagnon J. Physicians and knowledge translation of statistics: Mind the gap. *CMAJ*. 2016;188(1):11-12. doi:10.1503/cmaj.151254
60. Romano R, Gambale E. Statistics and medicine: the indispensable know-how of the researcher. *Transl Med UniSa*. 2013;5:28-31.
61. Kolachalama VB. Machine learning and pre-medical education. *Artificial Intelligence in Medicine*. 2022;129:102313. doi:10.1016/j.artmed.2022.102313
62. Cooper A, Rodman A. AI and Medical Education — A 21st-Century Pandora's Box. *N Engl J Med*. 2023;389(5):385-387. doi:10.1056/NEJMp2304993
63. Organization WH. *Big Data and Artificial Intelligence for Achieving Universal Health Coverage: An International Consultation on Ethics: Meeting Report, 12–13 October 2017, Miami, Florida, USA*. World Health Organization; 2018.
64. Dorado-Díaz PI, Sampedro-Gómez J, Vicente-Palacios V, Sánchez PL. Applications of Artificial Intelligence in Cardiology. The Future is Already Here. *Revista Española de Cardiología (English Edition)*. 2019;72(12):1065-1075. doi:10.1016/j.rec.2019.05.014
65. Kolachalama VB, Garg PS. Machine learning and medical education. *npj Digital Med*. 2018;1(1):54. doi:10.1038/s41746-018-0061-1
66. Yang YC, Islam SU, Noor A, Khan S, Afsar W, Nazir S. Influential Usage of Big Data and Artificial Intelligence in Healthcare. Asghar MZ, ed. *Computational and Mathematical Methods in Medicine*. 2021;2021:1-13. doi:10.1155/2021/5812499

8. Bibliografía

67. Essinger SD, Rosen GL. An introduction to machine learning for students in secondary education. In: *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*. IEEE; 2011:243-248. doi:10.1109/DSP-SPE.2011.5739219
68. Martins RM, Von Wangenheim CG, Rauber MF, Hauck JC. Machine Learning for All!—Introducing Machine Learning in Middle and High School. *Int J Artif Intell Educ*. Published online January 25, 2023. doi:10.1007/s40593-022-00325-y
69. Marques LS, Gresse Von Wangenheim C, Hauck JCR. Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. *Informatics in Education*. Published online June 15, 2020:283-321. doi:10.15388/infedu.2020.14
70. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor News*. 2009;11(1):10-18. doi:10.1145/1656274.1656278
71. Anil R, Capan G, Drost-Fromm I, et al. Apache mahout: Machine learning on distributed dataflow systems. *The Journal of Machine Learning Research*. 2020;21(1):4999-5004.
72. Poolsawad N, Kambhampati C, Cleland JGF. Balancing class for performance of classification with a clinical dataset. In: *Proceedings of the World Congress on Engineering*. Vol 1. ; 2014:1-6.
73. Kraiem MS, Sánchez-Hernández F, Moreno-García MN. Selecting the Suitable Resampling Strategy for Imbalanced Data Classification Regarding Dataset Properties. An Approach Based on Association Models. *Applied Sciences*. 2021;11(18):8546. doi:10.3390/app11188546
74. Kamalov F, Thabtah F, Leung HH. Feature Selection in Imbalanced Data. *Ann Data Sci*. Published online January 24, 2022. doi:10.1007/s40745-021-00366-5
75. Sampedro-Gómez J, Dorado-Díaz PI, Vicente-Palacios V, et al. Machine Learning to Predict Stent Restenosis Based on Daily Demographic, Clinical, and Angiographic Characteristics. *Canadian Journal of Cardiology*. 2020;36(10):1624-1632. doi:10.1016/j.cjca.2020.01.027
76. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825-2830.
77. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*. 2002;35(5-6):352-359. doi:10.1016/S1532-0464(03)00034-0
78. Mkhadri A, Celeux G, Nasroallah A. Regularization in discriminant analysis: an overview. *Computational Statistics & Data Analysis*. 1997;23(3):403-423. doi:10.1016/S0167-9473(96)00043-6
79. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324
80. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*. 2013;7:21.

8. Bibliografía

81. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. In: ; 2010. doi:10.25080/MAJORA-92BF1922-011
82. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2
83. Perez-Sanchez P, Vicente-Palacios V. IA-Cardiologia-husa/LAAC_Thrombus_detection_MLvsClassical: v1.0.0-alpha. Published online April 13, 2022. doi:10.5281/zenodo.6457213
84. Marchandot B, Matsushita K, Carmona A, Trimaille A, Morel O. ChatGPT: the next frontier in academic writing for cardiologists or a pandora's box of ethical dilemmas. *European Heart Journal Open*. 2023;3(2):oead007.
85. García-Holgado A, Vázquez-Ingelmo A, Alonso-Sánchez J, et al. User-centered design approach for a machine learning platform for medical purpose. In: *Iberoamerican Workshop on Human-Computer Interaction*. Springer; 2021:237-249.
86. Abadi M, Barham P, Chen J, et al. {TensorFlow}: a system for {Large-Scale} machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. ; 2016:265-283.
87. Vazquez-Ingelmo A, Garcia-Penalvo FJ, Theron R. Information Dashboards and Tailoring Capabilities - A Systematic Literature Review. *IEEE Access*. 2019;7:109673-109688. doi:10.1109/ACCESS.2019.2933472
88. Vázquez-Ingelmo A, Alonso-Sánchez J, García-Holgado A, et al. Bringing machine learning closer to non-experts: proposal of a user-friendly machine learning tool in the healthcare domain. In: *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*. ACM; 2021:324-329. doi:10.1145/3486011.3486469
89. García-Peñalvo F, Vázquez-Ingelmo A, García-Holgado A, et al. KoopaML: A Graphical Platform for Building Machine Learning Pipelines Adapted to Health Professionals. *IJIMAI*. 2023;In Press(In Press):1. doi:10.9781/ijimai.2023.01.006
90. Jääskeläinen R. Think-aloud protocol. *Handbook of translation studies*. 2010;1:371-374.
91. Van Someren M, Barnard YF, Sandberg J. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress*. 1994;11:29-41.
92. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. *International journal of human-computer studies*. 1994;41(3):385-397.
93. Della Rocca DG, Magnocavallo M, Di Biase L, et al. Half-dose direct oral anticoagulation versus standard antithrombotic therapy after left atrial appendage occlusion. *Cardiovascular Interventions*. 2021;14(21):2353-2364.
94. Flores-Umanzor EJ, Cepas-Guillen PL, Arzamendi D, Cruz-González I, Regueiro A, Freixa X. Rationale and design of a randomized clinical trial to compare two antithrombotic strategies after left atrial appendage occlusion: double antiplatelet therapy vs. apixaban (ADALA study). *J Interv Card Electrophysiol*. 2020;59(2):471-477. doi:10.1007/s10840-020-00884-x

8. Bibliografía

95. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc Natl Acad Sci U S A*. 2019;116(32):15849-15854. doi:10.1073/pnas.1903070116
96. Akbilgic O, Davis RL. The Promise of Machine Learning: When Will it be Delivered? *J Card Fail*. 2019;25(6):484-485. doi:10.1016/j.cardfail.2019.04.006
97. Budts W, Laenens D, Van Calenbergh F, et al. Left atrial appendage occlusion with the Amplatzer Cardiac Plug could improve survival and prevent thrombo-embolic and major bleeding events in atrial fibrillation patients with increased bleeding risk. *Acta Cardiol*. 2016;71(2):135-143. doi:10.2143/AC.71.2.3141842
98. Gyllingberg L, Birhane A, Sumpter DJT. The lost art of mathematical modelling. *Mathematical Biosciences*. Published online May 2023:109033. doi:10.1016/j.mbs.2023.109033
99. Taleb NN. *The Black Swan: The Impact of the Highly Improbable*. Vol 2. Random house; 2007.
100. Taleb NN. *Antifragile: Things That Gain from Disorder*. Vol 3. Random House Trade Paperbacks; 2014.
101. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018;284(6):603-619. doi:10.1111/joim.12822
102. Gomes B, Pilz M, Reich C, et al. Machine learning-based risk prediction of intrahospital clinical outcomes in patients undergoing TAVI. *Clin Res Cardiol*. 2021;110(3):343-356. doi:10.1007/s00392-020-01691-0
103. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259
104. O'Sullivan S, Nevejans N, Allen C, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *Int J Med Robotics Comput Assist Surg*. 2019;15(1):e1968. doi:10.1002/rcs.1968
105. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. *PLoS Med*. 2018;15(11):e1002689. doi:10.1371/journal.pmed.1002689
106. Fenech M, Strukelj N, Buston O. Ethical, social, and political challenges of artificial intelligence in health. *London: Wellcome Trust Future Advocacy*. 2018;12.
107. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. <https://www.boe.es/eli/es/lo/2018/12/05/3/con>
108. McGreevey JD, Hanson CW, Koppel R. Clinical, legal, and ethical aspects of artificial intelligence–assisted conversational agents in health care. *Jama*. 2020;324(6):552-553.
109. Datteri E. Predicting the Long-Term Effects of Human-Robot Interaction: A Reflection on Responsibility in Medical Robotics. *Sci Eng Ethics*. 2013;19(1):139-160. doi:10.1007/s11948-011-9301-3

8. Bibliografía

110. Boškoski I, Orlandini B, Papparella LG, et al. Robotics and artificial intelligence in gastrointestinal endoscopy: updated review of the literature and state of the art. *Current Robotics Reports*. 2021;2:43-54.
111. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *nature*. 2017;542(7639):115-118.

9. Anexos

9.1. Predictive power for thrombus detection after atrial appendage closure:
machine learning vs. classical methods

Article

Predictive Power for Thrombus Detection after Atrial Appendage Closure: Machine Learning vs. Classical Methods

Pablo Antúnez-Muñoz ^{1,2,*}, Víctor Vicente-Palacios ³, Pablo Pérez-Sánchez ^{1,2}, Jesús Sampredo-Gómez ¹, Antonio Sánchez-Puente ¹, Pedro Ignacio Dorado-Díaz ^{1,2}, Luis Nombela-Franco ⁴, Pablo Salinas ⁴, Hipólito Gutiérrez-García ⁵, Ignacio Amat-Santos ⁵, Vicente Peral ⁶, Antonio Morcuende ⁶, Lluís Asmarats ⁷, Xavier Freixa ⁸, Ander Regueiro ⁸, Berenice Caneiro-Queija ⁹, Rodrigo Estevez-Loureiro ⁹, Josep Rodés-Cabau ⁷, Pedro Luis Sánchez ^{1,2} and Ignacio Cruz-González ^{1,2}

¹ CIBERCV, University Hospital of Salamanca, 37007 Salamanca, Spain

² Biomedical Research Institute of Salamanca (IBSAL), 37007 Salamanca, Spain

³ Philips Ibérica, 28050 Madrid, Spain

⁴ Instituto Cardiovascular, Hospital Clínico San Carlos, IdISSC, 28040 Madrid, Spain

⁵ CIBERCV, Instituto de Ciencias del Corazón (ICICOR), Hospital Clínico Universitario de Valladolid, 47003 Valladolid, Spain

⁶ Department of Cardiology, Health Research Institute of the Balearic Islands (IdISBa), Hospital Universitari Son Espases, 07120 Palma, Spain

⁷ Quebec Heart and Lung Institute, Laval University, Quebec City, QC G1V 0A6, Canada

⁸ Institut Clinic Cardiovascular, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), 08036 Barcelona, Spain

⁹ University Hospital Alvaro Cunqueiro, 36312 Vigo, Spain

* Correspondence: pjantunez@gmail.com; Tel: +34-923291100



Citation: Antúnez-Muñoz, P.; Vicente-Palacios, V.; Pérez-Sánchez, P.; Sampredo-Gómez, J.; Sánchez-Puente, A.; Dorado-Díaz, P.I.; Nombela-Franco, L.; Salinas, P.; Gutiérrez-García, H.; Amat-Santos, I.; et al. Predictive Power for Thrombus Detection after Atrial Appendage Closure: Machine Learning vs. Classical Methods. *J. Pers. Med.* **2022**, *12*, 1413. <https://doi.org/10.3390/jpm12091413>

Academic Editor:
 Agata Bielecka-Dabrowa

Received: 14 June 2022
 Accepted: 21 August 2022
 Published: 30 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Device-related thrombus (DRT) after left atrial appendage (LAA) closure is infrequent but correlates with an increased risk of thromboembolism. Therefore, the search for DRT predictors is a topic of interest. In the literature, multivariable methods have been used achieving non-consistent results, and to the best of our knowledge, machine learning techniques have not been used yet for thrombus detection after LAA occlusion. Our aim is to compare both methodologies with respect to predictive power and the search for predictors of DRT. To this end, a multicenter study including 1150 patients who underwent LAA closure was analyzed. Two lines of experiments were performed: with and without resampling. Multivariate and machine learning methodologies were applied to both lines. Predictive power and the extracted predictors for all experiments were gathered. ROC curves of 0.5446 and 0.7974 were obtained for multivariate analysis and machine learning without resampling, respectively. However, the resampling experiment showed no significant difference between them (0.52 vs. 0.53 ROC AUC). A difference between the predictors selected was observed, with the multivariable methodology being more stable. These results question the validity of predictors reported in previous studies and demonstrate their disparity. Furthermore, none of the techniques analyzed is superior to the other for these data.

Keywords: left atrial appendage closure; device-related thrombosis; atrial fibrillation; machine learning; multivariable analysis; predictors

1. Introduction

Atrial fibrillation is the most prevalent arrhythmia worldwide, and more than one-third of the European population is expected to suffer from it in their lifetime [1]. This arrhythmia causes an inefficient atricular contraction, which increases the risk of thrombus formation, especially in the left atrium. It is worth noting that anticoagulation is indicated in those patients at high risk (CHA₂DS₂-VASc score ≥ 1 in men and ≥ 2 in women) to prevent thromboembolic events [2,3]. However, some patients are not suitable for this treatment, especially due to high risk of bleeding; therefore, percutaneous left atrial appendage

occlusion has emerged as an alternative in the last decade [4,5]. The vast majority (>95%) of the thrombi secondary to atrial fibrillation originate in the left atrial appendage secondary to a reduce flow velocity of the blood inside it during atrial fibrillation [1]. In this procedure, the appendage closure is performed by deploying a device inside it from a transeptal approach from a femoral venous access [4]. Different devices have been created, and their design has improved throughout the last decade. Different studies have shown that excluding this appendage was non-inferior to oral anticoagulation to reduce the risk of thromboembolic events; thus, anticoagulation treatment can be withdrawn. Moreover, the bleeding risk can be reduced compared with oral anticoagulation [6,7].

The success rate of left atrial appendage occlusion is high with a low rate of complications, especially the vascular ones in the access site. This rate continues to increase due to new emerging materials and techniques. However, between 3–4% cases present device-related thrombosis during the follow-up. It consists in the formation of a thrombus in the auricular face of the device before correct epithelization; this is the reason why anticoagulation or dual antiplatelet therapy should not be withdrawn during the first months after the procedure [8]. This complication, although not frequent, confers a high risk of stroke and systemic embolism, so anticoagulation should be reintroduced [9,10]. Moreover, this is especially an issue in those patients at high risk of bleeding. On the other hand, DRT can occur not only during the first months after left atrial appendage closure but at any time during follow-up, so high-risk patients could benefit from a closer follow-up [11–13]. Therefore, it is crucial to predict those patients at high risk of device-related thrombosis.

Several authors have recently searched for predictors of thrombus detection after left atrial appendage closure with quite disparate results [9–11,13–17]. This divergence is accompanied by the difficulty of the task and the methodology used to find these predictors: multivariable regression [18].

Although this classical statistical methodology is widely used in the medical field, some misconceptions can lead to misinterpretation of results. One of them, and perhaps the most repeated, is not distinguishing multivariable from multivariate terminology [19]. Whereas multivariate analysis refers to statistical models with two or more dependent variables, multivariable analysis refers to statistical models with multiple independent variables.

Another major problem with these techniques is that they do not report on the goodness-of-fit of the model [20]. In a vast majority of cases, the models focus mainly on the significance of the variables and do not provide a value that assesses the goodness-of-fit of the regression. Nonetheless, the assumption of providing goodness-of-fit does not preclude the model from being generalizable.

One solution to provide a proof that our model generalizes correctly is through resampling techniques [21] either with an external sample or with bootstrapping techniques.

Combining multivariate regression with propensity score analysis is another widespread technique [22]. However, although these techniques attempt to alleviate possible causal problems, they do not solve the aforementioned problems and are not free of controversy [23].

An alternative to find predictors is to apply machine learning models. Recent interest exists in applying machine learning techniques to clinical research, and cardiology is no exception [24].

Most machine learning applications applied to medicine are focused on trying to enable the physician to spend more time with the patient by eliminating routine tasks such as semi-automated labeling clinical reports [25] or assisting in the segmentation of medical images [26]. It is worth noting that the most successful applications are those in which human-machine interaction is part of the system [27]. However, in the case of precision medicine, machine learning techniques may not be entirely clear and can be misleading [28]. An example is the multitude of publications on the prediction of COVID-19 in medical imaging that have appeared in recent years and do not yield clear results or predictors [29].

Another drawback of thrombus analysis after atrial appendage closure is its low prevalence. This factor greatly affects the results reported with both classical statistical and machine learning methods. As discussed above for classical statistical techniques, one of the approaches to this problem is through resampling techniques [30]. However, in the case of machine learning, boosting and bagging techniques can be added [31].

Finally, the clinical databases available for this type of analysis are usually relatively small ($\sim 10^3$ cases) compared to those where machine learning is most effective ($\sim 10^6$ cases). In this context, and by combining the treatment of non-large databases with resampling techniques, steps have been taken towards a consistent clinical machine learning methodology [32,33].

Our aim is to determine whether any differences exist with respect to the predictive power between the classical and the machine learning methodology to predict thrombus after atrial appendage closure. Secondly, we want to extract the predictor variables of both methodologies and analyze their differences. To this end, we review in detail both methodologies, compare them with and without resampling techniques, and extract their predictors variables. Compared to previous studies, we report the goodness-of-fit of our models.

2. Materials and Methods

2.1. Data

A multicenter registry including 1150 consecutive patients who underwent left atrial appendage closure (LAAC) from 5 different international centers was utilized for this study [34–36]. Patient's follow-up and antithrombotic treatment at discharge were conducted according to each center's criteria. Both TEE and CT were performed at their respective centers to detect thrombus on the device during the first 3 months after left atrial appendage occlusion. The criteria used for the diagnosis were those proposed by Aminian et al. [13], defined as density on the left atrial aspect of the device not explained by imaging artifact or normal healing, visible in multiple planes, and in contact with the device. Patients without appropriately follow-up or with multiple missing data were excluded. Data were gathered retrospectively. Periprocedural variables of the registry were selected, including clinical and anatomic variables. However, the criteria for the diagnosis of device-related thrombosis were not universal for all the centers, and both transesophageal echocardiography and cardiac computerized tomography were used in the follow-up. Moreover, the diagnosis depended on each operator.

All data collection and the analysis fulfill the Declaration of Helsinki of 1975.

Other input variables such as clinical data have been already published in other studies [34–37].

Although both multivariable analysis and machine learning methodologies for predictor detection are widely known, it is important to highlight certain aspects to compare them. Both methodologies are described, and we detail the experiments performed.

2.2. Multivariable Analysis

The process for the selection of predictors is divided into two stages.

First, a univariate analysis of the analyzed variables is performed. In the case of continuous variables, one-way ANOVA or *t*-test is used. If the variables are categorical, the chi-square test is utilized. In the case of an ordinal variable, the Mann–Whitney U test is applied. Once each of the corresponding tests has been performed, the significant variables are selected (p -value < 0.05). This procedure represents a dimensional reduction process, especially in those cases where the number of variables is large.

Afterwards, these previously selected variables are incorporated into the multivariable regression model. Once the regression model is fitted, the significance of this set of variables is assessed. The final predictors are those that meet the chosen significance criterion (p -value < 0.05).

This process was performed on the entire database, and no resampling process was undertaken.

2.3. Machine Learning Feature Selection

Even though many methods for selecting predictors within machine learning exist, we chose one of the most widely used techniques: meta-transformer based on importance weights [38].

Meta-transformer based on importance weights works as follows: once the algorithm has been fitted, variables are selected based on their specific weight in the model. A threshold is set for this selection. Those variables that remain below this threshold are discarded from the model.

This process is always accompanied by either internal or external resampling depending on the characteristics of our data.

An independent external sample is ideal to demonstrate that our model is generalizable. However, in a clinical setting, it is not always possible either due to imbalanced data or size [39]. To alleviate this problem, we applied cross-validation. More specifically, we used k -folds with repetition [33] and shuffle split [40].

The k -fold procedure divides the main sample into k groups of samples or folds of equal size whenever possible. The algorithm is adjusted using the $k-1$ folds, and the fold that is left out is used for testing. This process is repeated n times with resampling.

The shuffle split method generates a defined number of independent splits of train and test data sets. The samples are first shuffled and then split into these subsets.

The suitability of the selected variables is made based on a chosen metric. Our chosen metric was the receiver operating characteristic (ROC AUC) curve.

2.4. Experiments Performed

To compare both methodologies, two lines of experiments were followed: multivariable and machine learning.

Logistic regression with regularization [41], random forest [42], and gradient boosting [43] were the machine learning algorithms used [44]. Logistic regression without regularization was utilized as the classic algorithm for the multivariable analysis.

Predictors were collected along with their associated values (p -values or weights). The area under the ROC (AUC ROC) curve, the specificity, and sensitivity of each experiment were also reported.

The cut-off point of the probability curve was determined by optimizing the ROC curve resulting from each experiment.

2.4.1. No Resampling Experiments

Both predictor selection methodologies were applied to the entire dataset.

2.4.2. Resampling Experiments

A 2-fold resampling technique with 5 repetitions was applied in addition to a shuffle split with a test size of 40% and 2 splits.

The selection of the resampling parameters is determined by the low prevalence of thrombus after atrial appendage closure. These parameters ensure a minimum number of cases of thrombus in each fold or split.

2.5. Experiments Design

Based on the two lines of research, 4 experiments were set up. Experiment I and II refer to research Section 2.4.1, while experiment III and IV refer to Section 2.4.2. Figure 1 shows the workflow of the 4 performed experiments.

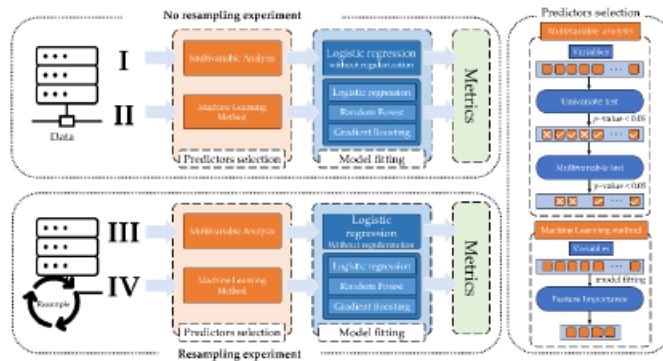


Figure 1. Experiment workflow.

2.6. Software

The programming language used in this article was Python, and the libraries used were scikit-learn [40], statsmodels [45], and Scipy [46].

The developed code is open source and can be consulted at https://github.com/IA-Cardiologia-husa/LAAC_Thrombus_detection_MLvsClassical (last accessed on 8 August 2022) [47].

3. Results

Out of the total of 1150 patients included in the registry, 813 were finally analyzed. Baseline characteristics of the patients are summarized in Table 1. Those patients with significant missing values or without adequate follow-up were excluded. Device-related thrombus was observed in 35 patients (4.31%).

Table 1. Baseline Characteristics. Values are mean ± SD or % (n/N).

Variables	All Patients (N = 813)
Age (Years)	75.62 ± 8.43 (813)
Gender (Male)	61.01 (496/813)
Weight (Kg)	74.96 ± 15.55 (806)
Height (cm)	164.73 ± 9.41 (804)
BMI *	27.55 ± 4.94 (804)
BSA *	1.81 ± 0.21 (804)
HTA * (Yes)	87.33 (710/813)
Diabetes (Yes)	33.7 (274/813)
Smoking (Yes)	14.37 (127/696)
Prior stroke (Yes)	35.36 (279/789)
Prior hemorrhagic stroke (Yes)	24.8 (185/746)
Prior systemic embolization (Yes)	9.85 (78/792)
Prior vascular disease (Yes)	20.54 (167/813)
Previous CAD * (Yes)	31.89 (258/809)
Previous MI * (Yes)	20.08 (154/767)
Prior mitral valvular surgery	2.70 (22/813)
CHADS2	2.96 ± 1.24 (768)
CHA2DS2VASc	4.44 ± 1.54 (813)
HAS-BLED	3.71 ± 1.05 (811)
Prior bleeding episodes	1.57 ± 1.15 (768)
Labile INR * (Yes)	11.07 (90/813)

Table 1. Cont.

Variables	All Patients (N = 813)
Previously on ASA* (Yes)	30.26 (246/813)
Previously on ADP inhibitor* (Yes)	11.07 (90/813)
Previous antiVITK treatment (Yes)	23.37 (190/813)
Previous severe mitral regurgitation (Yes)	15.23 (122/801)
Previous severe mitral stenosis (Yes)	0.26 (2/765)
Procedural device reposition (If available)	0.59 ± 1 (570)
Procedure contrast volume	134.03 ± 88.09 (695)
periprocedural pericardial effusion (Yes)	5.9 (48/813)
ADP inhibitor at discharge* (Yes)	62.64 (508/811)

*BMI, body mass index; BSA, body surface area; HTA, arterial hypertension; CAD, coronary artery disease; MI, myocardial infarction; INR, international normalized ratio; ASA, acetylsalicylic acid; ADP, adenosine diphosphate receptor.

Results are presented in two subsections: predictive power and predictor variables.

3.1. Predictive Power

Table 2 shows the results of the different experiments of ROC AUC, sensitivity, and specificity based on optimizing ROC AUC.

Table 2. Results of the four experiments. The adjustment of the metrics was performed by optimizing the ROC AUC value.

Experiment	Resample	Feature Selection	Model	ROC AUC	Sensitivity	Specificity
I	-	Multivariable analysis	LR*	0.5456	0.8857	0.2198
II	-	Select from model	LR	0.7974	0.6857	0.7776
			RF	1.0	1.0	1.0
			GB	1.0	1.0	1.0
III	Shuffle split k-fold	Multivariable analysis	LR*	0.4387 ± 0.0904	0.1811 ± 0.1143	0.6544 ± 0.2263
		Multivariable analysis	LR*	0.5174 ± 0.0531	0.4318 ± 0.3255	0.5504 ± 0.3173
IV	Shuffle split	Select from model	LR	0.4838 ± 0.0118	0.1697 ± 0.0724	0.6960 ± 0.0109
			RF	0.3989 ± 0.0240	0.0 ± 0.0	0.9014 ± 0.0639
	k-fold	Select from model	GB	0.4614 ± 0.0437	0.0444 ± 0.0459	0.9217 ± 0.0073
			LR	0.5325 ± 0.0349	0.2804 ± 0.0771	0.7634 ± 0.0518
			RF	0.4250 ± 0.354	0.0418 ± 0.0462	0.7049 ± 0.0193
			GB	0.4893 ± 0.0668	0.0754 ± 0.0501	0.9005 ± 0.0176

LR, logistic regression; RF, random forest; GB, gradient boosting. (*) LR without regularization.

In experiment I, sensitivity and specificity are imbalanced. The lack of regularization in the fitted logistic regression explains this result. On the other hand, experiment II presents a perfect fit for the non-linear algorithms (gradient boosting and random forest). Nonetheless, experiment II model is clearly overfitted for all samples.

Experiments III and IV were performed using the kfold technique with K = 2 and 5 repetitions and shuffle split method with two splits and 40% of test size. In contrast to experiments I and II, experiments III and IV gathered the values of the metrics for each of their iterations. In experiment III, the results show a better performance for the k-fold resampling scenario. In experiment IV, the best result in terms of ROC AUC is obtained with k-fold resampling in combination with logistic regression. Furthermore, this combination presents the best sensitivity—specificity balance.

Subsequent analyses were performed considering only the k-fold resampling method and logistic regression algorithm. This combination obtained the most plausible characteristics for experiment II and the most suitable for experiments III and IV.

Figure 2 shows the analyzed metrics in each of the iterations. Even though no significant differences between the two experiments in terms of the mean values of the analyzed

metrics were found, Figure 2 shows that experiment III is less variable than the experiment IV. Nevertheless, the low sensitivity obtained in experiment IV and its low AUC ROC proves a low predictive power.

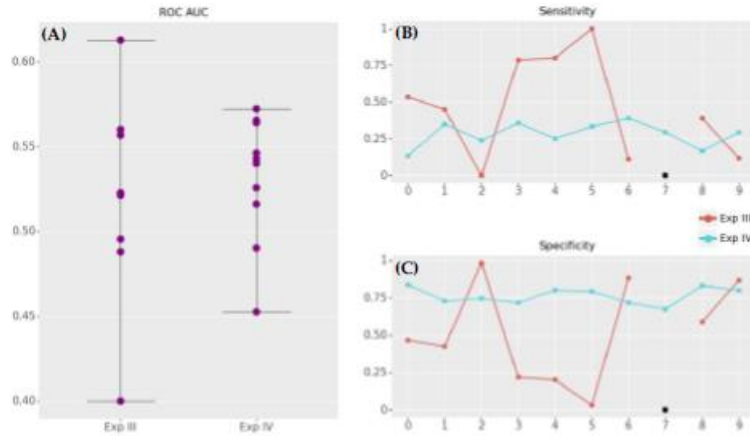


Figure 2. ROC AUC (A), sensitivity (B), and specificity (C) results for logistic regression in combination with k-fold resampling method (k = 2, 5 repetitions).

3.2. Predictor Variables

Table 3 shows the significant variables after univariate and multivariable analysis (experiment I). Only 3 variables out of 76 were significant in the univariate test. Among these three variables, just one (HAS-BLED score) was found to be significant in the multivariate test.

Table 3. Experiment I predictor variables selection. Results from univariate and multivariable test on classic method.

Var	Univariate p-Value	Multivariable p-Value
CHADS2	<0.01	0.852
CHA2DS2-VASC	<0.01	0.414
HAS-BLED	<0.01	<0.01

Figure 3 shows the variables selected by the machine learning methodology on experiment II. This figure illustrates the positive or negative importance of each variable on the variable to be predicted.

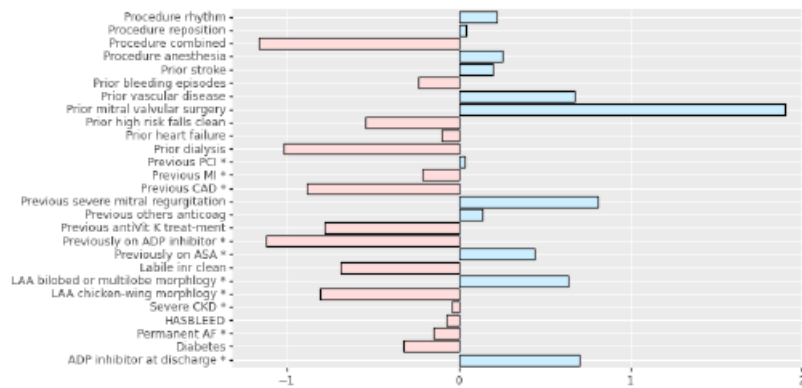


Figure 3. Variable weights results on experiment II using logistic regression. * AF, atrial fibrillation; ASA, acetylsalicylic acid; ADP, adenosine diphosphate receptor; CAD, coronary artery disease; CKD, chronic kidney disease; INR, international normalized ratio; LAA, left atrial appendage; MI, myocardial infarction; PCI, percutaneous coronary intervention.

Table 4 shows the number of occurrences the predictors were selected for experiments III and IV. In the case of experiment III, both the number of univariate and multivariable tests are shown.

Table 4. Occurrences of predictors for each feature selection method. Results for k-fold resampling method (k = 2, 5 repetitions).

Variable	Experiment III		Experiment IV
	Univariate	Multivariable	
Previous CAD *	-	-	10
HAS-BLED	10	8	10
Previous antiVtK treatment	-	-	10
CHADS2	10	-	6
CHA2DS2VASc	10	-	6
Chicken-wing LAA morphology	-	-	9
Previous severe mitral regurgitation	1	1	9
Prior bleeding episodes	-	-	9
Labile INR *	-	-	9
ADP inhibitor at discharge	-	-	8
Procedural device reposition	-	-	8
Severe CKD *	-	-	8
Periprocedural pericardial effusion	1	-	8
Diabetes	-	-	7
Prior stroke	-	-	7
Previously on ADP inhibitor *	-	-	7
Previously on ASA *	-	-	7
Prior systemic embolization	-	-	7
Prior vascular disease	-	-	6
Previous MI *	-	-	6
Prior hemorrhagic stroke	-	-	6
Procedure contrast volume	1	1	-
Prior valve surgery (mitral)	1	1	5

*CAD, coronary artery disease; INR, international normalized ratio; CKD, chronic kidney disease; ADP, adenosine diphosphate receptor; ASA, acetylsalicylic acid; MI, myocardial infarction.

Experiment III, similarly to experiment I, shows robustness in variable selection. In the case of experiment IV, only three variables were selected in each iteration (previous CAD, HAS-BLED, and pre-coumadin warfarin).

To analyze p -values (classical methodology) and weights (machine learning methodology) fluctuation due to resampling, a predictor variation rate for each variable was defined as described in Formula (1).

$$\text{predictor variation rate} = \frac{V_i - V_{ED}}{\sigma} \quad (i = \text{iteration}, ED = \text{entire dataset}) \quad (1)$$

V are p -values in the case of the classical methodology and weights for the machine learning methodology.

The numerator of Formula (1) is the difference between the value (p -value or weights) of each iteration of the resampling and the value obtained in the experiments performed on the entire database. The denominator is the standard deviation of the values collected for the predictor variable in the total number of iterations.

Figure 4 shows the predictor variational rate for the selected variables. To assess whether there were significant differences for the variance ratio between the two methodologies, a Mann–Whitney U test was performed, preceded by a Levene test. The Levene test was significant (p -value = 0.039), while the Mann–Whitney U test was not (p -value = 0.937).

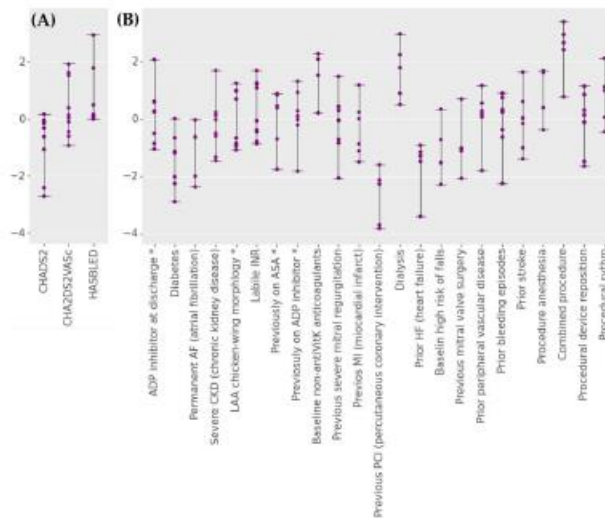


Figure 4. Predictor variation rate for each fold on the (A) experiment III and (B) experiment IV using logistic regression. Both experiments results used k -fold resampling method ($k = 2, 5$ repetitions). * ASA, acetylsalicylic acid; ADP, adenosine diphosphate receptor; LAA, left atrial appendage.

4. Discussion

Device-related thrombus continues to be one of the main challenges in LAA closure. Device endothelialization after the procedure is not well-known yet. Therefore, recent research has focused on DRT predictors identification and different treatment strategies after occlusion to prevent thrombosis. For instance, one prospective non-randomized study analyzed different antithrombotic treatments after LAAC and showed that half dose of direct oral anticoagulants (DOACs) can reduce DRT [48]. Furthermore, different prospective trials

are currently ongoing trying to elucidate this question. On the other hand, new materials and better devices are being developed in order to reduce the thrombotic risk, specially the first months before the endothelization is completed.

Different studies in the literature have tried previously to find predictors of DRT. However, each registry found different predictors and always with a suboptimal power of prediction. Classical statistical analysis was used in all of them. Simard et al. found both major and minor risk factors [10]. The major ones were iatrogenic pericardial effusion and a hypercoagulable state. On the other hand, the minor ones were deep LAAC implant (over 10 mm from the pulmonary ridge), renal insufficiency, and AF other than paroxysmal. They found that the presence of one major factor or two minor factors would double the risk of DRT [10]. Dukkupati et al. found different predictors such as history of TIA or stroke, vascular disease, and the LAA diameter [11]. Even though permanent AF was the only one risk factor found in both registries, the predictive power was modest. Moreover, other believed predictors such as pericardial effusion CKD or vascular disease were included in our models. Conversely, CHA2DS2VASC score was defined as DRT in a Watchman 2.5 registry [16].

In our study, we used a multicenter registry of highly trained centers. The incidence was 4.31%, similar to that described in previous registries, which seems to be very representative of the current situation worldwide [10]. Although its frequency is low, the clinical repercussion of it remains a serious and relevant issue in LAA occlusion.

We found that there is no significant difference in the predictive power of thrombus after atrial appendage closure between the classical multivariable and machine learning methodology. In both cases, the predictive power was low. Therefore, device-related thrombus remains difficult to predict after LAAC.

On the other hand, differences between the predictor variables extracted by the two methodologies were found. Whereas in the case of the multivariable methodology, the predictors were mainly constant, the machine learning methodology utilized several variables to fit its models.

Differentiated results were obtained within the two lines of experiments (without and with resampling). On the one hand, the predictive power of the machine learning methodology (LR: 0.7974, RF: 1.0, GB: 1.0 ROC AUC) was clearly higher than that of the multivariable methodology (0.5446 ROC AUC). On the other hand, the machine learning methodology presented a more balanced trade-off between specificity and sensitivity than the multivariable methodology. Although the difference is remarkable, it may not be representative because both the fitting and the evaluation were performed on the entire database, which clearly demonstrates overfitting [49]. This overfitting is more obvious in the case of gradient boosting and random forest algorithms.

Experiments with resampling revealed that no predictive power differences between the techniques were found. This result evidences the overfitting of the machine learning model without resampling. Nonetheless, it is noteworthy that in both methods, the ROC AUC values are low, minimally exceeding those of a random model. We should also mention the low sensitivity value of both models, which corroborates that neither method is able to predict thrombus.

These unsatisfactory results do not allow the identification of patients at increased risk of thromboembolic events. This problem has a high relevance since most of these patients also have a high risk of hemorrhage. Furthermore, although most of these complications manifest themselves within the first few months after left atrial appendage occlusion, some individuals may develop thrombus beyond the first six months [11]. Better predictive power could allow physicians to perform longer and more intensive follow-up through TEE and CT evaluation in those patients at higher risk.

Albeit no differences were found in the predictive power, differences were observed in the predictive variables selected by the different methods. The multivariable methodology proved to be more robust when selecting predictor variables. In both the experiments with and without resampling, the prominent variable was the scoring system for assessing

1-year risk of major bleeding in patients with atrial fibrillation (HAS-BLED). In contrast, the machine learning methodology employed a broader set of variables among which it repeated only three in both experiments: previous coronary artery disease, previous anticoagulant intake, and HAS-BLED. Although this divergence of choice occurred, the prediction of thrombus was similar.

Our findings question first the advantages of machine learning techniques over classical ones [50] and, second, prove the divergence of predictors in previous studies [9–11,13–16]. Older age [9], history of stroke [9,11], permanent atrial fibrillation, vascular disease, left atrial appendage diameter, left atrial appendage diameter [11], smoking, sex [14], and LAA orifice width [13] were some of the predictors reported. Moreover, even a randomized trial has shown that antithrombotic treatment at discharge with half dose DOACs has shown to reduce significantly the incidence of device-related thrombus, and none of them have reported it as a powerful predictor [48]. Among all the studies, only one performed bootstrapping [15], and in none of them was the predictive power reported.

As a main limitation, our study included only three machine learning algorithms and two resampling techniques. Additionally, hyperparameter tuning techniques were not employed. However, the main idea of the study was to compare methodologies (multivariable and machine learning). Therefore, the comparison of these with a common algorithm (logistic regression) is appropriate.

Another limitation is the lack of a structured protocol including postprocedural antithrombotic treatment and scheduled follow-up. However, different protocols for postprocedural antithrombotic treatment have been reported in LAAC studies and without a formal consensus. Furthermore, patients undergoing LAAC are very heterogeneous, which means that each patient may benefit from a different discharge treatment. Some of them undergo LAAC because of new thromboembolic events despite correct oral anticoagulation. On the other hand, others are at high risk of bleeding or had had a relevant bleeding episode on oral anticoagulation, so continuing with anticoagulation is forbidden. Finally, despite the large number of patients, the retrospective nature of the study is also a limitation, and only new hypotheses can be made about the findings observed.

We have been thorough in performing different experiments to make them comparable. Furthermore, we have provided the rate of variation of predictors to contrast both variable selection techniques.

In this article, we used a structured multicenter hemodynamic database similar to those used in recent publications [9–11,13–16]. Our database presented a clear imbalance between cases with and without thrombus after appendage closure. Given these circumstances, future research would have to address the problem attempting to balance the cases and incorporate other sorts of variables into the studies (e.g., image or hematologic data). Additionally, if machine learning is not capable of predicting DRT, there is something we still do not know about DRT. That is why several studies are trying to predict them with heterogeneous results. However, using higher databases or other anatomic characteristics, previous device rejection such as stent restenosis, procedural issues, or hematologic coagulation disorders could offer more insights into both classical and machine learning models.

We firmly believe that in order to report a predictor, it is necessary first to perform a cross or external validation and second to report the predictive power. These two conditions are independent of the method used and are sufficient to ensure the validity of the predictors.

5. Conclusions

DRT after LAA closure is rare but correlates with an increased risk of thromboembolism. We analyzed a multicentric registry, and we found that there is no significant difference in the predictive power of thrombus after atrial appendage closure between the classical multivariable and machine learning methodology. In both cases, the predictive power was low. These results question the validity of predictors reported in previous studies and demonstrate their disparity. Furthermore, none of the techniques analyzed is

superior to the other for these data. Nevertheless, machine learning models could achieve higher predicting power in larger registries.

Author Contributions: Conceptualization, V.V.-P., I.C.-G., P.P.-S. and P.A.-M.; methodology, V.V.-P.; software, P.P.-S., V.V.-P. and J.S.-G.; resources, P.S., L.N.-E., I.A.-S., H.G.-G., V.P., A.M., L.A., X.F., A.R., B.C.-Q., R.E.-L., J.R.-C. and I.C.-G.; data curation, V.V.-P. and J.S.-G.; writing—original draft preparation, V.V.-P., P.P.-S., P.A.-M. and I.C.-G.; writing—review and editing, A.S.-P., V.V.-P. and I.C.-G.; visualization, P.P.-S.; supervision, I.C.-G. and P.L.S.; project administration, P.L.D.-D.; funding acquisition, I.C.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been funded by Instituto de Salud Carlos III (ISCIII) through the project “PI19/00658. Modelo predictivo de trombosis sobre dispositivo de cierre de orejuela mediante machine learning e inteligencia artificial—Estudio TROMIA” and co-funded by the European Union. It has also been funded by Consejería de Sanidad de Castilla y León through the project “GRS 3031/A/19. Desarrollo de Modelo predictivo de trombosis sobre dispositivo de cierre de orejuela mediante inteligencia artificial”.

Institutional Review Board Statement: This study was approved by our local Ethic Committee de Ética de la Investigación con medicamentos del Área de Salud de Salamanca. Date of approval: 27th March 2022. Approval Code: PI 2019 09 366.

Informed Consent Statement: Informed consent was obtained for the procedure from all subjects involved in the study.

Conflicts of Interest: Ignacio Cruz-Gonzalez is proctor for Abbott, Boston Scientific, Lifetech. Victor Vicente-Palacios is a Philips Healthcare employee.

References

- Developed with the Special Contribution of the European Heart Rhythm Association (EHRA); Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS); Authors/Task Force Members; Camm, A.J.; Kirchhof, P.; Lip, G.Y.; Schotten, U.; Savelieva, I.; Ernst, S.; Van Gelder, I.C.; et al. Guidelines for the Management of Atrial Fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur. Heart J.* **2010**, *31*, 2369–2429. [PubMed]
- Lip, G.Y.H.; Nieuwlaet, R.; Pisters, R.; Lane, D.A.; Crijns, H.J.G.M. Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach. *Chest* **2010**, *137*, 263–272. [CrossRef] [PubMed]
- Rivera-Caravaca, J.M.; Roldan, V.; Esteve-Pastor, M.A.; Valdes, M.; Vicente, V.; Lip, G.Y.; Marin, F. Long-Term Stroke Risk Prediction in Patients With Atrial Fibrillation: Comparison of the ABC-Stroke and CHA₂DS₂-VAS_c Scores. *J. Am. Heart Assoc.* **2017**, *6*, e006490. [PubMed]
- Glikson, M.; Wolff, R.; Hindricks, G.; Mandrola, J.; Camm, A.J.; Lip, G.Y.; Fauchier, L.; Betts, T.R.; Lewalter, T.; Saw, J. EHRA/EAPCI Expert Consensus Statement on Catheter-Based Left Atrial Appendage Occlusion—An Update. *EP Eur.* **2020**, *22*, 184. [CrossRef]
- Cruz-González, I.; Trejo-Velasco, B. Percutaneous Left Atrial Appendage Occlusion in the Current Practice. *Kardiol. Pol. (Pol. Heart J.)* **2021**, *79*, 255–268. [CrossRef]
- Reddy, V.Y.; Doshi, S.K.; Kar, S.; Gibson, D.N.; Price, M.J.; Huber, K.; Horton, R.P.; Buchbinder, M.; Neuzil, P.; Gordon, N.T. 5-Year Outcomes after Left Atrial Appendage Closure: From the PREVAIL and PROTECT AF Trials. *J. Am. Coll. Cardiol.* **2017**, *70*, 2964–2975.
- Osmancik, P.; Herman, D.; Neuzil, P.; Hala, P.; Taborsky, M.; Kala, P.; Poloczek, M.; Stasek, J.; Haman, L.; Branny, M. 4-Year Outcomes After Left Atrial Appendage Closure Versus Nonwarfarin Oral Anticoagulation for Atrial Fibrillation. *J. Am. Coll. Cardiol.* **2022**, *79*, 1–14.
- Freeman, J.V.; Higgins, A.Y.; Wang, Y.; Du, C.; Friedman, D.J.; Daimee, U.A.; Minges, K.E.; Pereira, L.; Goldsweig, A.M.; Price, M.J.; et al. Antithrombotic Therapy After Left Atrial Appendage Occlusion in Patients With Atrial Fibrillation. *J. Am. Coll. Cardiol.* **2022**, *79*, 1785–1798. [CrossRef]
- Fauchier, L.; Cinaud, A.; Brigadeau, F.; Lepillier, A.; Pierre, B.; Abbey, S.; Fatemi, M.; Franceschi, F.; Guedeney, P.; Jacon, P.; et al. Device-Related Thrombosis After Percutaneous Left Atrial Appendage Occlusion for Atrial Fibrillation. *J. Am. Coll. Cardiol.* **2018**, *71*, 1528–1536. [CrossRef]
- Simard, T.; Jung, R.G.; Lehenbauer, K.; Piayda, K.; Pracoń, R.; Jackson, G.G.; Flores-Umanzor, E.; Faroux, L.; Korsholm, K.; Chun, J.K.R.; et al. Predictors of Device-Related Thrombus Following Percutaneous Left Atrial Appendage Occlusion. *J. Am. Coll. Cardiol.* **2021**, *78*, 297–313. [CrossRef]

11. Dukkipati, S.R.; Kar, S.; Holmes, D.R.; Doshi, S.K.; Swarup, V.; Gibson, D.N.; Maini, B.; Gordon, N.T.; Main, M.L.; Reddy, V.Y. Device-Related Thrombus After Left Atrial Appendage Closure: Incidence, Predictors, and Outcomes. *Circulation* **2018**, *138*, 874–885. [[CrossRef](#)]
12. Sedaghat, A.; Vij, V.; Al-Kassou, B.; Gloekler, S.; Galea, R.; Fürholz, M.; Meier, B.; Valgimigli, M.; O'Hara, G.; Arzamendi, D.; et al. Device-Related Thrombus After Left Atrial Appendage Closure. *Circ. Cardiovasc. Interv.* **2021**, *14*, 544–553. [[CrossRef](#)] [[PubMed](#)]
13. Aminian, A.; Schmidt, B.; Mazzone, P.; Berti, S.; Fischer, S.; Montorfano, M.; Lam, S.C.C.; Lund, J.; Asch, E.M.; Gage, R.; et al. Incidence, Characterization, and Clinical Impact of Device-Related Thrombus Following Left Atrial Appendage Occlusion in the Prospective Global AMPLATZER Amulet Observational Study. *JACC Cardiovasc. Interv.* **2019**, *12*, 1003–1014. [[CrossRef](#)] [[PubMed](#)]
14. Saw, J.; Tzikas, A.; Shakir, S.; Gafour, S.; Omran, H.; Nielsen-Kudsk, J.E.; Kefer, J.; Aminian, A.; Berti, S.; Santoro, G.; et al. Incidence and Clinical Impact of Device-Associated Thrombus and Peri-Device Leak Following Left Atrial Appendage Closure With the Amplatzer Cardiac Plug. *JACC Cardiovasc. Interv.* **2017**, *10*, 391–399. [[CrossRef](#)]
15. Budts, W.; Laenens, D.; Van Calenbergh, F.; Vermeersch, P.; De Potter, T.; Aminian, A.; Benit, E.; Stammen, F.; Kefer, J. Left Atrial Appendage Occlusion with the Amplatzer Cardiac Plug Could Improve Survival and Prevent Thrombo-Embolic and Major Bleeding Events in Atrial Fibrillation Patients with Increased Bleeding Risk. *Acta Cardiol.* **2016**, *71*, 135–143. [[CrossRef](#)]
16. Kaneko, H.; Neuss, M.; Weissenborn, J.; Butter, C. Predictors of Thrombus Formation after Percutaneous Left Atrial Appendage Closure Using the WATCHMAN Device. *Heart Vessels.* **2017**, *32*, 1137–1143. [[CrossRef](#)] [[PubMed](#)]
17. Vij, V.; Playda, K.; Nelles, D.; Gloekler, S.; Galea, R.; Fürholz, M.; Meier, B.; Valgimigli, M.; O'Hara, G.; Arzamendi, D.; et al. Clinical and Echocardiographic Risk Factors for Device-Related Thrombus after Left Atrial Appendage Closure: An Analysis from the Multicenter EUROCD-DRT Registry. *Clin. Res. Cardiol.* **2022**, 1–10. [[CrossRef](#)]
18. Sobh, M.; Cleophas, T.J.; Hadj-Chaib, A.; Zwiderman, A.H. Clinical Trials: Odds Ratios and Multiple Regression Models—Why and How to Assess Them. *Am. J. Ther.* **2008**, *15*, 44–52. [[CrossRef](#)]
19. Hidalgo, B.; Goodman, M. Multivariate or Multivariable Regression? *Am. J. Public Health* **2013**, *103*, 39–40. [[CrossRef](#)] [[PubMed](#)]
20. Ranganathan, P.; Pramesh, C.S.; Aggarwal, R. Common Pitfalls in Statistical Analysis: Logistic Regression. *Perspect. Clin. Res.* **2017**, *8*, 148–151. [[CrossRef](#)]
21. De Bin, R.; Janitza, S.; Sauerbri, W.; Boulesteix, A.-L. Subsampling versus Bootstrapping in Resampling-Based Model Selection for Multivariable Regression. *Biometrics* **2016**, *72*, 272–280. [[CrossRef](#)] [[PubMed](#)]
22. Kahlert, J.; Gribsholt, S.B.; Gammeler, H.; Dekkers, O.M.; Luta, G. Control of Confounding in the Analysis Phase—an Overview for Clinicians. *Clin. Epidemiol.* **2017**, *9*, 195–204. [[CrossRef](#)]
23. Pearl, J. Remarks on the Method of Propensity Score. *Stat. Med.* **2009**, *28*, 1415–1416. [[CrossRef](#)]
24. Dorado-Díaz, P.L.; Sampedro-Gómez, J.; Vicente-Palacios, V.; Sánchez, P.L. Applications of Artificial Intelligence in Cardiology. The Future Is Already Here. *Rev. Española de Cardiol. (Engl. Ed.)* **2019**, *72*, 1065–1075. [[CrossRef](#)]
25. Pérez-Sánchez, P.; Vicente-Palacios, V.; Barreiro, M.; Díaz-Peláez, E.; Sánchez Puente, A.; Sampedro-Gómez, J.; García-Galindo, A.; Dorado-Díaz, P.; Sánchez, P. Automatic Classification of Valve Diseases Through Natural Language Processing in Spanish and Active Learning. In *International Conference on Bioengineering and Biomedical Signal and Image Processing, Proceedings of the First International Conference, BIOMESIP 2021, Meloneras, Gran Canaria, Spain, 19–21 July 2021*; Springer: Cham, Switzerland, 2021; pp. 39–50. ISBN 978-3-030-88162-7.
26. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep Learning for Cardiac Image Segmentation: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [[CrossRef](#)] [[PubMed](#)]
27. Maadi, M.; Akbarzadeh Khorshidi, H.; Aickelin, U. A Review on Human–AI Interaction in Machine Learning and Insights for Medical Applications. *Int. J. Environ. Res. Public Health* **2021**, *18*, 2121. [[CrossRef](#)]
28. Wilkinson, J.; Arnold, K.F.; Murray, E.J.; van Smeden, M.; Carr, K.; Sippy, R.; de Kamps, M.; Beam, A.; Konigorski, S.; Lippert, C.; et al. Time to Reality Check the Promises of Machine Learning-Powered Precision Medicine. *Lancet. Digit. Health* **2020**, *2*, e677–e680. [[CrossRef](#)]
29. Roberts, M.; Driggs, D.; Thorpe, M.; Gilbert, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.L.; Etmann, C.; McCague, C.; Beer, L.; et al. Common Pitfalls and Recommendations for Using Machine Learning to Detect and Prognosticate for COVID-19 Using Chest Radiographs and CT Scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
30. Kraiem, M.S.; Sánchez-Hernández, E.; Moreno-García, M.N. Selecting the Suitable Resampling Strategy for Imbalanced Data Classification Regarding Dataset Properties. An Approach Based on Association Models. *Appl. Sci.* **2021**, *11*, 8546. [[CrossRef](#)]
31. Yaman, E.; Subasi, A. Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification. *Biomed. Res. Int.* **2019**, *2019*, 9152506. [[CrossRef](#)]
32. Avram, R.; Olgin, J.E.; Tison, G.H. The Rise of Open-Sourced Machine Learning in Small and Imbalanced Datasets: Predicting In-Stent Restenosis. *Can. J. Cardiol.* **2020**, *36*, 1574–1576. [[CrossRef](#)] [[PubMed](#)]
33. Sampedro-Gómez, J.; Dorado-Díaz, P.L.; Vicente-Palacios, V.; Sánchez-Puente, A.; Jiménez-Navarro, M.; Roman, J.A.S.; Galindo-Villardón, P.; Sánchez, P.L.; Fernández-Avilés, F. Machine Learning to Predict Stent Restenosis Based on Daily Demographic, Clinical, and Angiographic Characteristics. *Can. J. Cardiol.* **2020**, *36*, 1624–1632. [[CrossRef](#)] [[PubMed](#)]
34. Faroux, L.; Cruz-González, L.; Arzamendi, D.; Freixa, X.; Nombela-Franco, L.; Peral, V.; Caneiro-Queija, B.; Mangieri, A.; Trejo-Velasco, B.; Asmarats, L.; et al. Effect of Glomerular Filtration Rates on Outcomes Following Percutaneous Left Atrial Appendage Closure. *Am. J. Cardiol.* **2021**, *145*, 77–84. [[CrossRef](#)] [[PubMed](#)]

35. Faroux, L.; Cruz-González, I.; Arzamendi, D.; Freixa, X.; Nombela-Franco, L.; Peral, V.; Caneiro-Queija, B.; Mangieri, A.; Trejo-Velasco, B.; Asmarats, L.; et al. Short-Term Direct Oral Anticoagulation or Dual Antiplatelet Therapy Following Left Atrial Appendage Closure in Patients with Relative Contraindications to Chronic Anticoagulation Therapy. *Int. J. Cardiol.* **2021**, *333*, 77–82. [CrossRef]
36. Mesnier, J.; Cruz-González, I.; Arzamendi, D.; Freixa, X.; Nombela-Franco, L.; Peral, V.; Caneiro-Queija, B.; Mangieri, A.; Trejo-Velasco, B.; Asmarats, L.; et al. Early Discontinuation of Antithrombotic Treatment Following Left Atrial Appendage Closure. *Am. J. Cardiol.* **2022**, *171*, 91–98. [CrossRef]
37. Regueiro, A.; Cruz-Gonzalez, I.; Bethencourt, A.; Nombela-Franco, L.; Champagne, J.; Asmarats, L.; Jiménez-Quevedo, P.; Rodríguez-Gabella, T.; Rama-Merchan, J.C.; Puri, R.; et al. Long-Term Outcomes Following Percutaneous Left Atrial Appendage Closure in Patients with Atrial Fibrillation and Contraindications to Anticoagulation. *J. Intern. Card. Electrophysiol.* **2018**, *52*, 53–59. [CrossRef]
38. Saeyns, Y.; Inza, I.; Larrañaga, P.A. Review of Feature Selection Techniques in Bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [CrossRef]
39. Kamalov, F.; Thabtah, F.; Leung, H.H. Feature Selection in Imbalanced Data. *Ann. Data. Sci.* **2022**. [CrossRef]
40. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
41. Goodfellow, I.; Bengio, Y.; Courville, A. Machine Learning Basics. In *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; pp. 216–218.
42. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
43. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
44. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why Do Tree-Based Models Still Outperform Deep Learning on Tabular Data? *arXiv* **2022**, arXiv:2207.08815.
45. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. Available online: <https://pdfs.semanticscholar.org/3a27/6417e5350e29cb6bf04ea5a4785601d5a215.pdf> (accessed on 20 June 2022).
46. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
47. Pérez-Sánchez, P.; Vicente-Palacios, V.; Sampedro-Gómez, J. IA-Cardiologia-Husa/LAAC_Thrombus_detection_MLvsClassical: Update 2022. Available online: https://github.com/IA-Cardiologia-husa/LAAC_Thrombus_detection_MLvsClassical (accessed on 20 June 2022).
48. Della Rocca, D.G.; Magnocavallo, M.; Di Biase, L.; Mohanty, S.; Trivedi, C.; Tarantino, N.; Gianni, C.; Lavallo, C.; Van Niekerk, C.J.; Romero, J. Half-Dose Direct Oral Anticoagulation versus Standard Antithrombotic Therapy after Left Atrial Appendage Occlusion. *Cardiovasc. Interv.* **2021**, *14*, 2353–2364. [CrossRef] [PubMed]
49. Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling Modern Machine-Learning Practice and the Classical Bias-Variance Trade-Off. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15849–15854. [CrossRef]
50. Akbilgic, O.; Davis, R.L. The Promise of Machine Learning: When Will It Be Delivered? *J. Card. Fail.* **2019**, *25*, 484–485. [CrossRef]