

Clasificación del infarto de miocardio en mujeres

Ricardo Daniel Lozano Sánchez¹, María Dolores Torres Soto¹,
Aurora Torres Soto¹, Yoselin Esparza Monreal²,
Cinthy Judith López Ramírez², Esperanza Sánchez Alemán³

¹ Benemérita Universidad Autónoma de Aguascalientes,
Departamento de Sistemas de Información,
México

² Centenario Hospital Miguel Hidalgo,
Cardiología Clínica, Aguascalientes,
México

³ Benemérita Universidad Autónoma de Aguascalientes,
Departamento de Morfología, Centro de Ciencias Básicas,
México.

al236294@edu.uaa.mx, yosmonreal1992@hotmail.com,
dracinthyaramirezchmh@gmail.com,
{mdtorres, atorres, espesanchez}@correo.uaa.mx

Resumen. El infarto de miocardio es la principal causa de muertes en el mundo, siendo los mayores índices de mortalidad en mujeres de edades avanzadas. Proponemos el uso de dos algoritmos de aprendizaje automático para entrenar un modelo de clasificación que utiliza registros clínicos de mujeres con infarto para clasificar mujeres con alto riesgo de mortalidad y mejorar su esperanza de vida. Descubrimos que los datos clínicos de las mujeres pueden permitirnos crear un modelo basado en el aprendizaje automático para predecir la mortalidad en el 90 % de las pacientes que sufrieron de un infarto. Usando Create ML de Apple, entrenamos modelos basados en bosque aleatorio y máquina de soporte vectorial utilizando 105 registros de mujeres hospitalizadas del Instituto de Cardiología de Faisalabad y el Hospital Aliado en Faisalabad (Punjab, Pakistán). Nuestros resultados muestran que bosque aleatorio supera a máquina de soporte vectorial con una precisión del 90 % en comparación con el 85 % de precisión del modelo de Support Vector Machine. Esta herramienta muestra la capacidad de este tipo de algoritmos para crear modelos que permitan a los médicos proporcionar la atención necesaria a las mujeres que están en riesgo de muerte.

Palabras clave: Infarto de miocardio en mujeres, aprendizaje automático, create ML.

Myocardial Infarction Classification in Women

Abstract. Myocardial infarction is the leading cause of death in the world, being the highest mortality rates in older women. We propose the use of two machine learning algorithms to train a classification model that uses clinical records of women with heart attack to classify women at high risk of mortality and improve

their life expectancy. We discovered that women's clinical data can allow us to create a model based on machine learning to predict mortality in 90% of patients who suffered from a heart attack. Using Apple's Create ML, we train models based on Random Forest and Support Vector Machine using 105 records of hospitalized women from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad (Punjab, Pakistan). Our results show that random forest surpasses the vector support machine with an accuracy of 90% compared to the 85% accuracy of the Support Vector Machine model. This tool shows the ability of this type of algorithm to create models that allow doctors to provide the necessary care to women who are at risk of death.

Keywords: Myocardial infarction in women, machine learning, create ML.

1. Introducción

Según la Organización Mundial de la Salud (OMS), el infarto de miocardio (IM) es una de las principales causas de muerte en todo el mundo, afectando a 9 millones de personas en el 2019, representando el 16% de todas las muertes a nivel mundial [1]. El MI afecta a casi 2.7 millones de mujeres en los Estados Unidos, con una mayor tasa de prevalencia de MI en mujeres mayores y de poblaciones minoritarias [2]. Las mujeres tienden a experimentar síntomas atípicos del IM y a menudo se les diagnostica mal o se diagnostica demasiado tarde, debido a las diferencias fisiológicas en el corazón de una mujer que pueden plantear un desafío para los médicos [2].

En México, durante la pandemia de COVID-19, la principal causa de muerte en las mujeres fue por enfermedades relacionadas con el corazón, superando las 97,000 muertes por enfermedades cardiovasculares (ECV) en comparación con las 70,000 muertes causadas por el síndrome respiratorio agudo coronavirus (SARS-CoV-2) [3], tendencia que continuó al año siguiente, de enero a julio de 2021 [4].

Históricamente, el IM en las mujeres ha sido poco tratado y poco estudiado, lo que ha llevado a un diagnóstico erróneo y a un tratamiento inadecuado [5]. Además, las mujeres tienen menos probabilidades que los hombres de recibir tratamiento debido a los retrasos en el reconocimiento del infarto agudo de miocardio (IAM) por parte de los proveedores de atención médica o los pacientes [6].

Un estudio encontró que el 18% de los pacientes con IM sin elevación del ST recibieron otro diagnóstico inicial no relacionado con el corazón, y las mujeres con IM con elevación del ST tenían un 59% más de probabilidades de un diagnóstico erróneo inicial en comparación con los hombres [7-8]. Una atención pobre y retrasada pone a las mujeres en un alto riesgo de mortalidad [9].

Estas condiciones están relacionadas con el desconocimiento de la enfermedad, razones socioculturales y financieras que pueden resultar en retrasos prehospitales y menores tasas de acceso a la atención [10]. El conjunto de datos seleccionado contiene 105 registros de mujeres que sufrieron de un IM, estos registros abarcan de 40 a 95 años. Fue publicado por Davide Chicco, Giuseppe Jurman, contiene datos de 299 pacientes del Instituto de Cardiología de Faisalabad y del Hospital Aliado de Faisalabad (Punjab, Pakistán) durante abril a diciembre del 2015.

Sus resultados mostraron que la creatinina sérica y la fracción de eyección pueden determinar la mortalidad de un paciente diagnosticado con IAM, encontrando el

modelo de bosque aleatorio (RF) [11] con una precisión del 74% en la validación de su modelo [12]. Los autores proponen que únicamente la creatinina sérica y la fracción de eyección pueden predecir la supervivencia de los pacientes [12].

Nosotros utilizamos todas las características contenidas en el conjunto de datos, esto debido a que obtuvimos una mejor puntuación en la validación del modelo. Para la creación de la herramienta clasificadora utilizamos (RF) [11] y una máquina de soporte vectorial (SVM) [13] utilizando las herramientas Create ML de Apple [14].

El objetivo de este artículo es desarrollar una herramienta precisa de clasificación del riesgo de mortalidad en las mujeres que han experimentado un IM utilizando datos clínicos reales y lograr una tasa de precisión superior al 80%. Con la finalidad contribuir con una herramienta valiosa en el campo de la atención médica, reduciendo las tasas de mortalidad hospitalaria y mejorar la esperanza de vida de los pacientes con infarto de miocardio.

Profundizamos en el desarrollo y la viabilidad de un clasificador diseñado para distinguir a las mujeres que han sufrido un IM y evaluar su riesgo de mortalidad. Este documento contiene una sección de "marco teórico", aquí discutiremos algunas enfermedades relacionadas con el corazón, la enfermedad por MI, algunos factores de riesgo y críticos en las mujeres. Después de eso, en la sección "Material y método" discutiremos en profundidad el conjunto de datos elegido, los algoritmos utilizados y cómo se llevarán a cabo los experimentos, en la sección "Resultados" discutiremos cualquier hallazgo de la fase de experimentación.

2. Marco teórico

Históricamente, las enfermedades relacionadas con el corazón a menudo se han percibido como principalmente hombres de afecto, sin embargo, un cuerpo significativo de investigación ha encontrado que el impacto de las enfermedades vasculares cerebrovasculares, como la enfermedad arterial coronaria, la insuficiencia cardíaca y el accidente cerebrovascular, son la principal causa de mortalidad en las mujeres a nivel mundial [15-17].

2.1. Infarto de miocardio

Un infarto de miocardio es una afección en la que el tejido cardíaco sufre una falta de suministro de sangre (isquemia), lo que causa su muerte (necrosis); cuando una parte del tejido cardíaco tiene un infarto, puede producir complicaciones graves para la salud de alguien, incluido el riesgo de muerte. Los síntomas comunes incluyen dolor en el pecho, dificultad para respirar, dolor en el brazo o el cuello izquierdo, entre otros. La causa de un infarto de miocardio podría deberse a varios factores, como: la acumulación de placa, un paño o la vasoconstricción de las arterias coronarias del corazón, lo que resulta en que el músculo deje de funcionar, causando la muerte del tejido [18-19]:

1. **Disparidades de Género en las ECV.** A lo largo de los años, estudios han revelado profundas disparidades de género en la manifestación, el diagnóstico y el tratamiento del IM en las mujeres, a menudo exhiben síntomas únicos que difieren del clásico dolor torácico en hombres y en su lugar, pueden presentarse con signos sutiles como dificultad para respirar, náuseas, fatiga, síntomas gastrointestinales,

de ansiedad o emocionales [10,20]. Esta divergencia de síntomas ha llevado a un mal tratamiento, diagnósticos equivocados y un retraso en la atención [5,21].

2. **Factores de Riesgo y Causas Subyacentes del IAM.** Las ECV en mujeres tienen factores de riesgo adicionales a los que son comunes en los hombres, como sufrir complicaciones durante el embarazo, fluctuaciones hormonales, estrés crónico, entre otros [20]. Comprender esos factores de riesgo emergentes es esencial para una evaluación precisa del riesgo y permitir desarrollar estrategias de prevención específicas para las mujeres [10].
3. **Cambios Hormonales.** Las hormonas, específicamente los estrógenos contribuyen significativamente a la salud de las mujeres, siendo objeto de investigación sobre la salud del corazón en las mujeres. Los estrógenos tienen efectos cardiovasculares protectores, como el mantenimiento de niveles saludables de colesterol en sangre. Siendo las mujeres que tienen niveles más bajos de estrógenos, específicamente durante la menopausia, más propensas a las ECV, incluido el IAM [2,15].

2.2. Inteligencia artificial

La IA es la simulación de la inteligencia humana en las máquinas, capaz de aprender, resolver problemas y pensar como los humanos. La IA encierra una amplia gama de tecnologías y técnicas aplicables a problemas y tareas del mundo real, dirigidas específicamente a tareas que normalmente requieren inteligencia humana, tales tareas requieren la comprensión y la capacidad del lenguaje natural, el reconocimiento de imágenes, la resolución de problemas complejos y la toma de decisiones [22]:

1. **Aprendizaje Automático.** Es una técnica utilizada por la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos que permiten a las computadoras a aprender, hacer predicciones y clasificaciones basadas en datos. Los algoritmos de aprendizaje automático utilizan técnicas estadísticas para aprender patrones de grandes conjuntos de datos. Los algoritmos de aprendizaje automático se pueden clasificar en tres tipos: Aprendizaje supervisado: el modelo se entrena con datos etiquetados. Aprendizaje no supervisado: identifica patrones y estructuras, como la agrupación. y Aprendizaje por refuerzo: aprende a tomar decisiones interactuando con el entorno, recibe retroalimentación en forma de recompensas o sanciones [23].
2. **Algoritmo de Bosque Aleatorio.** Es un algoritmo de tipo supervisado utilizado para tareas de regresión y clasificación, pertenece a los métodos de aprendizaje conjunto. Combina predicciones de múltiples algoritmos de árboles de decisión para hacer una predicción más precisa y robusta que cualquier modelo individual. También tiene una característica clave para seleccionar cada característica para los nodos de cada árbol y asegura una predicción diversa y robusta [11] Actualmente se utiliza en la salud, las finanzas, la ecología y la clasificación de imágenes [24- 26].
3. **Algoritmo de Máquina de soporte Vectorial.** Es un algoritmo de aprendizaje automático supervisado que se utiliza para tareas de clasificación y regresión. Es muy adecuado para datos donde se pueden dibujar márgenes claros de separación entre diferentes clases o grupos, encuentra el hiperplano óptimo que mejor separa

Tabla 1. Características del conjunto de datos.

Nombre de la característica	Descripción
Age	La edad del paciente
Anaemia	Valor booleano dependiendo de si el paciente tiene anemia
Creatinine Phosphokinase	Niveles de enzima
Diabetes	CPK en la sangre, mcg/L
Ejection Fraction	Valor booleano dependiendo de si el paciente tiene diabetes
High Blood Pressure	Porcentaje de sangre que sale del hogar en cada contracción
Platelets	Valor booleano si el paciente tiene presión arterial alta
Serum Creatinine	Recuento de plaquetas en la sangre, kiloplaquetas/mL
Serum Sodium	Nivel de creatinina sérica en la sangre mg/dl
Sex	Nivel de sodio sérico en la sangre mEq/L
Smoking	Valor binario si el paciente es una mujer u hombre
Time	Valor booleano si el paciente fuma
Death Event	El período de seguimiento en días

diferentes clases con el margen máximo, que es la distancia entre el hiperplano y los puntos de datos más cercanos que separa cada clase. Maximizar el margen garantiza una mejor generalización y robustez del modelo. Además, el truco del kernel permite que los datos lineales y no lineales se manejen transformando las características de entrada en un espacio dimensional más alto [13,27].

4. **Diseño factorial de experimentos.** Enfoque que utiliza la repetición de diferentes experimentos para estudiar la influencia de múltiples variables simultáneamente. Este método permite explorar los efectos individuales de varias variables en una variable dependiente. Como son conceptos básicos, tenemos factores: denotados como las variables que se están estudiando, niveles como los diferentes valores de esos factores y combinaciones [28-29].

3. Material y método

Presentamos un relato detallado de las metodologías empleadas en nuestro estudio sobre la clasificación de las mujeres adultas en riesgo de mortalidad después de tener un IAM, utilizando técnicas de aprendizaje automático y utilizando algoritmos SVM y RF.

3.1. Conjunto de datos

El conjunto de datos publicado en el artículo “La creatinina sérica y la fracción de eyección pueden predecir la supervivencia de los pacientes por insuficiencia cardíaca

contiene 105 mujeres y 194 hombres, que oscilan entre 40 y 95 años, este conjunto de datos se creó utilizando los datos recopilados en el Instituto de Cardiología de Faisalabad y en el Hospital Aliado en Faisalabad durante abril y diciembre del 2015. De los 299 pacientes, solo 96 murieron (32,10 %), específicamente, de las 105 mujeres, solo 34 murieron (32,38%), por otro lado, de los 194 hombres, 62 murieron, lo que corresponde a un 31,95 %.

Los 299 pacientes tenían disfunción sistólica del ventrículo izquierdo y tenían insuficiencia cardíaca previa [12]. Características como el sexo y el tiempo (Tabla 1) no son necesarias, nos centraremos en pacientes mujeres y esas características no proporcionan información valiosa.

3.2. Descripción del conjunto de datos

Comparamos un conjunto de características categóricas con la variable dependiente, este proceso se hace para tener una mejor comprensión del conjunto de datos. La Figura 1 no muestra que las mujeres anémicas sean especialmente más propensas a tener un resultado fatal.

Como muestra la Figura 2, las mujeres con IAM son igualmente propensas a la muerte si se les diagnostica diabetes. La Figura 3, revela que las mujeres fumadoras tienen una mayor incidencia de IAM mortal.

3.3. Algoritmos

Create ML de Apple es un conjunto de algoritmos de aprendizaje automático diseñados específicamente para desarrolladores que utilizan plataformas macOS y iPadOS.

Introducido por Apple en 2019; este marco permite a los desarrolladores construir, entrenar y probar modelos de aprendizaje automático. Se puede utilizar su aplicación fácil de usar o directamente mientras se codifica utilizando su librería de Swift [14]:

- 1. Selección de algoritmos.** El artículo “Modelo mejorado de predicción de enfermedades cardiovasculares utilizando un algoritmo bosque aleatorio”. Logrando una precisión de más del 99% al clasificar a los pacientes con ECV, superando a la SVM, Regresión Logística, K-Means, entre otros [26]. Aunque SVM no funcionó tan bien como RF; el artículo “Early Coronary Heart Disease Deciphered via SVM: Insights from Experiments” demuestra que SVM es un algoritmo adecuado para un conjunto de datos de pacientes con IAM, siendo capaz de clasificar a los pacientes con una precisión del 87.8% usando el kernel sigmoide [30]. Decidimos usar la implementación de la herramienta Create ML de SVM y de RF. Ambos algoritmos proporcionan modelos de aprendizaje eficientes, complementa la facilidad de uso y aprovecha la capacidad de la arquitectura ARM para entrenar de manera eficiente [31].

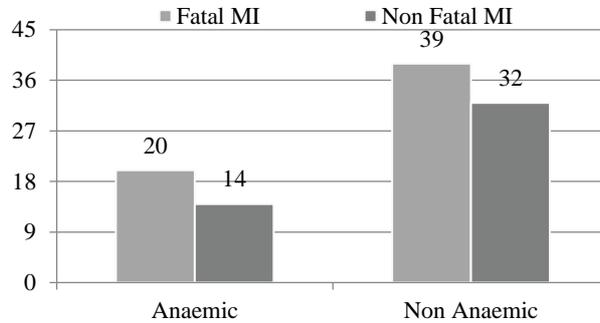


Fig. 1. Evento de muerte.

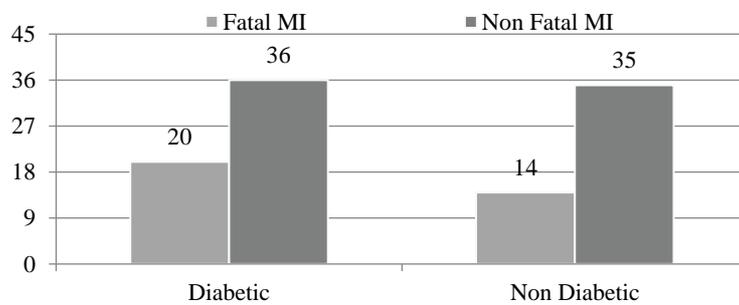


Fig. 2. Evento de muerte.

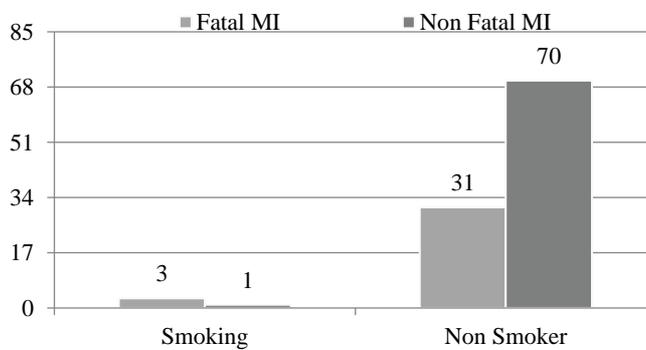


Fig. 3. Evento de muerte.

3.4. Experimentos

El conjunto de datos antes mencionado es utilizado para entrenar el modelo con los algoritmos de SVM y RF. Primero se lleva a cabo un análisis empírico para encontrar el mejor conjunto de parámetros para cada algoritmo, para enseguida, en base a los resultados anteriores, se realiza una búsqueda exhaustiva de parámetros utilizando un diseño de experimento factorial, luego se evalúa el modelo. La figura 4 representa los

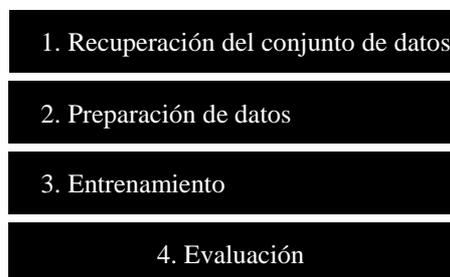


Fig. 4. Metodología.

pasos tomados para diseñar, entrenar y evaluar el modelo utilizando ambos algoritmos, SVM y RF:

1. **Recuperación del conjunto de datos.** El conjunto de datos se obtuvo durante la investigación destinada a predecir IAM utilizando una aplicación basada en el aprendizaje automático, centrada en las mujeres. Este conjunto de datos se obtuvo del documento "Machine learning can predict the survival of patients with heart failure from serum creatinine and the ejection fraction alone" de Chicco y Jurman (2020). Estos datos comprenden información clínica recopilada de pacientes en el Instituto de Cardiología de Faisalabad en Faisalabad (Punjab, Pakistán) entre abril y diciembre de 2015. De los 299 registros de este conjunto de datos, 105 pertenecen a mujeres, de los cuales 34 murieron desafortunadamente durante el período de estudio.
2. **Preparación de datos.** Nuestro enfoque se centra en los casos femeninos; para garantizarlo, hemos filtrado el conjunto de datos completo para incluir únicamente registros femeninos, creando un nuevo conjunto de datos específicamente adaptado para el propósito de este artículo. Para mantener una distribución de la función "DEATH_EVENT", asegurando una proporción aproximada de 3:10 entre las muertes y la supervivencia, hemos seleccionado cuidadosamente el conjunto de datos. Este conjunto de datos equilibrado se divide en entrenamiento y prueba, que sirven como base para pasos siguientes. Establecimos la división de entrenamiento comprenderá el 80% de este conjunto de datos filtrado, y la división de pruebas abarca el 20% restante. Con el fin de eliminar las características no correlativas, refinamos aún más este conjunto de datos, eliminamos las características de "sexo" y "tiempo", ya que el sexo no proporcionó ningún dato significativo, ya que todos los registros restantes donde la "mujer" y el período de seguimiento no brindan información sobre la mortalidad de un individuo en este contexto.
3. **Entrenamiento.** Nuestro enfoque inicial implicó una exploración empírica para identificar los parámetros óptimos para SVM [13] y RF [11]. Después de encontrar un conjunto de parámetros para cada algoritmo individualmente, empleamos el diseño factorial de experimentos [29] utilizando una amplia gama de parámetros, cada resultado de cada combinación se registró en un archivo CSV, lo cual nos permitió encontrar la configuración de parámetros más efectiva. Durante esta fase, se observó que Create ML genera automáticamente una división de validación. Aunque Apple no ha revelado detalles específicos, como los parámetros utilizados

Tabla 2. Parámetros de SVM para el diseño factorial de experimentos.

Parámetros	Valores	Valor k
Max Iterations	100 and 1,000	$k_1 = 2$
Penalty	25 to 65, steps by 0.1	$k_2 = 400$
Convergence Threshold	0.001	$k_3 = 1$
Feature Rescaling	1	$k_4 = 1$

Tabla 3. Resultados del diseño factorial de experimentos para SVM.

Parámetros	Matriz de confusión	Precisión
Max Iterations: 100 Penalty: 35.0 Convergence Threshold: 0.001 Feature Rescaling: true	True\Pred false true false 14 0 true 3 4	85.71%

en este mecanismo de división [14], observamos que sirve como retroalimentación para cada iteración durante la fase de entrenamiento, mejorando la precisión de los modelos.

4. **Evaluación.** En esta fase, evaluamos los resultados obtenidos de la frase anterior, revisando los archivos creados con la precisión de cada resultado combinado del Diseño del Experimento Factorial e implicamos rigurosas pruebas y optimización para nuestros modelos de aprendizaje automático, con esta información.

4. Resultados

El algoritmo SVM obtiene una puntuación del 85% de precisión después de evaluar el modelo utilizando el conjunto de datos de prueba, después de eso, el algoritmo RF dio la mejor puntuación de evaluación con un 90% de precisión.

4.1. Support Vector Machine

Al probar el parámetro de iteraciones máximas de SVM como parámetro de convergencia, encontramos que 100 iteraciones máximas dieron resultados prometedores, estos valores fueron suficientes para que el algoritmo proporcionara un modelo maduro con resultados prometedores, aunque, al probar algunos valores necesitaba más iteraciones; decidimos también probar con 1,000 iteraciones máximas, para el parámetro de penalización encontramos que los valores de 25 a 65 el algoritmo estaba dando resultados prometedores. El umbral de convergencia se mantuvo en un valor de 0.001 para asegurar resultados, el parámetro kernel se utilizó para todos los experimentos. Los parámetros para el diseño factorial de experimentos se muestran en la tabla 2:

Table 4. Random forest parameters for the factorial experiment design.

Parámetros	Valores	Valor k
Max Depth	4 to 6, steps of 1	$k_1 = 3$
Max Iterations	15 to 130, steps of 15	$k_2 = 8$
Min Loss Reduction	0.001 to 0.41, steps by 0.05	$k_3 = 9$
Min Child Weight	0.001 to 0.41, steps by 0.05	$k_4 = 9$
Row Subsample	80%	$k_5 = 1$
Column Subsample	80%	$k_6 = 1$
Random Seed	46	$k_7 = 1$

Table 5. Factorial experiment design results for random forest.

Parámetros	Matriz de confusión	Precisión
Max Depth: 4		
Max Iterations: 60		
Min Loss Reduction: 0.001	True\Pred False True	
Min Child Weight: 0.001	False 14 0	90.48%
Random Seed: 46	False 2 5	
Row Subsample: 0.8		
Column Subsample: 0.8		

$$30(2 \times 400) = 24,000 . \tag{1}$$

Como muestra la fórmula (1), el número de iteraciones del experimento factorial es de 24.000. Este experimento nos dio el mejor resultado, como se revela en la Tabla 3. Los resultados del modelo en la Tabla 3 no cumplen con nuestras expectativas.

4.2. Bosque aleatorio

Cuando se entrenó el algoritmo de RF con mil iteraciones máximas para converger, nos dio peores resultados, ya que era el parámetro de parada. Encontramos que el número de iteraciones que nos dieron resultados prometedores para este parámetro estaba entre 19 y 120 iteraciones máximas. Para el parámetro de profundidad máxima, los mejores resultados que obtuvimos de 4 y el algoritmo se detuvo para dar buenos resultados al nivel de profundidad 6. La reducción de la pérdida mínima y los parámetros de peso mínimo del niño dieron los mejores resultados entre 0.001 y 0.4.

Por último, los parámetros de submuestra de fila y submuestra de columna obtuvieron el mejor rendimiento a 0.8. Realizamos múltiples iteraciones para encontrar los parámetros óptimos para la convergencia. Encontramos que el uso de mil iteraciones condujo a malos resultados. Descubrimos que el punto óptimo para el número de iteraciones que proporcionaba mejores resultados, oscilaba entre 19 y 120. Del mismo modo, para determinar el parámetro de máxima profundidad, encontramos al algoritmo produjo constantemente resultados favorables dentro del rango de 4 a 6 niveles.

Específicamente, el algoritmo dejó de producir mejoras significativas más allá del sexto nivel. Además, el ajuste de los parámetros de “Min Loss Reduction” y “Min Child

Weight” resultó crítico. Los mejores resultados se lograron cuando estos valores se establecieron entre 0.001 y 0.4. Además, observamos un rendimiento óptimo al establecer los parámetros de "submuestra de fila" y "submuestra de columna" en 0.8. Estos hallazgos destacan la importancia de un ajuste meticuloso de los parámetros para aprovechar el verdadero potencial del algoritmo del bosque aleatorio. Teniendo en cuenta los resultados, los parámetros utilizados en el diseño factorial de experimentos [28-29] se muestran en la Tabla 4:

$$30(3 \times 8 \times 9 \times 9) = 58,320. \quad (2)$$

Como muestra la fórmula (2), el número de iteraciones del experimento factorial, que es de 58,320. En nuestra configuración actual, tardó varias horas en completarse, pero finalmente los resultados. Estos resultados finales mostraron en la Tabla 5 que RF superó a SVM con una mejor precisión del modelo en comparación con los resultados de la Tabla 3 de SVM, pero, como se ve en la matriz de confusión de los resultados de ambos modelos (Tabla 3 y Tabla 5), los modelos tuvieron problemas en identificar casos verdaderos, esto debido a la escasa exposición de los algoritmos a los casos verdaderos, lo que lo hace más débil en esta clase. En este caso, el algoritmo de RF superó a SVM, esto podría deberse al espacio de búsqueda estrecho para SVM, que en el caso de RF el espacio de búsqueda para los hiperparámetros dio buenos resultados sin modificar el espacio de búsqueda.

5. Conclusiones

Este modelo funciona de manera efectiva, logrando predicciones precisas en pacientes dentro de grupos de edad y características similares. Además, es crucial reconocer que la raza juega un papel importante en la influencia de la precisión del modelo, teniendo experiencia únicamente para esta población.

5.1. Trabajo futuro

La expansión del alcance de este clasificador médico implica el potencial de mejorar la esperanza de vida de las mujeres con IAM. Teniendo al alcance un grupo más grande y diverso de pacientes, que permita capturar un espectro amplio de perfiles de salud, lo cual podría permitir que el modelo reconozca patrones y correlaciones con otros factores que puedan ser significativos para la enfermedad. Es importante tomar en cuenta rango de edad, lo cual permitirá tener en cuenta los factores fisiológicos únicos en las diferentes etapas de la vida. Además, considerar una región geográfica más amplia puede introducir variables que podrían tener un impacto significativo en los resultados de salud.

Referencias

1. World Health Organization: The top 10 causes of death. <http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (2024)

2. Williams, L., McKnight, E., Gillard, C.: Myocardial infarction and older women. *U.S. Pharmacist*. <http://www.uspharmacist.com/article/myocardial-infarction-and-older-women> (2016)
3. Instituto Nacional de Estadística y Geografía: Características de las defunciones registradas en México durante 2020. Comunicado de Prensa, no. 402, vol. 21 (2021)
4. Instituto Nacional de Estadística y Geografía: Estadística de defunciones registradas de enero a junio de 2022. Comunicado de Prensa, no. 29, vol. 23 (2023)
5. Wenger, N.K.: Women and Coronary Heart Disease: A Century After. *Herrick: Understudied, Underdiagnosed, and Undertreated*. *Circulation*, vol. 126, no. 5, pp. 604–611 (2012). DOI: 10.1161/circulationaha.111.086892.
6. Stehli, J., Martin, C., Brennan, A., Dinh, D.T., Lefkovits, J., Zaman, S.: Sex Differences Persist in Time to Presentation, Revascularization, and Mortality in Myocardial Infarction Treated with Percutaneous Coronary Intervention. *Journal of the American Heart Association*, vol. 8, no. 10 (2019). DOI: 10.1161/jaha.119.012161.
7. Wu, J., Gale, C.P., Hall, M., Dondo, T.B., Metcalfe, E., Oliver, Batin, G., Hemingway, P.D., Timmis, H., West, A., Robert, M.: Editor's Choice - Impact of Initial Hospital Diagnosis on Mortality for Acute Myocardial Infarction: A National Cohort Study. *European Heart Journal: Acute Cardiovascular Care*, vol. 7, no. 2, pp. 139–148 (2016). DOI: 10.1177/2048872616661693.
8. Kwok, C.S., Bennett, S., Azam, Z., Welsh, V., Potluri, R., Loke, Y.K., Mallen, C.D.: Misdiagnosis of Acute Myocardial Infarction: A Systematic Review of the Literature. *Critical Pathways in Cardiology: A Journal of Evidence-Based Medicine*, vol. 20, no. 3, pp. 155–162 (2021). DOI: 10.1097/hpc.0000000000000256.
9. Arnstein, P.M., Buselli, E.F., Rankin, S.H.: Women and Heart Attacks: Prevention, Diagnosis, and Care. *The Nurse Practitioner*, vol. 21, no. 5, pp. 57–71 (1996). DOI: 10.1097/00006205-199605000-00005.
10. Chandrasekhar, J., Gill, A., Mehran, R.: Acute Myocardial Infarction in Young Women: Current Perspectives. *International Journal of Women's Health*, vol. 10, pp. 267–284 (2018). DOI: 10.2147/ijwh.s107371.
11. Breiman, L.: Random Forests. *Machine Learning*, vol. 45, no. 1, pp. 5–32 (2001). DOI: 10.1023/a:1010933404324.
12. Chicco, D., Jurman, G.: Machine Learning can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone. *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 16 (2020). DOI: 10.1186/s12911-020-1023-5.
13. Amari, S., Wu, S.: Improving Support Vector Machine Classifiers by Modifying Kernel Functions. *Neural Networks*, vol. 12, no. 6, pp. 783–789 (1999). DOI: 10.1016/s0893-6080(99)00032-5.
14. Apple Developer Documentation: Create ML—Create Machine Learning Models for Use in your App. <http://developer.apple.com/documentation/createml> (2023)
15. Fedorova, E.L., Bondareva, Z.G., Kuimov, A.D., Nesterenko, E.V.: Myocardial Infarction in Women: Risk Factors and Clinical Features. *Klinicheskaya Meditsina (Moskva)*, vol. 81, no. 6, pp. 28–32 (2003)
16. Arroyo-Quiroz, C., Barrientos-Gutierrez, T., O'Flaherty, M., Guzman-Castillo, M., Palacio-Mejia, L., Osorio-Saldarriaga, E., Rodriguez-Rodriguez, A.Y.: Coronary Heart Disease Mortality is Decreasing in Argentina, and Colombia, but Keeps Increasing in Mexico: A

- Time Trend Study. *BMC Public Health*, vol. 20, no. 1, pp. 162 (2020). DOI: 10.1186/s12889-020-8297-5.
17. World Health Organization: Deaths by Sex and Age Group for a Selected Country or Area and Year. WHO Mortality Database. <http://platform.who.int/mortality/themes/theme-details/MDB/all-causes> (2024)
 18. Libby, P., Theroux, P.: Pathophysiology of Coronary Artery Disease. *Circulation*, vol. 111, no. 25, pp. 3481–3488 (2005). DOI: 10.1161/circulationaha.105.537878.
 19. Institute of Medicine (US): Committee on Social Security Cardiovascular Disability Criteria: Cardiovascular Disability: Updating the Social Security Listings. National Academies Press (2010). DOI: 10.17226/12940.
 20. Wizemann, T.M., Pardue, M.-L.: Exploring the Biological Contributions to Human Health: does Sex Matter? The National Academies Collection: Reports Funded by National Institutes of Health (2001). DOI: 10.17226/10028.
 21. Vest, A.R., Cho, L.: No Woman Left Behind: Recognizing and Responding to Cardiogenic Shock in Younger Women. *Circulation: Heart Failure*, vol. 13, no. 10 (2020). DOI: 10.1161/circheartfailure.120.007782.
 22. Hunt, E.B.: Artificial Intelligence. Academic Press (1975)
 23. Zhou, Zhi-Hua: Machine Learning. Springer Singapore (2021). DOI: 10.1007/978-981-15-1967-3
 24. Cutler, D. Richard, Edwards, Thomas C., Beard, Karen H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J.: Random Forests for Classification in Ecology. *Ecology*, vol. 88, no. 11, pp. 2783–2792 (2007). DOI: 10.1890/07-0539.1.
 25. Belgiu, M., Drăguț, L.: Random Forest in Remote Sensing: A Review of Applications and Future Directions. (ISPRS) *Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31 (2016). DOI: 10.1016/j.isprsjprs.2016.01.011.
 26. Al-Manaseer, H., Abualigah, L., Alsoud, A. R., Zitar, R.A., Ezugwu, A.E., Jia, H.: A Novel Big Data Classification Technique for Healthcare Application using Support Vector Machine, Random Forest and J48. *Classification Applications with Deep Learning and Machine Learning Technologies, Studies in Computational Intelligence*, Springer, Cham, vol. 1071, pp. 205–215 (2022). DOI: 10.1007/978-3-031-17576-3_9.
 27. Suthaharan, S.: Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. *Integrated Series in Information Systems*, Springer US (2016). DOI: 10.1007/978-1-4899-7641-3.
 28. Yates, F.D.: The Design and Analysis of Factorial Experiments. Imperial Bureau of Soil Science, pp. 96 (1978)
 29. Spall, J.C.: Factorial Design for Efficient Experimentation. *IEEE Control Systems*, vol. 30, no. 5, pp. 38–53 (2010). DOI: 10.1109/mcs.2010.937677.
 30. Akhtar, F., Heyat, B.B., Parveen, S., Singh, P., Hassan, M.F.U.I, Parveen, S., Hayat, M.A.B., Sayeed, E., Ali, A., Li, J.P., Sawan, M.: Early Coronary Heart Disease Deciphered Via Support Vector Machines: Insights from Experiments. In: *Proceedings of the International Computer Conference on Wavelet Active Media Technology and Information Processing*, pp. 1–7 (2023). DOI: 10.1109/iccwamtip60502.2023.10387051.
 31. Machine Learning Research: Deploying Transformers on the Apple Neural Engine. *Computer Vision, Research Area Speech and Natural Language Processing*. <http://machinelearning.apple.com/research/neural-engine-transformers> (2022)