Analysis of cardiological data with machine learning tools

Nicolás Seivane[1] and Andrea Rey[1]

Laboratorio de Investigación y Desarrollo Experimental en Computación (LIDEC), Universidad Nacional de Hurlingham (UNAHUR), Buenos Aires, Argentina
seivanenicolas@gmail.com, andrea.rey@unahur.edu.ar

Abstract. In the current landscape, where medical practices, such as cardiology, generate an unprecedented volume of data whose complexity often exceeds the capabilities of conventional analysis, machine learning tools emerge as a resource capable of extracting phenotypic patterns hidden within clinical datasets, thereby obtaining additional information for a holistic view of the patient's cardiovascular status. In the present work, a comparative analysis of machine learning techniques applied to cardiological data is conducted. Classical algorithms were implemented and evaluated, including Naïve Bayes, Logistic Regression, Random Forest, and Support Vector Machines. The study includes predictive models for heart failure and fetal cardiotocography. The results indicate that the best overall performing method is Random Forest, achieving a high generalization capacity in both scenarios. In this context, the implementation of advanced computational models could be used to design preventive strategies based on algorithmic evidence and the analysis of historical data. It is important to highlight that machine learning does not aim to replace clinical judgment, but rather to strengthen it, functioning as a support tool for the interpretation of diagnostic tests, optimizing response times in critical cardiac intensive care environments.

Keywords: cardiology, machine learning, classification

Análisis de datos cardiológicos con herramientas de aprendizaje automático

Abstract. En el escenario actual donde prácticas médicas, como cardiología, generan un volumen de datos sin precedentes, cuya complejidad a menudo excede las capacidades del análisis convencional, las herramientas de aprendizaje automático emergen como un recurso capaz de extraer patrones fenotípicos ocultos en conjuntos de datos clínicos, obteniendo así información adicional para una visión holística del estado cardiovascular del paciente. En el presente trabajo se realiza un análisis comparativo de técnicas de aprendizaje automático aplicadas a datos cardiológicos. Se

implementaron y evaluaron algoritmos clásicos, como Naïve Bayes, Regresión Logística, *Random Forest* y Máquinas de Vectores de Soporte. El estudio incluye modelos de predicción para insuficiencia cardíaca y para cardiotocografía fetal. Los resultados indican que el método con mejor desempeño general es *Random Forest*, logrando una alta capacidad de generalización en ambos escenarios. En este contexto, la implementación de modelos computacionales avanzados podría utilizarse para diseñar estrategias preventivas basadas en la evidencia algorítmica y el análisis de datos históricos. Es importante resaltar que el aprendizaje automático no pretende sustituir el juicio clínico, sino fortalecerlo, funcionando como una herramienta de soporte para la interpretación de pruebas diagnósticas, optimizando los tiempos de respuesta en entornos críticos de cuidados intensivos cardiológicos.

Keywords: cardiología, aprendizaje automático, clasificación

1 Introducción

Las enfermedades cardiovasculares representan la principal causa de muerte a nivel mundial, cobrando aproximadamente $17,9$ millones de vidas al año, lo que equivale al 31% de las defunciones globales. El proceso de diagnóstico tradicional, aunque efectivo, puede ser extenso y depende del análisis de parámetros clínicos estructurados, señales cardíacas como el electrocardiograma (ECG) e imágenes médicas. Herramientas como el ECG, las pruebas de esfuerzo y los ecocardiogramas son el estándar clínico actual, pero requieren una interpretación obligatoria experta que puede beneficiarse de la agilidad de los modelos computacionales de Aprendizaje Automático (AA). El objetivo central de este estudio es evaluar el rendimiento de técnicas de AA al realizar predicciones sobre la posible presencia de insuficiencia cardíaca. Se exploran no solo casos de clasificación binaria (enfermo/sano), sino también el caso multiclase para el estudio de la cardiotocografía (CTG), que evalúa el bienestar fetal. Basándose en el estado del arte (Isaksen et al., 2025; Kumar & Kumar, 2021), que destaca el uso de modelos como Máquinas de Vectores de Soporte (SVM, por sus siglas en inglés) (Cortes & Vapnik, 1995) y Naïve Bayes (NB) (Hand & Yu, 2001), para la detección de arritmias y cardiopatías, este trabajo busca validar estas técnicas en un entorno comparativo con conjuntos de datos reales, bajo métricas de calidad estándar. Diversos estudios han demostrado una efectividad moderada, por lo que se requiere del escrutinio médico sobre los resultados obtenidos al aplicar algoritmos de AA y métodos de ensamble en problemas de clasificación médica. En particular, *Random Forest* (RF, por sus siglas en inglés) (Breiman et al., 2017), ha mostrado un alto desempeño debido a su capacidad para modelar relaciones no lineales y reducir el sobreajuste mediante la combinación de múltiples árboles de decisión. Asimismo, técnicas de interpretabilidad de la explicabilidad de variables (también llamadas características o atributos), como la importancia

por permutación, permiten identificar aquellas variables más influyentes, contribuyendo a la comprensión del modelo, siendo éstas las más importantes para el modelo y no en el estudio, el cual debe estar determinado y diseñado por un experto del área de salud.

2 Materiales y métodos

En este trabajo se utilizaron dos conjuntos de datos: por un lado, *HeartFailure* (fedesoriano, 2021), compuesto por 918 registros con dos posibles etiquetas, correspondientes a un problema de clasificación binaria, y por otro lado, *Cardiotocography* (Campos, 2000), con 2115 registros, incluyendo las clases 'normal', 'sospechoso' y 'patológico', representando así a un problema de clasificación multiclase. Ambos conjuntos contienen variables clínicas relevantes utilizadas para el análisis y predicción de condiciones cardiovasculares, como por ejemplo, la presión sanguínea, el nivel de colesterol en sangre, la presencia de dolor de pecho, la frecuencia cardíaca, entre otras. Para cada uno de estos conjuntos, se realizó un proceso de preparación incluyendo análisis exploratorio, verificación de la ausencia de valores faltantes, validación de la consistencia de tipos de datos y codificación de variables categóricas cuando fue necesario. Este proceso permitió asegurar la calidad de los datos antes del entrenamiento de los modelos.

Se implementaron cuatro algoritmos de clasificación: Regresión Logística (RL) (Cramer, 2002), SVM, NB Gaussiano, y RF. Cada modelo requiere de la optimización de hiperparámetros (valores que debe fijar el usuario antes de construir los modelos) que influyen en su capacidad de generalización y en el equilibrio entre sesgo y varianza, incluyendo la compensación entre un posible sobreajuste (*overfitting*) o un posible subajuste (*underfitting*). La optimización de hiperparámetros se realizó mediante búsqueda en grilla combinada con validación cruzada *k-folds*, con $k = 5$, lo que permite evaluar el modelo en múltiples particiones del conjunto de datos y obtener estimaciones sólidas de su desempeño. Para la evaluación de los clasificadores, se utilizaron métricas estándar de clasificación, incluyendo *accuracy*, que mide la proporción de registros correctamente clasificados; F1-*Score*, definido como la media armónica entre la calidad de predicción de la clase positiva y la capacidad de detección, y Área Bajo la Curva ROC (AUC por sus siglas en inglés). Además, se aplicó el método de importancia por permutación (Breiman et al., 2017), que consiste en medir la disminución del desempeño del modelo al alterar aleatoriamente cada variable, permitiendo estimar su contribución relativa.

3 Resultados y discusión

Luego de optimizar los hiperparámetros, se obtuvo la mejor configuración para SVM con kernel radial en el caso binario y polinómico de grado 2 en el caso multiclase, con coeficiente igual a $0,1$ y a 1 en los casos binario y multiclase, respectivamente, y costo de violación de restricción igual a 1 y a $0,1$ al considerar dos o tres clases, respectivamente. El modelo RF se optimizó utilizando el criterio

de la entropía para el caso binario y con el criterio de Gini en el caso multiclase, con una profundidad máxima de 7 en el caso binario y sin restricción para el caso multiclase, y 5 o 2 muestras como mínimo en cada partición para los casos binario y multiclase, respectivamente. En el caso de NB, los diferentes valores de suavizado no introdujeron modificaciones en el rendimiento del modelo. Los resultados de la Tabla 1 para el caso binario (izq.) y para el caso multiclase (der.) muestran que RF tuvo el mejor desempeño en ambos casos, alcanzando los valores más altos de *accuracy* (0, 8782 y 0, 9432), F1-*Score* (0, 8775 y 0, 9412) y AUC (0, 9315 y 0, 9868), tanto en el caso binario como en el multiclase. El modelo SVM con kernel radial también mostró un desempeño competitivo, especialmente en el caso binario, mientras que RL presentó resultados sólidos en el problema multiclase, aunque con menor capacidad para capturar relaciones no lineales complejas.

Tabla 1. Métricas de evaluación en los casos binario (izq.) y multiclase (der.).

Modelo	***Accuracy***	**F1-*Score***	**AUC**
RL	0, 8485	0, 8481	0, 9050
SVM	0, 8616	0, 8609	0, 9232
NB	0, 8420	0, 8420	0, 9138
RF	0, 8780	0, 8775	0, 9315

Modelo	***Accuracy***	**F1-*Score***	**AUC**
RL	0, 8485	0, 8481	0, 9050
SVM	0, 8616	0, 8609	0, 9232
NB	0, 8420	0, 8420	0, 9138
RF	0, 8780	0, 8775	0, 9315

El análisis de importancia por permutación permitió evaluar cuánto influye cada variable en el desempeño del modelo adoptado. Como se observa en la Fig. 1, el F1-*Score* aumenta en la medida en que se incorporan los atributos en orden decreciente de importancia, lo que indica que las variables mejor posicionadas aportan la mayor capacidad predictiva. Las primeras características generan el mayor incremento en el desempeño, mientras que las restantes contribuyen con mejoras marginales, evidenciando que una parte reducida de los atributos concentra la mayor parte de la información relevante. El análisis de importancia por permutación permitió identificar las variables más influyentes en las predicciones, donde la pendiente del segmento ST en el ECG presentó la mayor contribución en el conjunto *HeartFailure*, mientras que el porcentaje de tiempo con variabilidad anormal a corto plazo fue el atributo más relevante en la base *Cardiotocography*.

4 Conclusiones y trabajos futuros

Se concluye que el modelo RF es la herramienta más robusta para el tipo de datos bajo análisis debido a su capacidad para capturar relaciones no lineales y reducir el sobreajuste mediante el voto mayoritario de los árboles que lo integran. SVM también mostró un rendimiento competitivo, especialmente con *kernels* radiales y polinómicos. El estudio permitió identificar atributos críticos, como la pendiente del segmento ST en el ECG y el porcentaje de tiempo con variabilidad anormal a corto plazo, para el diagnóstico de riesgo cardíaco. Como

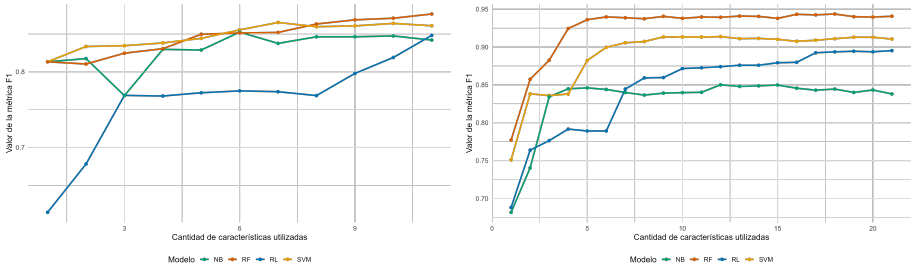

Fig. 1. Importancia de variables para los casos binario (izq.) y multiclase (der.).

trabajos futuros, se propone el uso de optimización bayesiana para un ajuste de hiperparámetros más fino y técnicas de reducción de dimensionalidad, como *Principal Component Analysis*, para simplificar los modelos sin perder precisión.

Esta línea de trabajo se basa en la hipótesis de que el uso de herramientas de Aprendizaje Automático en entornos sanitarios puede ser útil en la asistencia a los profesionales de ciencias de la salud, tanto al momento de realizar el diagnóstico, como en el diseño de estudios capaces de medir indicadores con importancia para el entrenamiento de un algoritmo, favoreciendo incluso prácticas no invasivas para los pacientes. Asimismo, se busca potenciar este soporte de diagnóstico, entendiendo al modelo no como una herramienta aislada, sino como un recurso para relacionar e integrar datos del paciente de forma más ágil.

Disclosure of Interests. Los autores declaran no tener conflictos de intereses.

Referencias

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Chapman; Hall/CRC.

Campos, J., D. y Bernardes. (2000). Cardiotocography.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Cramer, J. S. (2002). *The origins of logistic regression* (tech. rep.). Tinbergen Institute discussion paper.

fedesoriano. (2021, September). Heart failure prediction dataset. https://url-shortener.me/NFNJ

Hand, D. J., & Yu, K. (2001). Idiot's Bayes—not so stupid after all? *International statistical review*, *69*(3), 385–398.

Isaksen, J. L., Nørregaard, M., Manninger, M., et al. (2025). Evaluating artificial intelligence-enabled medical tests in cardiology: Best practice. *IJC Heart & Vasculature*, *60*, 101783.

Kumar, N., & Kumar, D. (2021). Machine learning based heart disease diagnosis using non-invasive methods: A review. *Journal of Physics: Conference Series*, *1950*(1), 012081.