

Predicción de Riesgo de Evento Cardíaco con Señales de ECG Usando la Base de Datos PTB Diagnostic ECG y PTB-XL

Guillermo Bustamante
Universidad de Talca
Curicó, Chile
gbustamante21@alumnos.utalca.cl

Abstracto— Este proyecto se enfoca en predecir el riesgo de eventos cardíacos utilizando señales de ECG de las bases de datos PTB Diagnostic ECG y PTB-XL. El objetivo es desarrollar un modelo de aprendizaje profundo para identificar pacientes en riesgo mediante características extraídas de los ECG. Se implementaron tres tipos de extracción de características: estadísticas generales, segmentos QRST y ST, y una variante con SMOTE para balancear los datos. Se utilizaron modelos como Random Forest, redes neuronales, AdaBoost y XGBoost, obteniendo una precisión de hasta 94.89% en los modelos segmentados. Aunque el modelo no cumple con estándares médicos, es útil para priorizar la atención de pacientes en riesgo.

Palabras Clave— Adaboost, Aprendizaje Profundo, ECG, Extracción de características, Machine Learning, QRST, Random Forest, Redes Neuronales, ST, STEMI, XGBoost

I. INTRODUCCIÓN

Los eventos cardíacos son una de las principales causas de muerte en todo el mundo. La detección temprana de pacientes en riesgo de sufrir eventos cardíacos es fundamental para prevenir complicaciones y mejorar la calidad de vida de los pacientes. En este sentido, el análisis de los registros de ECG (electrocardiogramas) es una herramienta útil para identificar pacientes en riesgo de sufrir eventos cardíacos.

El problema abordado es la identificación temprana de infartos al miocardio (STEMI) utilizando modelos de aprendizaje automático. Se asumió que los datos están libres de errores significativos y son representativos de la población objetivo. Para resolver el desafío, se implementaron técnicas de preprocesamiento, extracción de características y modelado predictivo, evaluando la robustez de los resultados a través de métricas como precisión, recall y F1-score.

Este trabajo demuestra cómo el uso de ciencia de datos y aprendizaje automático puede proporcionar herramientas útiles para priorizar la atención médica, incluso si no cumple con estándares clínicos.

II. MATERIALES Y METODOS

A. Datos

Los datos utilizados provienen de las bases PTB Diagnostic ECG y PTB-XL de PhysioNet. Estas bases contienen señales ECG con distintas condiciones cardíacas:

- PTB Diagnostic ECG: La base de datos contiene 549 registros de 290 sujetos de entre 17 y 87 años. La base de datos incluye diferentes clases diagnósticas con el siguiente número de sujetos: infarto de miocardio (148), miocardiopatía/insuficiencia cardíaca (18), bloqueo de rama (15), disritmia (14), hipertrofia miocárdica (7), enfermedad valvular (6), miocarditis (4), casos diversos (4) y controles sanos (52).
- PTB-XL: Comprende 21,799 registros clínicos de ECG de 12 derivaciones con una duración de 10 segundos de 18,869 pacientes, donde el 52% son hombres y el 48% son mujeres, con edades que cubren todo el rango de 0 a 95 años. El conjunto de datos contiene registros clasificados en diferentes superclases, con los siguientes detalles: 9514 registros correspondientes a ECG normales (NORM), 5469 de infarto de miocardio (MI), 5235 con cambios ST/T (STTC), 4898 con trastornos de conducción (CD) y 2649 de hipertrofia (HYP).

Se realizó un preprocesamiento exhaustivo en la base de datos PTB Diagnostic ECG que incluyó:

- Eliminación de canales innecesarios (posición de electrodos).
- Reducción de frecuencia de muestreo a 500 Hz para homogeneización.
- Filtrado de ruido con un filtro Butterworth entre 0.67 y 30 Hz.[1]

B. Método

La metodología incluyó las siguientes etapas:

- **Análisis exploratorio de datos (EDA):** Identificación de propiedades y distribución de clases en los registros. Durante la segregación de registros en la base de datos **PTB Diagnostic ECG**, se descubrió un desbalance entre las clases, lo que afectaba la representación de los pacientes sanos. Para abordar este problema, se introdujo la base de datos **PTB-XL**, con el objetivo de balancear la clase minoritaria, que en este caso era la de los pacientes sanos. Luego de realizar el balanceo, se obtuvieron 346 registros STEMI (Infarto al miocardio) y HC (Control sanos)
- **Extracción de características:** Se realizaron tres tipos de extracción de características en el análisis de los registros ECG. El primero consistió en la extracción de estadísticas generales por canal, incluyendo la media, desviación estándar, valores mínimos, máximos y la cantidad de muestras. El segundo tipo de extracción se centró en obtener características específicas de los segmentos QRST y ST de cada registro y canal. El tercer tipo de extracción fue similar al anterior, pero con la adición de **SMOTE** para balancear las características con menos datos, específicamente las de los registros sanos, que representaban la clase minoritaria.
- **Modelado:** Se entrenaron modelos como Random Forest, XGBoost y Redes Neuronales para cada tipo de extracción de características.
- **Evaluación:** Las métricas utilizadas incluyeron precisión, recall y F1-score para medir el desempeño.

III. RESULTADOS

A. Desempeño de modelos

En la extracción general de características, se observa un desempeño sobresaliente en los modelos **Redes Neuronales**, **Random Forest** y **XGBoost**:

- **Redes Neuronales** logra el mejor rendimiento global con **91.30%** en F1-Score, **91.90%** en Precisión y **91.36%** en Recall, lo que indica una alta capacidad de predicción y equilibrio entre las métricas.
- **Random Forest** y **XGBoost** obtienen resultados similares, con un F1-Score de **89.53%** y **89.00%**, respectivamente. Ambos presentan buena precisión y recall, con ligeras ventajas en precisión para XGBoost.
- **AdaBoost** queda por debajo con un F1-Score de **85.79%**, mostrando un desempeño moderado.
- **SVM** es el modelo con el peor rendimiento, con un F1-Score de **65.33%**, lo que sugiere dificultades para generalizar sobre los datos.

En este método de extracción general de características, las Redes Neuronales lideran el desempeño, seguidas de Random Forest y XGBoost, mientras que SVM presenta limitaciones claras.

TABLE I. RESULTADOS EXTRACCIÓN GENERAL

Modelo	Métricas		
	F1-Score (%)	Precisión (%)	Recall (%)
Random Forest	89.53	90.33	89.45
SVM	65.33	66.79	65.05
XGBoost	89.00	90.62	88.91
AdaBoost	85.79	86.89	85.67
Redes Neuronales	91.30	91.90	91.36

En la extracción de segmentos QRST, los modelos mantienen un rendimiento superior, con diferencias notables en las métricas entre clases (0 y 1):

- **XGBoost** sobresale con un F1-Score de **93.39%**, logrando una precisión notable del **99.79%** para la clase 1 (STEMI). Sin embargo, presenta limitaciones para la clase 0 (controles sanos) con un F1-Score bajo.
- **Redes Neuronales** obtiene un F1-Score general de **92.20%**, con buen equilibrio entre precisión y recall, mostrando un desempeño más consistente entre clases comparado con otros modelos.
- **Random Forest** y **AdaBoost** también alcanzan métricas altas (**91.70%** y **93.02%** en F1-Score), aunque presentan dificultades con la clase 0 (valores bajos de precisión).

En este método de extracción específica de características de los segmentos QRST, **XGBoost** y **Redes Neuronales** destacan, con un rendimiento excepcional para la clase 1, aunque persiste un desbalance en la clasificación de controles sanos (clase 0).

TABLE II. RESULTADOS EXTRACCIÓN QRST

Modelo	Métricas				
	F1-Score (%)	Precisión (%)		Recall (%)	
	Weighted avg	0	1	0	1
Random Forest	91.70	23.48	97.01	43.84	92.72
XGBoost	93.39	59.97	95.443	6.29	99.79
AdaBoost	93.02	43.83	95.27	2.78	99.82
Redes Neuronales	92.20	79.84	94.09	25.02	99.47

En la extracción de segmentos **ST**, los modelos mejoran aún más en general:

- **Redes Neuronales** lidera nuevamente con un F1-Score de **94.89%**, alcanzando una precisión alta del **99.59%** para la clase 1 y un rendimiento equilibrado entre las métricas.
- **XGBoost** sigue de cerca con un F1-Score de **92.23%**, con un alto recall del **99.50%** en la clase 1.

- **Random Forest** y **AdaBoost** mantienen buenos resultados, aunque con menor rendimiento en comparación con XGBoost y Redes Neuronales.

La extracción ST mejora la capacidad predictiva, con Redes Neuronales como el modelo dominante y XGBoost mostrando un desempeño cercano.

TABLE III. RESULTADOS EXTRACCIÓN ST

Modelo	Métricas				
	<i>F1-Score (%)</i>	<i>Precisión (%)</i>		<i>Recall (%)</i>	
	<i>Weighted avg</i>	<i>0</i>	<i>1</i>	<i>0</i>	<i>1</i>
Random Forest	90.36	29.78	96.97	57.90	90.81
XGBoost	92.23	64.70	94.48	13.62	99.50
AdaBoost	91.40	65.59	94.03	5.88	99.79
Redes Neuronales	94.89	89.72	95.80	44.83	99.59

Al aplicar **SMOTE** en la extracción QRST, se observan reducciones significativas en las métricas globales, debido al equilibrio artificial introducido en los datos:

- **Random Forest** y **XGBoost** muestran un descenso en F1-Score a **80.80%** y **78.21%**, respectivamente, con un recall alto pero una precisión muy baja para la clase 0 (controles sanos).
- **Redes Neuronales** se mantiene más estable con un F1-Score de **78.52%**, mostrando un equilibrio moderado entre clases.
- **AdaBoost** registra el rendimiento más bajo con un F1-Score de **75.57%**.

La aplicación de **SMOTE** mejora la cobertura de la clase minoritaria (0), pero afecta negativamente la precisión global, especialmente en modelos como Random Forest y XGBoost.

TABLE IV. RESULTADOS EXTRACCIÓN QRST SMOTE

Modelo	Métricas				
	<i>F1-Score (%)</i>	<i>Precisión (%)</i>		<i>Recall (%)</i>	
	<i>Weighted avg</i>	<i>0</i>	<i>1</i>	<i>0</i>	<i>1</i>
Random Forest	80.80	12.26	98.20	73.62	73.16
XGBoost	78.21	11.29	98.34	77.13	69.13
AdaBoost	75.57	9.95	98.08	74.90	64.47
Redes Neuronales	78.52	78.64	78.39	78.35	78.68

La extracción ST con **SMOTE** muestra un comportamiento similar al caso QRST con SMOTE:

- **Redes Neuronales** lidera con un F1-Score de **82.94%**, mostrando la mejor estabilidad entre precisión (**83.92%**) y recall (**81.53%**) para ambas clases.

- **Random Forest** y **XGBoost** logran un rendimiento moderado, con F1-Scores de **76.90%** y **75.92%**, respectivamente. Ambas muestran un recall alto para la clase 1, pero con precisión muy baja para la clase 0.
- **AdaBoost** presenta el rendimiento más bajo con un F1-Score de **74.94%**.

En la extracción ST con SMOTE, Redes Neuronales sobresalen como el modelo más equilibrado, mientras que Random Forest y XGBoost sufren pérdidas en precisión.

TABLE V. RESULTADOS EXTRACCIÓN ST SMOTE

Modelo	Métricas				
	<i>F1-Score (%)</i>	<i>Precisión (%)</i>		<i>Recall (%)</i>	
	<i>Weighted avg</i>	<i>0</i>	<i>1</i>	<i>0</i>	<i>1</i>
Random Forest	76.90	14.52	98.10	80.44	68.14
XGBoost	75.92	12.98	97.33	72.59	67.26
AdaBoost	74.94	11.92	96.85	68.08	66.16
Redes Neuronales	82.94	83.92	82.02	81.53	84.36

IV. DISCUSIÓN

Los resultados obtenidos validan la robustez de los modelos propuestos en la identificación de riesgos cardíacos:

- **Interpretación:** Los modelos evaluados, especialmente **Redes Neuronales**, **Random Forest** y **XGBoost**, demostraron una **alta capacidad predictiva**, alcanzando precisiones de hasta **94.89%** en la extracción de características específicas (QRST y ST). Esto confirma que los modelos son efectivos para distinguir entre pacientes con **infarto al miocardio (STEMI)** y **pacientes sanos**, cumpliendo el objetivo central del proyecto.

La implementación de tres métodos de extracción de características (general, segmentos QRST y ST) permitió analizar cómo distintas representaciones de los datos impactan el rendimiento de los modelos.

Los segmentos QRST y ST, más específicos, lograron mejores resultados que la extracción general, mostrando que la segmentación aporta más información relevante para la clasificación.

El uso de **SMOTE** ayudó a equilibrar la clase minoritaria (pacientes sanos), abordando uno de los desafíos iniciales del proyecto. Sin embargo, esto afectó la precisión global, especialmente en modelos como **Random Forest** y **XGBoost**.

A pesar de esta limitación, **Redes Neuronales** mantuvieron un rendimiento más equilibrado, lo que valida su capacidad de adaptación frente a datos balanceados artificialmente.

- **Limitaciones:** Una de las limitaciones del proyecto fue la decisión deliberada de trabajar con la base de datos **PTB Diagnostic ECG** a pesar de su desbalance de clases. Esta restricción fue impuesta con el objetivo de **conocer y aprender más sobre el manejo de bases de datos desbalanceadas**, lo que permitió explorar técnicas como la incorporación de registros adicionales de **PTB-XL** y el uso de **SMOTE**. Si bien esto añadió un desafío adicional al desarrollo, también ofreció una oportunidad valiosa para comprender cómo este tipo de problemas impactan en el rendimiento de los modelos y cómo pueden mitigarse.

- **Aplicaciones:** Los resultados obtenidos en este proyecto tienen aplicaciones prácticas significativas, especialmente en entornos clínicos y de investigación donde el diagnóstico temprano de eventos cardíacos es crucial. Aunque el modelo no cumple con estándares médicos, su utilidad radica en los siguientes aspectos:

1) Herramienta de priorización de pacientes: El modelo puede integrarse en sistemas de monitoreo hospitalario para analizar señales de ECG de manera automática y rápida. Esto permitiría priorizar la atención de pacientes con riesgo potencial, optimizando recursos en entornos con limitaciones como emergencias o áreas rurales con escaso personal especializado.

2) Soporte a profesionales de la salud: Proporciona una herramienta complementaria para cardiólogos y personal médico, facilitando la detección preliminar de anomalías cardíacas como infartos al miocardio (STEMI) y otros trastornos. Esto no reemplaza la evaluación médica, pero ayuda a reducir tiempos de diagnóstico y proporciona una segunda opinión basada en análisis automatizado.

3) Entrenamiento y educación: Los métodos utilizados en el proyecto, incluyendo el manejo de datos desbalanceados y la aplicación de técnicas como SMOTE, pueden servir como base para la formación de estudiantes e investigadores en áreas de ciencia de datos aplicada a la medicina. Además, el uso de señales ECG reales fomenta un aprendizaje más práctico y relevante.

4) Investigación en inteligencia artificial aplicada a la salud: Este trabajo proporciona un marco metodológico reproducible que puede extenderse a otros problemas relacionados con la detección de enfermedades cardíacas y el análisis de señales fisiológicas. Es un punto de partida para desarrollar modelos más robustos y cumplir con estándares clínicos en el futuro.

5) Implementaciones en dispositivos portátiles: Los avances en modelos de aprendizaje automático permiten que este tipo de soluciones puedan implementarse en dispositivos portátiles, como monitores cardíacos o wearables (smartwatches), para el monitoreo continuo de pacientes con factores de riesgo.

V. CONCLUSIONES

El desarrollo del proyecto permitió obtener resultados prometedores en la predicción del riesgo de eventos cardíacos utilizando registros de ECG y técnicas de aprendizaje profundo. Los modelos de clasificación desarrollados lograron identificar pacientes en riesgo de sufrir eventos cardíacos con una precisión superior al 90% en algunos casos. Pero debido a la naturaleza del proyecto, no se recomienda su uso como un sistema autónomo clasificador de pacientes con riesgo de eventos cardíacos, sino como una herramienta de apoyo para los profesionales de la salud para priorizar pacientes en su atención.

Esto debido a la existencia de falsos negativos, es decir, casos en los que el modelo clasifica erróneamente a pacientes enfermos como si estuvieran sanos. Esta situación puede resultar particularmente peligrosa en un contexto clínico, ya que un diagnóstico incorrecto puede retrasar la atención médica oportuna y el tratamiento necesario, aumentando así el riesgo de complicaciones graves o incluso la mortalidad. Los falsos negativos podrían generar una falsa sensación de seguridad tanto en el paciente como en el profesional de la salud, lo que podría tener consecuencias críticas en la toma de decisiones. Por lo tanto, es fundamental que los resultados de los modelos sean interpretados por especialistas, quienes pueden combinar estos datos con otras evaluaciones clínicas para realizar diagnósticos más precisos y garantizar un manejo adecuado de los pacientes.

REFERENCIAS

- [1] M. Bodini, M. W. Rivolta, y R. Sassi, "Interpretability Analysis of Machine Learning Algorithms in the Detection of ST-Elevation Myocardial Infarction," *Computing in Cardiology*, vol. 47, pp. 1–2, September 2020.