Mejora de la clasificación de latidos cardíacos en señales ECG mediante preentrenamiento autosupervisado a gran escala

Sara Artal [1], Julia Ramírez [1,2,3], Antonio Miguel [1], Juan Pablo Martínez [1,2]

[1] Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, Zaragoza, España, sartal@unizar.es

[2] Centro de Investigación Biomédica en Red, Biomateriales, Bioingeniería y Nanomedicina, España

[3] William Harvey Research Institute, Queen Mary University of London, London, United Kingdom

Resumen

Los latidos ectópicos, anomalías del ritmo cardíaco, son indicadores importantes de arritmias. El electrocardiograma (ECG) de 12 derivaciones es una herramienta económica y no invasiva, ideal para el diagnóstico de estas patologías. Recientemente, las redes neuronales (RN) se han propuesto en cardiología para la predicción del riesgo cardíaco. Sin embargo, su uso sigue siendo limitado debido a la escasez de datos etiquetados. El objetivo de este proyecto es evaluar si el preentrenamiento autosupervisado mejora el rendimiento de las RN en una tarea de aprendizaje supervisado para la clasificación de tres tipos de latidos (normales, ectópicos supraventriculares aislados, o combinados con ventriculares), usando datos de entrenamiento limitados con una distribución desequilibrada de casos. Para ello, se han descargado y preprocesado señales ECG disponibles públicamente, dando lugar a 14.662.014 señales ECG. A continuación, se ha implementado una RN convolucional multicapa (CNN) para la clasificación de latidos. Después se ha desarrollado y preentrenado, de forma autosupervisada, una RN basada en codificación predictiva contrastiva. Los dos modelos han mostrado un buen rendimiento. Al utilizar una CNN simplificada junto con una RN preentrenada de forma autosupervisada, la precisión balanceada mejoró de 0,701 a 0,832, la puntuación F1 balanceada ascendió de 0,897 a 0,902, y la macro área bajo la curva aumentó de 0,835 a 0,966. De acuerdo a nuestros hallazgos, el método de aprendizaje supervisado basado en características autosupervisadas mejora el rendimiento y permite simplificar la arquitectura y el entrenamiento del modelo puramente supervisado.

1. Introducción

Las arritmias cardíacas son alteraciones del ritmo del corazón causadas por anomalías en la generación o propagación del impulso eléctrico cardíaco. Estas alteraciones pueden manifestarse como latidos ectópicos, que se generan en zonas del corazón diferentes al nodo sinoauricular. Según su origen, los latidos ectópicos se clasifican en supraventriculares, si se generan en el nodo auriculoventricular o en las aurículas, y en ventriculares, si se originan en los ventrículos. También existen los latidos de fusión, resultado de la coincidencia casi simultánea de estos impulsos de diferente origen. En este contexto, el electrocardiograma (ECG) es una herramienta diagnóstica estándar, no invasiva y altamente efectiva para el análisis de arritmias, ya que muchas de ellas se reflejan en secuencias de latidos con intervalos irregulares o morfologías atípicas, lo que hace que la clasificación automática de latidos sea un paso crucial para su identificación.

En los últimos años, las redes neuronales (RN) han abierto nuevas oportunidades para el análisis de la señal ECG, identificando automáticamente los aspectos morfológicos más relevantes que contribuyen al riesgo cardíaco [1]. Sin embargo, la necesidad de acceder a grandes volúmenes de datos etiquetados, que habitualmente son difíciles y costosos de obtener, especialmente en el ámbito médico, así como la sensibilidad de estos modelos al desequilibrio de los datos, limitan su rendimiento [2, 3].

En otras aplicaciones, como el procesado del lenguaje natural, se ha demostrado que el preentrenamiento de una RN con datos no etiquetados, por ejemplo, mediante codificación predictiva contrastiva (CPC) [4], mejora el rendimiento de dicha RN cuando se cuenta con datos etiquetados limitados. La idea principal detrás de CPC es aprender representaciones que codifiquen la información subyacente compartida entre distintas partes de la señal, descartando al mismo tiempo la información local y de bajo nivel considerada como ruido.

En este trabajo, planteamos la hipótesis de que el rendimiento de una RN, entrenada con una cantidad reducida de datos etiquetados (aprendizaje supervisado) para clasificar tres tipos de latidos cardíacos (normales, ectópicos supraventriculares y ectópicos ventriculares más latidos de fusión), puede mejorar notablemente partiendo de una RN preentrenada para resolver tareas más simples con señales ECG no etiquetadas (aprendizaje autosupervisado). A esta estrategia, que combina modelos autosupervisados con modelos supervisados, la denominaremos aprendizaje supervisado basado en características autosupervisadas.

2. Materiales y métodos

2.1. Materiales

En este proyecto, se han utilizado diferentes bases de datos disponibles en los repositorios de libre acceso PhysioNet y Zenodo. La Tabla 1 detalla sus principales características.

2.2. Preprocesado de las señales ECG

Todas las señales se remuestrearon, cuando fue necesario, a 500 Hz mediante un filtro polifásico en cuadratura. Para eliminar el ruido de línea base, se utilizó un filtro IIR paso alto

ISBN: 978-84-09-80259-3

Base de datos	Pacientes	Registros	Derivaciones	Fs (Hz)	Duración
ICENTIA-11K	11.000	542.157	1	250	70 min
ECG-ARRHYTHMIA	45.152	45.152	12	500	10 s
PTB-XL	18.869	21.799	12	500	10 s
GEORGIA	15.742	10.344	12	500	10 o 5 s
CHALLENGE-17	8.528	8.528	1	300	10 a 60 s
CPSC	9.458	6.877	12	500	6 a 60 s
CPSC EXTRA	9.458	3.453	12	500	6 a 60 s
MIMIC-IV ECG	161.352	800.035	12	500	10 s
CODE-15 %	233.770	345.779	12	400	10.24 s
MITDB	47	48	2	360	30 min
SVDB	78	78	2	128	30 min

Tabla 1. *Bases de datos utilizadas en el proyecto. Nota: Fs=frecuencia de muestreo.*

Butterworth de orden 4 y frecuencia de corte 0,3 Hz.

2.3. Modelo de aprendizaje supervisado

El objetivo de la tarea de aprendizaje supervisado, utilizada para verificar la hipótesis planteada, es clasificar tres tipos de latidos cardíacos: normales, ectópicos supraventriculares y ectópicos ventriculares más latidos de fusión. La clasificación se realiza a partir de segmentos ECG de 4 segundos, donde el latido de interés se encuentra en el segundo 3 del fragmento.

Para lograrlo, se ha utilizado una RN convolucional que consta de dos componentes principales, un extractor de características y un clasificador. Como se ilustra en la Figura 1, la extracción de características se realiza mediante 6 capas convolucionales 1D, cada una con 128 unidades ocultas, un tamaño de kernel de 3 y un *stride* (desplazamiento) de 2. Estas capas operan en una dimensión de la señal de entrada, la dimensión temporal del ECG, lo que implica un diezmado global de la señal por un factor 64. A continuación, se incorporaron 2 capas LSTM bidireccionales [5], que permiten capturar y retener información relevante del contexto temporal de la señal. Además, se aplica normalización por lotes [6] entre las capas convolucionales para mejorar la velocidad de convergencia y la estabilidad de la red. La función de activación utilizada, con el fin de modelar las relaciones no lineales presentes en los datos, es ReLU. Finalmente, para la clasificación se emplea una única capa completamente conectada.

Para el entrenamiento, validación y test del modelo se emplearon las bases de datos MITDB y SVDB, siguiendo la división propuesta por Llamedo et al. [7] para facilitar la comparación con un trabajo del estado del arte. El conjunto de entrenamiento incluye 22 registros de MITDB y el de test otros 22, sin solapamiento de pacientes entre ambos, lo que evita sesgos y asegura que el modelo aprenda patrones generales de los latidos. Para validación se empleó la base de datos SVDB. La Tabla 2 muestra el número de ejemplos por clase para las tres cohortes descritas.

	N	S	V'
Entrenamiento	45.742	941	4.192
Validación	161.977	12.180	9.947
Test	44.158	1.835	3.603

Tabla 2. Distribución de ejemplos por clase para entrenamiento (MITDB - partición 1), validación (SVDB) y test (MITDB - partición 2). Nota: N=latidos normales, S=latidos ectópicos supraventriculares y V'=latidos ectópicos ventriculares más latidos de fusión.

El modelo se entrenó durante 200 iteraciones, cada una con 2.700 ejemplos. Para mitigar el desbalance de clases, en cada iteración se utilizó un conjunto equilibrado con 900 ejemplos por clase y se aplicó aumento de datos mediante la adición de ruido blanco gaussiano. La tasa de aprendizaje inicial fue de 0,0001, reduciéndose un 5 % cada 50 iteraciones. El entrenamiento se realizó con lotes de 64 ejemplos y fragmentos de ECG de 4 segundos (2.000 muestras a 500 Hz).

Para evaluar el rendimiento del modelo en la fase de test, se calcularon la precisión balanceada, la puntuación F1 balanceada y la macro área bajo la curva (AUC). Para cada clase se obtuvo su sensibilidad (Se) y valor predictivo positivo (P^+).

2.4. Modelo de aprendizaje autosupervisado

La arquitectura del modelo CPC desarrollado, mostrada en la Figura 1, puede dividirse en dos bloques. En primer lugar, consta de un codificador no lineal que extrae las características de las señales ECG de entrada. Este mapea la secuencia de muestras crudas de entrada a una secuencia de representaciones latentes (z) [4]. En nuestro caso, está compuesto por 6 capas convolucionales 1D, todas con 1.024 unidades ocultas, un tamaño de kernel de 3 y un *stride* de 2. Como consecuencia, la resolución temporal de las representaciones latentes, z, será potencialmente menor. Nuevamente, se emplearon la normalización por lotes y la función de activación ReLU entre las capas convolucionales.

La salida del codificador (z) constituye la entrada del segundo bloque. Este es un bloque autorregresivo (AR), que está compuesto por una capa GRU [8] que almacena información relevante de varios instantes de tiempo pasados mediante una variable de estado, con el fin de generar predicciones de cómo evolucionará la señal en muestras futuras. A la salida de la capa GRU se obtiene la información de contexto (c), denominada representación latente contextual. Esta se utilizará para realizar predicciones de k pasos futuros. En nuestro caso, se obtiene una predicción distinta para cada uno de ellos, siendo k igual a 8. La capa de predicción está formada por 2 capas convolucionales 1D con un tamaño de kernel igual a 1, y una función de activación ReLU entre ellas. De esta manera, se aplica una transformación no lineal a la información contextual.

Para el preentrenamiento del modelo autosupervisado se utilizaron todas las bases de datos de la Tabla 1, excepto las empleadas por el modelo supervisado (MITDB y SVDB). En total, se dispuso de 14.662.014 señales ECG, de las cuales 14.661.045 se asignaron al conjunto de entrenamiento y 969 al conjunto de test. Además, el modelo fue entrenado durante 4.000 iteraciones con 50.000 muestras por iteración, así como con una tasa de aprendizaje con un valor inicial de 0,0001, que se redujo cada 40 iteraciones en un 5 %. Los datos se procesaron con un tamaño de batch real de 256 unidades. Sin embargo, mediante acumulación de gradientes cada 4 pasos se simuló un batch efectivo de 1.024 unidades. Las entradas a la RN fueron segmentos de ECG de longitud variable entre 3 y 7 segundos, lo que equivale a 1.500 y 3.500 muestras, respectivamente, para una frecuencia de muestreo de 500 Hz.

ISBN: 978-84-09-80259-3

2.5. Modelo de aprendizaje supervisado basado en características autosupervisadas

El objetivo del modelo de aprendizaje supervisado basado en características autosupervisadas es evaluar el valor añadido de las representaciones aprendidas por este último en la tarea de clasificación de tres tipos de latidos. Para ello, se utilizaron las representaciones internas del modelo autosupervisado, z (representación latente), c (representación latente contextual) o la concatenación de ambas, como entrada para el modelo supervisado. En consecuencia, se modificó el número de canales de la primera capa convolucional del modelo supervisado para que coincidiera con la dimensión de la representación de salida del modelo autosupervisado (1.024 en el caso de z o c, y 2.048 en el caso de la concatenación de z y c). La Figura 1 muestra un diagrama detallando la arquitectura y las dimensiones de la combinación de ambos modelos.

Además, se realizaron múltiples experimentos en los que se modificó el número de capas convolucionales en la sección de extracción de características del modelo supervisado. Partiendo del modelo supervisado original, que contaba con 6 capas convolucionales, se implementaron nuevas versiones en las que el número de capas varió desde 6 hasta ninguna. En todos los casos, después de las capas convolucionales se mantuvieron las capas LSTM, excepto en el último, donde se dejó únicamente la capa de clasificación lineal. Durante el entrenamiento, el modelo autosupervisado se mantiene congelado para evitar su actualización.

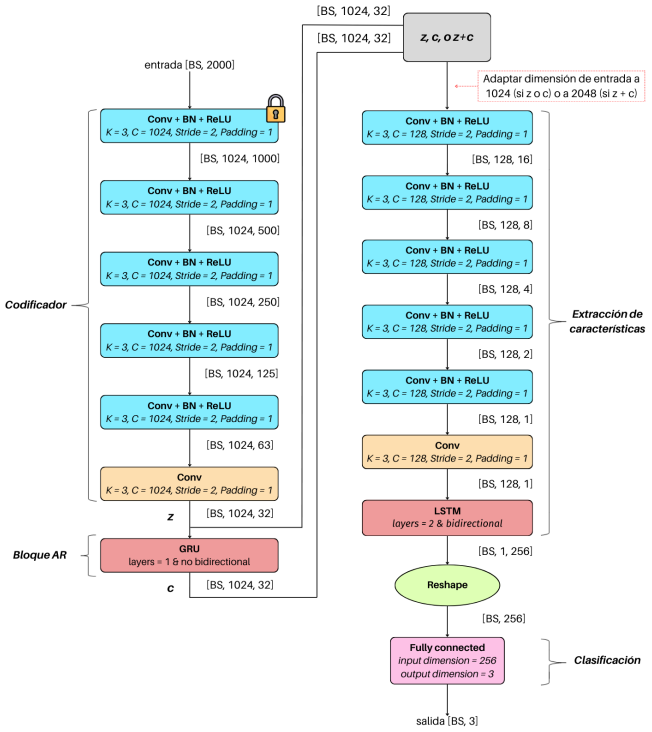


Figura 1. *Arquitectura del modelo de aprendizaje supervisado basado en características autosupervisadas. A la izquierda, la arquitectura del modelo autosupervisado y, a la derecha, la del modelo supervisado con 6 capas convolucionales 1D para la extracción de características.*

3. Resultados

El modelo supervisado implementado para la clasificación de tres tipos de latidos cardíacos (normales, ectópicos supraventriculares y ectópicos ventriculares más latidos de fusión) obtuvo en test una precisión balanceada de 0,701, una puntuación F1 balanceada de 0,897 y una macro AUC de 0,835 (Tabla 3, métricas obtenidas utilizando las muestras de ECG crudas como entrada para el modelo supervisado de 6 capas convolucionales). No obstante, el desbalance de clases afecta al rendimiento global, ya que el modelo tiende a mostrar una menor Se y P^+ en las clases menos representadas (Tabla 4).

Los resultados de los experimentos realizados para evaluar los modelos supervisados basados en características autosupervisadas también se muestran en la Tabla 3. De todas las combinaciones estudiadas, la que obtuvo los mejores resultados fue la unión del modelo supervisado original y el modelo autosupervisado, pasando la variable z (representación latente) como entrada al modelo supervisado, a la vez que se redujo el número de capas convolucionales en la parte de extracción de características supervisada de 6 a 3. De esta forma, los resultados mejoraron considerablemente con respecto a los obtenidos por el modelo puramente supervisado. Siguiendo este nuevo enfoque, se obtuvo una precisión balanceada de 0,832, una puntuación F1 balanceada de 0,902 y una macro AUC de 0,966.

Al evaluar la Se y el P^+ por clase (Tabla 4), se observa una mejora notable en la identificación de ectópicos supraventriculares (S), la clase minoritaria, donde la Se aumenta de 0,245 a 0,641 y el P^+ de 0,294 a 0,553. Asimismo, los resultados en términos de precisión balanceada son comparables con los reportados por Llamedo et al. [7]. Cabe destacar que dicho trabajo emplea características derivadas de las señales ECG, como el intervalo RR, mientras que en nuestro enfoque las redes operan directamente sobre las señales ECG.

4. Discusión y conclusiones

En este trabajo, analizamos la mejora del rendimiento obtenida mediante un modelo de codificación predictiva, preentrenado de manera autosupervisada, en una tarea de clasificación multiclase supervisada. Nuestro principal hallazgo es que dicho modelo es capaz de proporcionar mejoras notables con respecto a la tarea supervisada original, en un entorno donde los datos disponibles son limitados y desequilibrados.

La implementación de esta técnica supone una mejora metodológica significativa. La calidad de los resultados obtenidos ha aumentado, alcanzando mayor precisión y consistencia mediante el uso de una arquitectura supervisada más simple. En el enfoque supervisado original se empleaban 6 capas convolucionales, mientras que, al incorporar representaciones autosupervisadas, logramos superar esos resultados incluso utilizando una sola capa conectada en el modelo supervisado (tanto con z como con c). Además, al simplificar la arquitectura supervisada se reducen los requisitos de memoria y tiempo de entrenamiento, así como el riesgo de sobreajuste. Este ahorro no se extiende a test, donde ambos modelos, autosupervisado y supervisado, deben evaluarse.

ISBN: 978-84-09-80259-3

Métrica	Señales de entrada	Número de capas convolucionales en el modelo supervisado						
		6	5	4	3	2	1	0
Precisión balanceada	x (muestras)	0,701	0,684	0,688	0,695	0,648	0,650	0,518
	z	0,652	0,677	0,691	**0,832**	0,657	0,702	**0,724**
	c	0,705	0,721	0,705	0,692	0,763	0,767	**0,785**
	z + c	0,717	0,744	0,731	0,723	0,757	0,755	0,713
Puntuación F1 balanceada	x (muestras)	0,897	0,903	0,879	0,896	0,894	0,879	0,776
	z	0,907	0,915	0,885	**0,902**	0,908	0,921	**0,902**
	c	0,890	0,877	0,898	0,864	0,908	0,899	**0,909**
	z + c	0,885	0,941	0,908	0,908	0,895	0,924	0,876
Macro AUC	x (muestras)	0,835	0,846	0,870	0,926	0,896	0,899	0,739
	z	0,890	0,906	0,934	**0,966**	0,863	0,944	**0,952**
	c	0,935	0,914	0,857	0,929	0,959	0,939	**0,962**
	z + c	0,884	0,894	0,889	0,929	0,956	0,953	0,959

Tabla 3. *Clasificación de tres tipos de latidos cardíacos midiendo la precisión balanceada, la puntuación F1 balanceada y la macro área bajo la curva (AUC) para diferentes arquitecturas de modelos supervisados y modelos supervisados basados en características autosupervisadas.*

Experimento	Test								
	N		S		V'		Precisión balanceada	Puntuación F1 balanceada	Macro AUC
	Se	P^+	Se	P^+	Se	P^+			
Puramente supervisado	0,916	0,964	0,245	0,294	0,942	0,555	0,701	0,897	0,835
Supervisado basado en características autosupervisadas	0,895	0,983	0,641	0,553	0,961	0,474	0,832	0,902	0,966
Llamedo et al.	0,950	0,980	0,770	0,390	0,810	0,870	0,840	-	-

Tabla 4. *Sensibilidad (Se) y valor predictivo positivo (P^+) para los diferentes enfoques considerados. Nota: N=latidos normales, S=latidos ectópicos supraventriculares y V'=latidos ectópicos ventriculares más latidos de fusión.*

Sin embargo, nuestro enfoque no logra superar los resultados de Llamedo et al. [7], que emplean características morfológicas e intervalos RR. Esto podría deberse a que dichas variables encapsulan información fisiológica crítica que nuestro modelo debe aprender implícitamente a partir de la señal ECG cruda, un proceso más complejo que puede requerir más datos y entrenamiento. No obstante, se han obtenido resultados muy cercanos.

Como futura ampliación, resultaría útil incorporar técnicas de visualización para analizar las representaciones internas aprendidas por el modelo CPC, con el fin de estudiar los patrones extraídos de los ECG y su relación con características fisiológicas relevantes, aumentando así la confiabilidad del modelo y favoreciendo su integración clínica. Además, podría explorarse el uso de señales ECG más largas durante el preentrenamiento autosupervisado, para proporcionar al modelo una representación más completa de la señal y la posibilidad de capturar información vinculada al intervalo RR.

El método propuesto abre la puerta al uso de RN en problemas con disponibilidad limitada de señales ECG, como ocurre en las enfermedades cardiovasculares de baja prevalencia.

Agradecimientos

Este trabajo ha contado con el apoyo de los proyectos PID2021-128972OA-100, TED2021-130459B-I00, PID2022-140556OB-I00, PID2023-148975OB-I00 y CNS2023-143599 financiados por el Ministerio de Ciencia e Innovación (MCIN) y FEDER, y del Grupo T39-23R de Interpretación de Señales Biomédicas y Simulación Computacional (BSICoS). SA agradece la beca del Instituto de Investigación en Ingeniería de Aragón (I3A) para la realización del Trabajo Fin de Máster. JR agradece la financiación de la beca RYC2021-031413-I de MCIN. Los cálculos se realizaron utilizando ICTS NANBIOSIS.

Referencias

[1] A. Kashou, A. May, and P. Noseworthy, "Artificial intelligence-enabled ecg: a modern lens on an old technology," *Current cardiology reports*, vol. 22, pp. 1–8, 2020.

[2] H. Liu, Z. Zhao, and Q. She, "Self-supervised ecg pre-training," *Biomedical Signal Processing and Control*, vol. 70, p. 103010, 2021.

[3] M. Buda, A. Maki, and M. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *CoRR*, vol. abs/1710.05381, 2017.

[4] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, 1997.

[6] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv:1502.03167*, 2015.

[7] M. Llamedo and J. P. Martínez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 616–625, 2010.

[8] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.