Universidad Tecnológica Nacional
MAYO - AGOSTO 2026 / Año 24 - Nº 56

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

Arquitectura Híbrida Interpretable basada en Wavelets para la Clasificación Multiclase de Latidos ECG

Interpretable Wavelet-Based Hybrid Architecture for Multiclass ECG Beat Classification

Presentado: 09/03/2026

Aprobado: 14/05/2026

Publicado: 10/06/2026

Cesar Christian Holote Larrosa
https://orcid.org/0009-0007-6335-9466
Universidad Tecnológica Nacional, Facultad Regional La Rioja, Argentina.
ccholote@rec.utn.edu.ar

Marcelo Gómez
Universidad Tecnológica Nacional, Facultad Regional La Rioja, Argentina.
mgomez_ar@hotmail.com

Daniel Turra
Universidad Tecnológica Nacional, Facultad Regional La Rioja, Argentina.
danielnicolast@gmail.com

Resumen

Se presenta una arquitectura híbrida de aprendizaje profundo para la clasificación multiclase de latidos electrocardiográficos (ECG) utilizando la base MIT-BIH Arrhythmia. El enfoque combina bloques convolucionales multiescala tipo Inception con convoluciones separables en profundidad (depthwise separable convolutions), codificadores Transformer unidimensionales (Transformer 1D), unidades recurrentes bidireccionales basadas en Gated Recurrent Units (BiGRU) y mecanismos de atención contextual. La representación de cada latido se enriquece mediante características estadísticas obtenidas a partir de transformadas wavelet discretas multiescala y mediante una rama temporal decimada derivada de la señal original. Para abordar el desbalance de clases se emplean estrategias de aumentación fisiológica controlada, Borderline-SMOTE y una función de pérdida focal ponderada. El modelo alcanzó 99,88 % de exactitud en entrenamiento y 98,97 % en validación, mostrando además elevada separabilidad en el espacio latente y coherencia interpretativa mediante análisis SHapley Additive exPlanations (SHAP). Los resultados sugieren que la integración de representaciones espectro-temporales explícitas con modelado profundo híbrido constituye

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

una alternativa robusta, interpretable y computacionalmente eficiente para la clasificación automática de arritmias cardíacas.

Palabras clave: ECG; clasificación de arritmias cardíacas; transformada wavelet; aprendizaje profundo; transformer; redes neuronales profundas.

Abstract

A hybrid deep learning architecture for multiclass electrocardiographic (ECG) beat classification using the MIT-BIH Arrhythmia database is presented. The proposed approach combines multiscale Inception-based convolutional blocks with depthwise separable convolutions, one-dimensional Transformer encoders (1D Transformers), bidirectional recurrent units based on Gated Recurrent Units (BiGRU), and contextual attention mechanisms. Each heartbeat representation is enriched through statistical features extracted from multiscale discrete wavelet transforms, together with a decimated temporal branch derived from the original signal. To address class imbalance, controlled physiological augmentation strategies, Borderline-SMOTE, and a weighted focal loss function are jointly employed. The model achieved 99.88% training accuracy and 98.97% validation accuracy, while also exhibiting high latent-space separability and consistent interpretability patterns through SHapley Additive exPlanations (SHAP) analysis. The results suggest that integrating explicit spectro-temporal representations with hybrid deep modeling provides a robust, interpretable, and computationally efficient alternative for automatic cardiac arrhythmia classification.

Keywords: ECG; cardiac arrhythmia classification; wavelet transform; deep learning; transformer; deep neural networks.

Introducción

Las enfermedades cardiovasculares continúan siendo una de las principales causas de morbimortalidad a nivel mundial (World Health Organization, 2023), y dentro de ellas las arritmias representan una condición clínica de alto interés por su asociación con eventos agudos y deterioro progresivo del estado hemodinámico. El electrocardiograma (ECG) constituye el estudio de referencia para su identificación; sin embargo, la inspección manual de grandes volúmenes de registros puede resultar lenta, costosa y susceptible a variabilidad interobservador (Jun et al., 2018; Zhao, 2023).

En los últimos años, el aprendizaje profundo ha mostrado un progreso sostenido en la clasificación automática de señales fisiológicas. Las redes convolucionales han demostrado gran capacidad para capturar patrones morfológicos locales del complejo P-QRS-T, como se evidencia en trabajos pioneros basados en redes neuronales convolucionales (Convolutional Neural Networks, CNN) aplicadas a señales ECG (Acharya et al., 2017). Asimismo, las arquitecturas recurrentes y los mecanismos de atención aportan modelado contextual y dependencias temporales de mayor alcance (Meng et al., 2022; Zhao, 2023). No obstante, muchos enfoques centrados exclusivamente en el procesamiento de la señal cruda no explotan de forma explícita la información multirresolución disponible en el dominio wavelet, la cual resulta especialmente útil para describir componentes morfológicos y rítmicos del ECG (He et al., 2025). En este contexto, los enfoques híbridos que integran representaciones multiescala con modelado profundo han comenzado a mostrar ventajas frente a arquitecturas

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

puramente convolucionales o recurrentes, al permitir combinar descriptores espectro-temporales explícitos con mecanismos de aprendizaje capaces de capturar relaciones complejas en la señal (He et al., 2025; Meng et al., 2022; Islam et al., 2023).

A pesar de los avances recientes, numerosos enfoques continúan presentando limitaciones relevantes en escenarios multiclase altamente desbalanceados. Las arquitecturas puramente convolucionales suelen concentrarse principalmente en patrones morfológicos locales, mientras que los modelos recurrentes tradicionales pueden presentar dificultades para capturar dependencias temporales extensas y variabilidad inter-paciente compleja. Por otra parte, los modelos basados exclusivamente en Transformers suelen requerir mayores volúmenes de entrenamiento y complejidad computacional elevada, dificultando su aplicación eficiente en entornos biomédicos con recursos limitados (Meng et al., 2022; Zhao, 2023).

Asimismo, aunque diversos trabajos recientes han comenzado a incorporar transformadas wavelet y mecanismos de atención en clasificación ECG, todavía existe una exploración limitada sobre arquitecturas híbridas capaces de integrar simultáneamente representación multiescala explícita, modelado contextual profundo e interpretabilidad fisiológica dentro de un único esquema de aprendizaje. En particular, la combinación de descriptores wavelet estadísticos con módulos Transformer, mecanismos de atención contextual y análisis explicativos basados en SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017) continúa siendo escasamente estudiada en escenarios de clasificación multiclase de arritmias cardíacas (He et al., 2025; Islam et al., 2023).

En línea con esta tendencia, el presente trabajo propone una arquitectura híbrida que integra tres niveles complementarios de representación: i) descriptores wavelet estadísticos obtenidos a partir de una descomposición multiescala, ii) extracción convolucional multiescala tipo Inception con convoluciones separables (Ismail Fawaz et al., 2020; Chollet, 2017) y iii) modelado global y secuencial mediante codificadores Transformer unidimensionales, unidades recurrentes bidireccionales basadas en Gated Recurrent Units (BiGRU) y mecanismos de atención contextual. Además, el entrenamiento incorpora estrategias de aumentación fisiológica de datos, balanceo mediante Borderline-SMOTE y optimización mediante pérdida focal ponderada, con el objetivo de mejorar la sensibilidad del modelo en clases minoritarias (Han et al., 2005; Lin et al., 2017).

El aporte principal del trabajo consiste en integrar en una misma arquitectura información temporal, espectral y contextual, junto con un análisis de interpretabilidad basado en SHAP y una evaluación estructural mediante estudio de ablación. Los resultados experimentales obtenidos sobre la base MIT-BIH Arrhythmia muestran que la arquitectura propuesta alcanza 98,97 % de exactitud en validación, situándose dentro del estado del arte en clasificación multiclase de latidos ECG, al tiempo que mantiene una complejidad computacional compatible con entornos de investigación aplicada y prototipado clínico.

A partir de este enfoque, las principales contribuciones de este trabajo pueden resumirse en los siguientes puntos:

- Propuesta de una arquitectura híbrida multiescala que integra descriptores wavelet estadísticos con extracción convolucional tipo Inception, codificadores Transformer 1D y modelado secuencial mediante BiGRU, permitiendo combinar información espectral, temporal y contextual de la señal ECG.
- Estrategia integral para manejo del desbalance de clases, basada en aumentación fisiológica de señales, balanceo mediante Borderline-SMOTE y optimización con pérdida focal ponderada, orientada a mejorar la sensibilidad en clases minoritarias de arritmias.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

- Análisis interpretativo del modelo, mediante técnicas SHAP aplicadas a la representación wavelet, que permite identificar las combinaciones de características más influyentes en la clasificación y aportar evidencia sobre la plausibilidad fisiológica de las decisiones del modelo.
- Evaluación estructural mediante estudio de ablación, que permite identificar los componentes arquitectónicos con mayor impacto en el rendimiento y analizar el rol relativo de los distintos módulos del modelo propuesto.

Desarrollo

Base de datos y preprocesamiento

La evaluación se realizó sobre la base MIT-BIH Arrhythmia, compuesta por 48 registros de aproximadamente 30 minutos cada uno, anotados manualmente y ampliamente utilizados como referencia para clasificación de latidos (Goldberger et al., 2000). Para cada latido se extrajo un segmento de 300 muestras centrado en el pico R, de modo de preservar la morfología del complejo QRS y parte de su contexto temporal inmediato.

Se eliminaron las anotaciones no diagnósticas o asociadas a artefactos (~, + y |), conservando únicamente clases clínicamente interpretables. Posteriormente, la señal se normalizó mediante z-score utilizando estadísticas calculadas exclusivamente sobre el conjunto de entrenamiento, con el objetivo de evitar fuga de información.

Para incrementar la diversidad del conjunto de entrenamiento y mejorar la capacidad de generalización del modelo frente a variaciones fisiológicas reales, se aplicaron estrategias de aumentación controlada sobre las señales ECG. Las transformaciones fueron diseñadas bajo restricciones conservadoras orientadas a preservar la plausibilidad morfológica de los latidos y evitar distorsiones incompatibles con la fisiología cardíaca.

La aumentación incluyó tres operaciones principales: i) incorporación de ruido gaussiano de baja amplitud proporcional a la desviación estándar de la señal ($\sigma \approx 0{,}01$), con el objetivo de emular perturbaciones de adquisición y ruido instrumental; ii) desplazamientos temporales aleatorios acotados a ±5 muestras, orientados a representar pequeñas variaciones de alineación temporal del complejo QRS; y iii) transformaciones moderadas de estiramiento/compresión temporal dentro del rango 0,95–1,05, destinadas a simular fluctuaciones leves de frecuencia cardíaca y variabilidad inter-paciente. Estas operaciones fueron aplicadas de manera aleatoria y controlada, preservando la estructura global del complejo P-QRS-T y evitando modificaciones morfológicas extremas, en línea con estrategias de aumentación y representación robusta previamente utilizadas en clasificación automática de señales ECG (Kachuee et al., 2018).

Extracción de características wavelet y balanceo

La extracción de rasgos espectro-temporales se realizó mediante transformada wavelet discreta multiescala (Discrete Wavelet Transform, DWT) de nivel 4, utilizando múltiples familias wavelet discretas, incluyendo Haar, Biorthogonal (bior3.5), Reverse Biorthogonal (rbio2.2, rbio3.1, rbio3.3, rbio4.4, rbio5.5 y rbio6.8), Symlet (sym5), Coiflet (coif5) y Discrete Meyer (dmey). Para cada latido ECG se obtuvieron coeficientes de aproximación y detalle asociados a diferentes bandas temporales y frecuenciales de la señal, permitiendo capturar simultáneamente información morfológica local y comportamiento energético multiescala (Mallat, 1989; Addison, 2005).

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

La utilización simultánea de múltiples familias wavelet tuvo como objetivo explotar propiedades matemáticas complementarias de representación. Las wavelets biortogonales y reverse biorthogonal fueron seleccionadas debido a su simetría aproximada y buena localización temporal, propiedades particularmente útiles para representar complejos QRS abruptos minimizando distorsiones de fase. La wavelet Haar fue incorporada por su elevada sensibilidad frente a discontinuidades rápidas de amplitud, mientras que las familias Symlet y Coiflet aportan mayor regularidad espectral y suavidad en la representación. Por su parte, la wavelet Discrete Meyer (dmey) proporciona una representación más estable de componentes suaves de baja frecuencia y transiciones progresivas de la señal ECG (Daubechies, 1992; Addison, 2005).

A partir de cada subbanda wavelet se extrajeron catorce métricas estadísticas orientadas a caracterizar propiedades energéticas, morfológicas y dinámicas de la señal. Entre ellas se incluyen energía, coeficiente de variación, entropía espectral, curtosis, sesgo, tasa de zero-crossing, percentiles y métricas de pendiente máxima y pendiente media. Estas características permiten describir dispersión, complejidad, asimetría, impulsividad y comportamiento oscilatorio de los coeficientes wavelet, enriqueciendo la representación espectro-temporal del latido ECG y favoreciendo la discriminación entre morfologías cardíacas complejas (Yu & Chen, 2007; He et al., 2025).

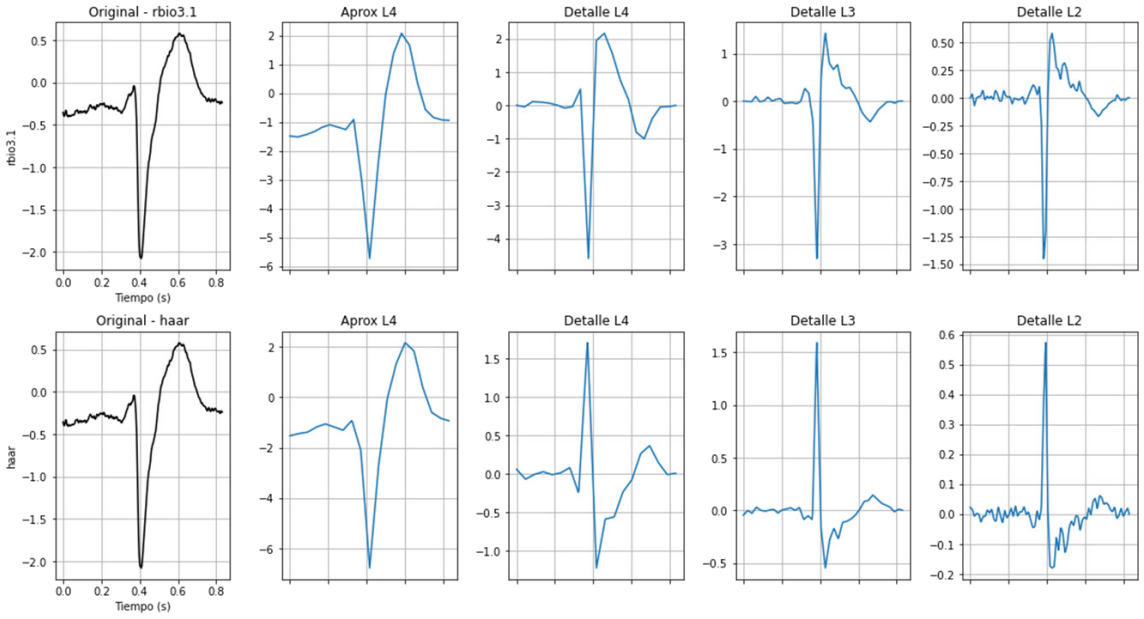

Figura 1. Ejemplo de descomposición wavelet discreta multiescala (nivel 4) aplicada a un latido ECG de la base MIT-BIH Arrhythmia, mostrando coeficientes de aproximación y detalle utilizados para extracción de características espectro-temporales.

Para mitigar el desbalance residual entre clases se aplicó Borderline-SMOTE sobre la representación reorganizada de características, limitando el sobremuestreo a un máximo de 2000 instancias por clase. A diferencia del SMOTE convencional, Borderline-SMOTE concentra la generación sintética en regiones cercanas a las fronteras de decisión, donde suelen

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

concentrarse los ejemplos más difíciles de clasificar (Han et al., 2005). Esta estrategia permitió reforzar particularmente las clases minoritarias con elevada superposición morfológica, reduciendo parcialmente el sesgo inducido por la distribución desigual de muestras. Una vez completado el balanceo, las muestras fueron reestructuradas a la forma requerida por la red neuronal.

Arquitectura propuesta

La arquitectura propuesta fue diseñada para explotar simultáneamente la morfología local del latido, las relaciones de largo alcance dentro de la secuencia y la información multirresolución extraída por wavelets. La Figura 2 resume el flujo completo del modelo, mientras que la Tabla 1 presenta la arquitectura funcional resumida.

La entrada principal del sistema corresponde a un segmento ECG de 300 muestras centrado en el pico R. A partir de este segmento se construyen dos ramas de representación. La primera rama conserva la señal temporal decimada, útil para preservar relaciones morfológicas directas en el dominio temporal. La segunda rama realiza una descomposición wavelet de nivel 4 y calcula 14 estadísticas por subbanda, lo que genera una representación espectro-temporal rica en información de energía, forma y variabilidad local.

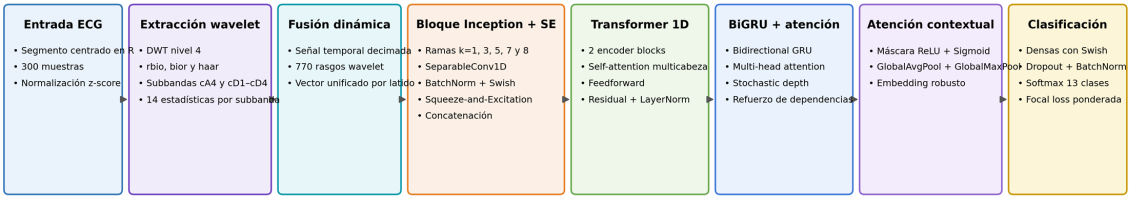

Figura 2. Arquitectura híbrida propuesta para la clasificación multiclase de latidos ECG.

Ambas ramas son unificadas en un módulo de fusión dinámica, que produce un tensor de características apto para el procesamiento profundo posterior. Esta decisión de diseño permite evitar que la red aprenda exclusivamente desde la forma cruda del latido y, al mismo tiempo, le aporta descriptores explícitos con alto valor discriminativo frente a ruido, artefactos y similitud morfológica entre clases.

El bloque de entrada profundo está conformado por una arquitectura multiescala tipo Inception, inspirada en el enfoque original de convoluciones paralelas multiescala y en su adaptación moderna a la clasificación profunda de series temporales (Szegedy et al., 2015; Ismail Fawaz et al., 2020), con seis ramas de convolución paralela y kernels de distinta longitud (2, 3, 4, 5, 6 y 7 muestras). Esta configuración permite capturar simultáneamente patrones morfológicos locales y dependencias temporales de diferente alcance presentes en la señal ECG. Cada rama emplea entre 32 y 48 filtros convolucionales y es seguida por normalización por lotes (Batch Normalization) y activación Swish (Ramachandran et al., 2017).

Posteriormente, las salidas de las ramas son concatenadas y recalibradas mediante módulos Squeeze-and-Excitation (SE) (Hu et al., 2018), los cuales introducen mecanismos de atención a nivel de canal y destacan automáticamente aquellas combinaciones de filtros con mayor relevancia discriminativa para la clasificación de arritmias. En etapas posteriores de la arquitectura se incorporan convoluciones separables en profundidad (depthwise separable convolutions), que separan la filtración espacial por canal de la proyección lineal entre

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

canales y reducen significativamente el número total de parámetros y el costo computacional sin degradar la capacidad representacional del modelo (Chollet, 2017).

Tras la concatenación multiescala, el tensor se refina mediante convoluciones adicionales y regularización espacial. Sobre esta representación se apilan dos codificadores Transformer unidimensionales (Transformer 1D) con mecanismos de autoatención multi-cabeza (multi-head self-attention), redes feedforward, normalización por capas y conexiones residuales. Estos bloques permiten capturar dependencias globales a lo largo de la secuencia, incluyendo relaciones entre regiones morfológicas distantes y variaciones rítmicas representadas dentro de la ventana temporal analizada (Meng et al., 2022; Zhao, 2023).

El modelado secuencial se completa mediante una capa bidireccional basada en Gated Recurrent Units (BiGRU), seleccionada por su capacidad para representar dependencias temporales bidireccionales con menor complejidad computacional respecto de arquitecturas Long Short-Term Memory (LSTM) equivalentes. Sobre la salida recurrente se incorpora una segunda etapa de atención multi-cabeza regularizada mediante stochastic depth, estrategia que mejora la generalización al omitir aleatoriamente ciertas conexiones residuales durante el entrenamiento (Huang et al., 2016).

Finalmente, el modelo incorpora un bloque de atención contextual que genera una máscara aprendida sobre el mapa de características y la combina con Global Average Pooling (GAP) y Global Max Pooling (GMP). Esta agregación dual produce un embedding final robusto, capaz de resumir simultáneamente activaciones salientes e información estadística global. La etapa de clasificación utiliza capas densas con activación Swish, normalización por lotes y dropout, seguidas por una salida softmax de 13 clases. El entrenamiento se optimiza mediante el algoritmo Adam y pérdida focal ponderada, estrategia particularmente efectiva en escenarios con fuerte desbalance de clases (Lin et al., 2017; Pasupa et al., 2020).

En su configuración final, la red contiene 577.781 parámetros entrenables y aproximadamente 90 operaciones funcionales dentro del flujo inferencial. Este compromiso entre profundidad arquitectónica, mecanismos de atención y convoluciones separables permitió alcanzar un elevado rendimiento predictivo manteniendo una complejidad computacional moderada y compatible con entornos de investigación biomédica sin infraestructura aceleradora especializada.

Etapa	Componentes	Función
Entrada	**Segmento ECG centrado en R**	**Representación inicial del latido**
Wavelet	**DWT multiescala nivel 4**	**Descriptores espectro-temporales**
Fusión	**Integración señal + wavelets**	**Tensor combinado**
Extracción local	**Bloques Inception + convoluciones separables**	**Captura morfología multiescala**
Modelado global	**2 × Transformer Encoder (self-attention)**	**Dependencias temporales largas**
Modelado secuencial	**BiGRU bidireccional**	**Contexto temporal**
Atención	**Atención contextual + multi-head**	**Refinamiento discriminativo**
Agregación	**Global Average Pooling (GAP) + Global Max Pooling (GMP)**	**Embedding final**
Clasificación	**Capas densas + softmax**	**Predicción multiclase**

Tabla 1. Arquitectura funcional resumida del modelo propuesto

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

Pseudocódigo científico del modelo

El Algoritmo 1 resume la secuencia operativa del pipeline de entrenamiento y validación empleado en este estudio.

Algoritmo 1. Entrenamiento del modelo híbrido propuesto

Entrada: señales ECG anotadas D, etiquetas y, hiperparámetros H

1: Extraer segmentos x_i de 300 muestras centrados en el pico R

2: Eliminar etiquetas no diagnósticas (~, +, |)

3: Aplicar normalización z-score con estadísticas del conjunto de entrenamiento

4: Generar aumentación fisiológica: ruido gaussiano, desplazamientos y deformaciones temporales controladas

5: Calcular DWT multiescala (nivel 4) con wavelets seleccionadas

6: Obtener 14 estadísticas por subbanda y concatenarlas en un vector wavelet

7: Construir representación temporal decimada de la señal original

8: Fusionar dinámicamente la rama temporal y la rama wavelet

9: Balancear el conjunto de entrenamiento mediante Borderline-SMOTE

10: Alimentar el tensor resultante al bloque Inception + SE

11: Propagar la salida por dos codificadores Transformer 1D

12: Modelar dependencias bidireccionales con BiGRU

13: Aplicar atención multi-cabeza, stochastic depth y atención contextual

14:Obtener embedding final con GlobalAveragePooling (GAP) + GlobalMaxPooling (GMP)

15: Clasificar con capas densas + softmax de 13 clases

16: Optimizar con Adam y focal loss ponderada hasta convergencia

Salida: modelo entrenado M y probabilidades p(y|x) por latido

Configuración experimental

El entrenamiento se realizó en Python utilizando TensorFlow/Keras, con tamaño de lote (batch size) de 128 muestras, optimizador Adam y una tasa de aprendizaje inicial de 0,0005. Se estableció un máximo de 100 épocas, incorporando early stopping con paciencia de 5 épocas y reducción adaptativa de la tasa de aprendizaje (learning rate) ante mesetas de validación.

La función de pérdida empleada fue focal loss con $\gamma = 2$ y pesos α ajustados según la frecuencia de cada clase. Esta formulación reduce la contribución relativa de ejemplos fáciles y concentra el aprendizaje en instancias difíciles o subrepresentadas (Lin et al., 2017). La combinación con Borderline-SMOTE y aumentación fisiológica controlada permitió mejorar la sensibilidad en clases minoritarias sin degradar el rendimiento global del modelo.

El entrenamiento fue ejecutado sobre un entorno de hardware compuesto por procesador Intel Core i7-5600U @ 2.60 GHz, 8 GB de memoria RAM DDR3 y gráficos integrados Intel HD

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

Graphics 5500, sin utilización de GPU dedicada. El proceso convergió mediante early stopping luego de 51 épocas, alcanzando un tiempo total aproximado de entrenamiento de 323 minutos, con un tiempo promedio cercano a 6,3 minutos por época.

A pesar de la complejidad híbrida de la arquitectura propuesta, la utilización de convoluciones separables en profundidad y mecanismos de reducción dimensional permitió mantener una complejidad computacional moderada, compatible con entornos de investigación aplicada y prototipado biomédico sin requerir infraestructura aceleradora especializada.

Resultados y discusión

El modelo propuesto alcanzó 99,88 % de exactitud en entrenamiento y 98,97 % en validación, lo que indica una elevada capacidad de aprendizaje y una generalización estable sobre la base MIT-BIH Arrhythmia. Las métricas globales del modelo se resumen en la Tabla 2. En el conjunto de validación se obtuvo 98,97 % de precisión, 98,88 % de recall y un F1-score global de 98,92 %, lo que confirma un desempeño robusto y equilibrado incluso en presencia de desbalance entre clases. Estos resultados sitúan el rendimiento del modelo dentro del estado del arte para clasificación multiclase de latidos ECG en MIT-BIH, obteniendo un valor superior al 98,21 % de exactitud reportado por el método HARDC de Islam et al. (2023) en un escenario comparable de 13 clases.

Métrica	Valor
Exactitud de entrenamiento	**99,88 %**
Exactitud de validación	**98,97 %**
Precisión en validación	**98,97 %**
Recall en validación	**98,88 %**
F1-score en validación	**98,92 %**
Parámetros entrenables	**577.781**
Batch size	**128**

Tabla 2. Métricas globales del modelo propuesto.

La matriz de confusión normalizada (Figura 3) muestra que la mayoría de las clases presentan tasas de acierto superiores al 93 %, con un desempeño particularmente sólido en las categorías más frecuentes y mejoras sostenidas en clases minoritarias gracias a la combinación de aumentación fisiológica, Borderline-SMOTE y pérdida focal ponderada.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

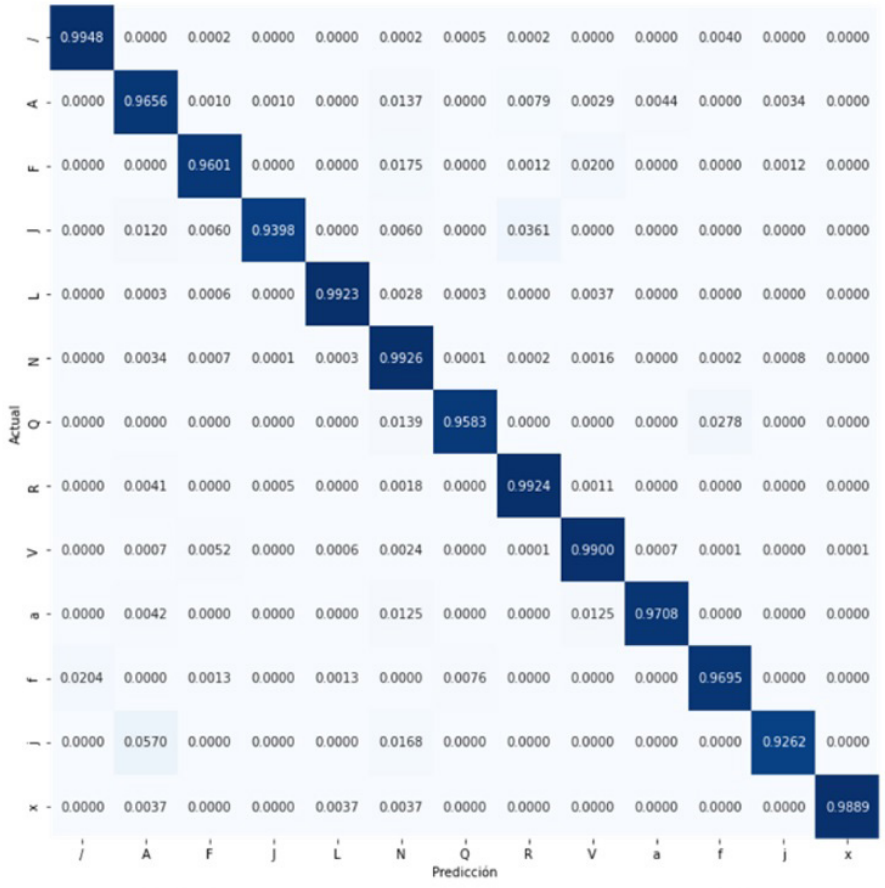

Figura 3. Matriz de confusión normalizada para la clasificación de 13 clases de latidos ECG.

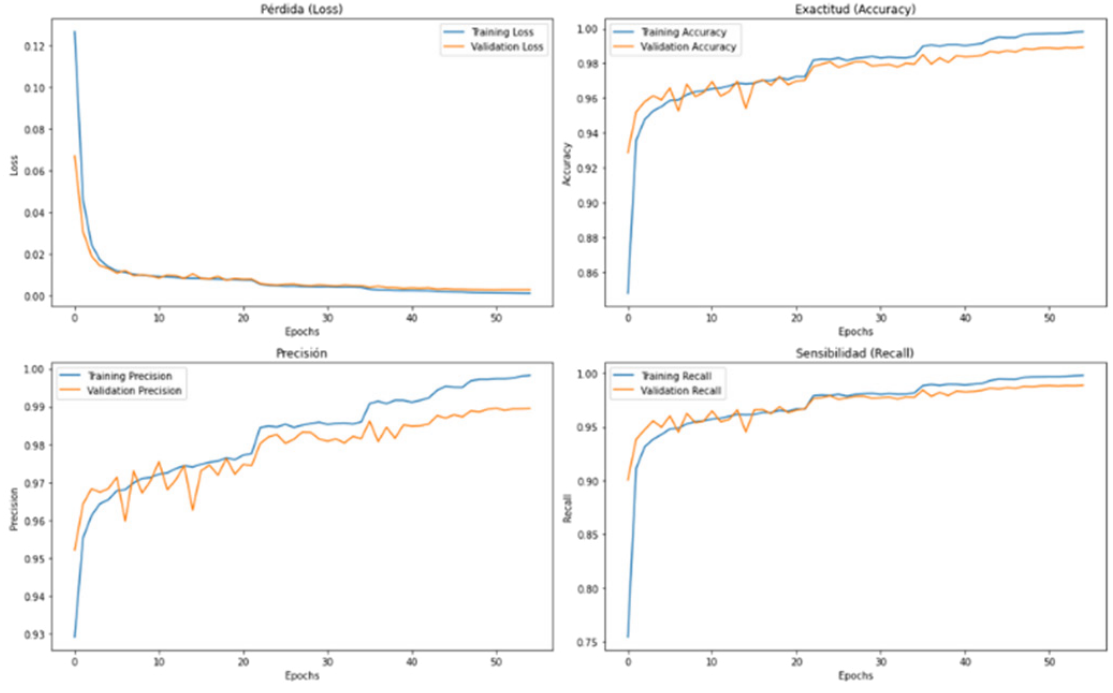

Figura 4. Evolución de la pérdida, la exactitud, la precisión y el recall durante entrenamiento y validación.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

Debe destacarse que varias de las clases minoritarias presentes en MIT-BIH presentan una elevada variabilidad intra-clase y fuerte similitud morfológica inter-clase, particularmente en latidos supraventriculares y fusionados. Este comportamiento incrementa la dificultad discriminativa incluso en modelos profundos, ya que pequeñas variaciones temporales o energéticas pueden modificar significativamente la representación latente. En este contexto, la combinación de descriptores wavelet multiescala y mecanismos de atención permitió estabilizar parcialmente la separación entre clases minoritarias, aunque persisten regiones de solapamiento observables en la proyección t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten & Hinton, 2008).

Las curvas de aprendizaje presentadas en la Figura 4 evidencian una convergencia rápida y estable del proceso de optimización. La evolución paralela de la pérdida y las métricas de validación sugiere que el modelo logra capturar patrones discriminativos sin incurrir en sobreajuste significativo. Este comportamiento resulta consistente con la presencia de mecanismos de regularización estructural, tales como dropout, stochastic depth y normalización por lotes, que contribuyen a estabilizar el entrenamiento en arquitecturas profundas.

Desde el punto de vista representacional, la proyección t-distributed Stochastic Neighbor Embedding (t-SNE) del espacio latente (Figura 5) revela la formación de agrupamientos relativamente bien definidos para varias de las clases analizadas. El algoritmo t-SNE proyecta representaciones latentes de alta dimensión hacia un espacio bidimensional preservando relaciones probabilísticas de vecindad local. En consecuencia, las coordenadas obtenidas no representan variables fisiológicas originales sino relaciones de similitud aprendidas por la red neuronal.

La separación observada sugiere que la arquitectura logra construir representaciones diferenciadas a partir de la integración de información temporal, espectral y contextual. Sin embargo, persisten regiones parciales de solapamiento entre algunas clases minoritarias, particularmente en latidos supraventriculares y fusionados, comportamiento consistente con la elevada variabilidad intra-clase y similitud morfológica inter-clase presente en la base MIT-BIH Arrhythmia. Estas dificultades discriminativas también han sido reportadas en trabajos previos sobre clasificación multiclase de arritmias ECG (He et al., 2025; Islam et al., 2023).

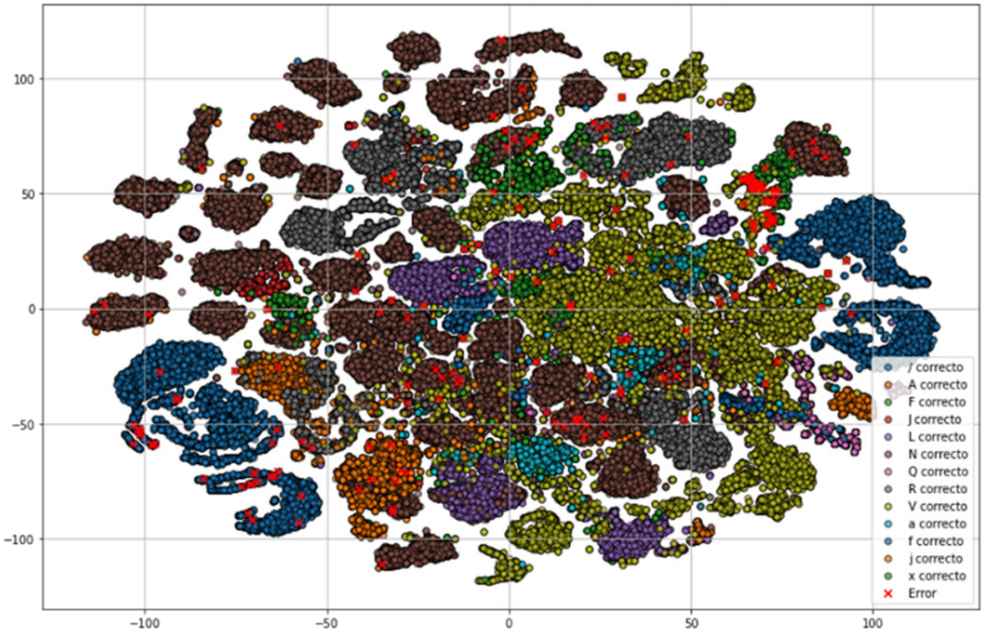

Figura 5. Proyección t-SNE del espacio latente aprendido por el modelo para las 13 clases analizadas.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

Una comparación cualitativa con enfoques recientes se presenta en la Tabla 3. Esta comparación permite situar el desempeño del modelo dentro del estado del arte. En particular, frente al método HARDC propuesto por Islam et al. (2023), que reporta aproximadamente 98,21 % de exactitud en un escenario comparable de 13 clases sobre MIT-BIH, la arquitectura presentada en este trabajo alcanza 98,97 % de exactitud de validación, incorporando además mecanismos explícitos de interpretabilidad basados en análisis wavelet y explicaciones SHapley Additive exPlanations (SHAP). Estos resultados sugieren que la integración de descriptores espectro-temporales explícitos con modelado profundo híbrido puede aportar ventajas frente a enfoques dependientes exclusivamente de la señal cruda.

Referencia	Base	Rasgos explícitos	Atención / secuencia	Comentario
Jun et al. (2018)	**MIT-BIH**	**No**	**Red neuronal convolucional 2D**	**Buen punto de partida basado en imagen de ECG**
Meng et al. (2022)	**ECG multiclase**	**No**	**Transformer temporal**	**Fuerte modelado global de secuencias**
He et al. (2025)	**ECG multiclase**	**Sí, DWT**	**CNN-BiGRU con atención**	**Integra wavelets y dependencia temporal**
Islam et al. (2023)	**MIT-BIH, 13 clases**	**No**	**CNN residual con atención**	**98,21 % de validación reportada**
Propuesto	**MIT-BIH, 13 clases**	**Sí, DWT multiescala**	**Inception + Transformer + BiGRU + atención**	**98,97 % de validación y análisis interpretativo**

Tabla 3. Comparación cualitativa con enfoques recientes relacionados.

Desde una perspectiva arquitectónica, los resultados sugieren que la combinación de tres niveles de representación resulta particularmente efectiva para el análisis de señales ECG. En primer lugar, la transformada wavelet multiescala proporciona una descripción compacta del contenido frecuencial del latido, permitiendo capturar variaciones energéticas y morfológicas relevantes en distintas bandas. En segundo lugar, los bloques Inception con convoluciones separables permiten modelar la morfología del complejo P-QRS-T en diferentes escalas temporales. Finalmente, los módulos Transformer y BiGRU aportan capacidad para capturar dependencias globales y dinámicas secuenciales, lo que resulta especialmente útil para distinguir latidos con morfologías similares, pero contexto rítmico diferente.

El análisis de ablación presentado en el anexo técnico respalda esta interpretación. En particular, la eliminación del módulo de atención multi-cabeza produce la mayor degradación en la precisión media, mientras que los bloques BiGRU e Inception también muestran contribuciones relevantes al rendimiento global del modelo, lo que indica que estos componentes constituyen el núcleo estructural de la arquitectura. Los mecanismos de atención y regularización, incluyendo Squeeze-and-Excitation y stochastic depth, actúan principalmente como refinadores de la representación, contribuyendo a mejorar la robustez del modelo frente a ruido, variabilidad intra-clase y similitud morfológica entre arritmias.

Un aspecto especialmente relevante del enfoque propuesto es su interpretabilidad. El análisis SHAP (Lundberg & Lee, 2017) realizado sobre la representación wavelet revela patrones consistentes en las combinaciones de características que más influyen en las decisiones del modelo. En particular, se observa una recurrencia significativa de combinaciones

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

como rbio3.1 + cD3 + coef_var, lo que sugiere que la red explota información contenida en frecuencias intermedias estrechamente asociadas a la morfología del complejo QRS y de la onda T. Este comportamiento aproxima la lógica interna del modelo a una interpretación fisiológicamente plausible, lo que constituye una ventaja frente a modelos de caja negra puramente basados en señal cruda.

Desde una perspectiva más amplia, los resultados obtenidos se alinean con una tendencia creciente en la literatura reciente: la transición desde arquitecturas puramente convolucionales hacia modelos híbridos que integran representaciones multiescala, mecanismos de atención y modelado secuencial profundo. En este contexto, la arquitectura propuesta puede interpretarse como una aproximación que combina explícitamente información espectral, temporal y contextual, lo que contribuye a mejorar la discriminación entre morfologías complejas de latidos.

No obstante, el estudio presenta algunas limitaciones. En primer lugar, la evaluación se realizó exclusivamente sobre la base MIT-BIH Arrhythmia, ampliamente utilizada en la literatura, pero limitada en términos de diversidad clínica y condiciones de adquisición. La validación en bases adicionales, como INCART, CUDB o PTB-XL, permitiría evaluar la capacidad de generalización del modelo en contextos más heterogéneos. En segundo lugar, si bien la arquitectura fue diseñada con convoluciones separables y un número moderado de parámetros, el trabajo no aborda en profundidad estrategias de compresión o cuantización del modelo para despliegue en dispositivos embebidos o sistemas de monitoreo continuo.

En conjunto, los resultados sugieren que la combinación de descriptores wavelet multiescala con arquitecturas profundas híbridas constituye una estrategia prometedora para el análisis automático de señales ECG. Este enfoque no solo permite mejorar el rendimiento en tareas de clasificación multiclase, sino que también facilita la interpretación de los mecanismos que sustentan las decisiones del modelo, aspecto cada vez más relevante en aplicaciones clínicas basadas en inteligencia artificial.

Conclusiones

Se desarrolló una arquitectura híbrida profunda para clasificación multiclase de latidos electrocardiográficos (ECG) que integra representación wavelet multiescala, extracción convolucional tipo Inception con convoluciones separables en profundidad, codificadores Transformer unidimensionales (Transformer 1D), unidades recurrentes bidireccionales basadas en Gated Recurrent Units (BiGRU) y mecanismos de atención contextual.

Los resultados obtenidos demuestran que la integración explícita de información espectral, temporal y contextual constituye una estrategia efectiva para escenarios con morfologías heterogéneas, elevada variabilidad intra-clase y desbalance significativo entre categorías de arritmias. Asimismo, el análisis interpretativo basado en SHAP sugiere que el modelo aprende patrones fisiológicamente consistentes asociados a subbandas wavelet relevantes del complejo QRS, reforzando la plausibilidad clínica de las decisiones generadas por la arquitectura propuesta.

Desde el punto de vista computacional, la arquitectura mantuvo una complejidad moderada y pudo ser entrenada sobre hardware sin aceleración GPU dedicada, lo que refuerza su potencial aplicabilidad en entornos de investigación biomédica y prototipado clínico con recursos limitados.

Como línea futura, se propone validar el modelo sobre bases de datos de mayor diversidad

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

poblacional y complejidad clínica, explorar reemplazos de BiGRU mediante Temporal Convolutional Networks (TCN) (Bai et al., 2018) y evaluar estrategias de compresión y optimización orientadas al despliegue en sistemas embebidos o plataformas de monitoreo cardíaco en tiempo real.

En conjunto, los resultados sugieren que las arquitecturas híbridas multirrepresentación constituyen una dirección prometedora para el desarrollo de sistemas de diagnóstico asistido por inteligencia artificial aplicados al análisis automático de señales ECG.

Disponibilidad de datos y material respaldatorio

Los datos primarios utilizados en este estudio corresponden a la base pública MIT-BIH Arrhythmia Database, disponible en PhysioNet (Goldberger et al., 2000). El conjunto de scripts, parámetros experimentales y salidas de entrenamiento que respaldan los resultados puede ser provisto por el autor de correspondencia para fines académicos y de verificación metodológica.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

Referencias

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., Tan, R. S. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, 89, 389–396. https://doi.org/10.1016/j.compbiomed.2017.08.022

Addison, P. S. (2005). Wavelet transforms and the ECG: A review. *Physiological Measurement*, 26(5), R155–R199. https://doi.org/10.1088/0967-3334/26/5/R01

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 1251-1258. https://doi.org/10.1109/CVPR.2017.195

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics (SIAM). https://doi.org/10.1137/1.9781611970104

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215-e220. https://doi.org/10.1161/01.CIR.101.23.e215

Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In: Huang, DS., Zhang, XP., Huang, GB. (eds). *Advances in Intelligent Computing. ICIC 2005. Lecture Notes in Computer Science, vol 3644.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/11538059_91

He, X., Hu, C., Ma, K., Huang, J., & He, H. (2025). ECG-based arrhythmia classification using discrete wavelet transform and attention-enhanced CNN-BiGRU model. *Physical and Engineering Sciences in Medicine*, 48, 1995–2009. https://doi.org/10.1007/s13246-025-01639-6

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, K. Q. (2016). Deep networks with stochastic depth. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds). *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9908.* Springer, Cham. https://doi.org/10.1007/978-3-319-46493-0_39

Islam, M. S., Hasan, K. F., Sultana, S., Uddin, S., Lio', P., Quinn, J. M. W., & Moni, M. A. (2023). HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN. *Neural Networks,* 162, 271–287. https://doi.org/10.1016/j.neunet.2023.03.004

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P. A., & Petitjean, F. (2020). InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery, 34*(6), 1936–1962. https://doi.org/10.1007/s10618-020-00710-y

Jun, T. J., Nguyen, H. M., Kang, D., Kim, D., Kim, Y.-H., & Kim, D. (2018). ECG arrhythmia classification using a 2-D convolutional neural network. arXiv preprint arXiv:1804.06812.

Kachuee, M., Fazeli, S., & Sarrafzadeh, M. (2018). ECG heartbeat classification: A deep transferable representation. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 443–444. https://doi.org/10.1109/ICHI.2018.00092

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

Lin, T. -Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision,* Venice, Italy. 2980-2988. https://doi.org/10.1109/ICCV.2017.324

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*. https://arxiv.org/abs/1705.07874

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693. https://doi.org/10.1109/34.192463

Meng, L., Tan, W., Ma, J., Wang, R., Yin, X., & Zhang, Y. (2022). Enhancing dynamic ECG heartbeat classification with lightweight transformer model. *Artificial Intelligence in Medicine, 124*, 102236. https://doi.org/10.1016/j.artmed.2022.102236

Pasupa, K., Vatathanavaro, S., & Tungjitnob, S. (2020). Convolutional neural network based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification. arXiv preprint arXiv:2001.03329.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605. https://www.jmlr.org/papers/v9/vandermaaten08a.html

World Health Organization. (2023). Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Yu, S. -N., & Chen, Y. -H. (2007). Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. *Pattern Recognition Letters, 28*(10), 1142–1150. https://doi.org/10.1016/j.patrec.2007.01.017

Zhao, Z. (2023). Transforming ECG Diagnosis: An In-depth Review of Transformer-based Deep Learning Models in Cardiovascular Disease Detection. *arXiv preprint arXiv:2306.01249*.

Contribución de los Autores

Nombres y Apellidos del autor	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Holote Larrosa	x		x	x	x	x	x	x	x	x	x	x	x	x
Gómez		x	x	x		x	x	x				x	x	x
Turra			x				x	x					x	x

1-Administración del proyecto, 2-Adquisición de fondos, 3-Análisis formal, 4-Conceptualización, 5-Curaduría de datos, 6-Escritura - revisión y edición, 7-Investigación, 8-Metodología, 9-Recursos, 10-Redacción - borrador original, 11-Software, 12-Supervisión, 13-Validación, 14-Visualización.

Conflicto de intereses y consideraciones éticas

Los autores declaran no poseer conflictos de intereses. El estudio se realizó exclusivamente sobre una base de datos pública y anonimizada (MIT-BIH Arrhythmia), por lo que no implicó intervención directa sobre pacientes ni recolección de datos sensibles identificables.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026- ISSN 1666-6933

Anexo técnico complementario

La Tabla 5 resume el impacto observado al remover componentes clave de la arquitectura. Esta síntesis permite identificar qué bloques sostienen mayor parte del rendimiento y cuáles podrían simplificarse en escenarios con restricciones computacionales.

Variante	Exactitud global (%)	Clases más afectadas	Interpretación
Modelo completo	**99,88 %**	**—**	**Referencia de máximo desempeño**
Sin BiGRU	**99,75 %**	**A, f**	**Pérdida de contexto bidireccional**
Sin Inception	**99,76 %**	**A, F**	**Menor captación multiescala**
Sin Multi-Head Attention	**99,40 %**	**A, F, f**	**Mayor degradación ante morfologías complejas**
Sin Stochastic Depth	**99,80 %**	**A, V**	**Menor regularización estructural**
Sin atención contextual	**99,83 %**	**A, F**	**Menor refinamiento contextual**
Sin Transformer 1D	**99,78 %**	**V, f**	**Menor codificación global**
Sin bloque Squeeze-and-Excitation	**99,85 %**	**F, f**	**Menor recalibración por canal**
Sin pooling combinado	**99,79 %**	**A, F**	**Se pierde complementariedad estadística**
Sin densas intermedias	**99,03 %**	**—**	**Simplifica el modelo sin degradación significativa**

Tabla 5. Estudio de ablación de los principales componentes de la arquitectura propuesta.

Las Figuras 6 y 7 ilustran dos hallazgos interpretativos relevantes del análisis SHAP agregado: la jerarquía de wavelets con mayor contribución total y la dominancia de la subbanda cD3 sobre el resto de las subbandas consideradas.

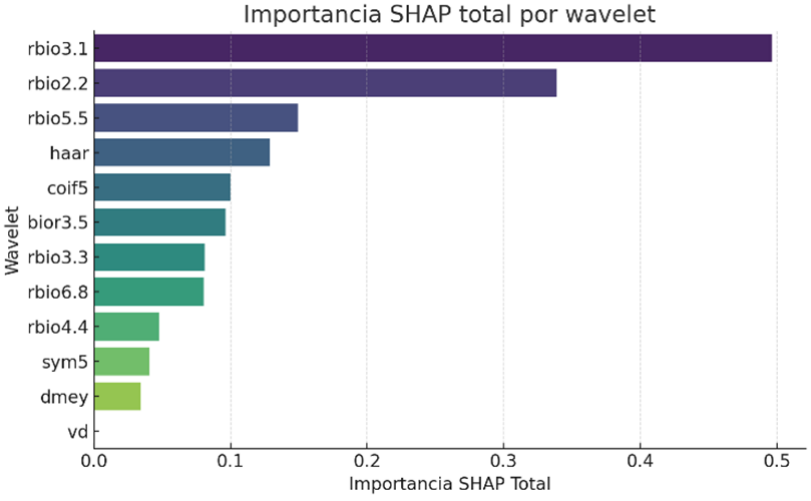

Figura 6. Importancia SHAP total agregada por wavelet utilizada en la representación multiescala.

Revista Tecnología y Ciencia
DOI: https://doi.org/10.33414/rtyc.56.22-39.2026 - ISSN 1666-6933

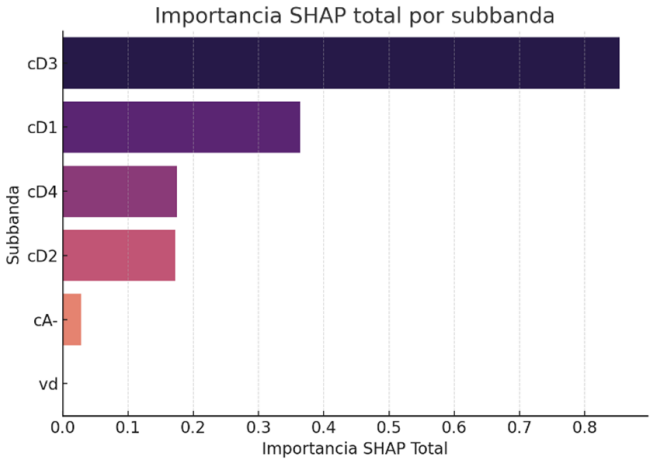

Figura 7. Importancia SHAP total agregada por subbanda wavelet; se destaca el peso relativo de cD3.

En conjunto, el anexo confirma que los bloques Inception y BiGRU conforman el núcleo más sensible de la propuesta, mientras que la atención multi-cabeza y la regularización estocástica aportan robustez adicional. La evidencia interpretativa también apoya la selección de wavelets biortogonales y de la subbanda cD3 como componentes particularmente informativos para la clasificación de arritmias.

A nivel desagregado por clase, el análisis interpretativo muestra que no todas las combinaciones wavelet-subbanda-estadístico aportan de la misma forma a la decisión final. La Figura 8 resume, para cada clase, las cinco características con mayor contribución SHAP agregada. Este patrón confirma que la red no depende de un único descriptor global, sino de un conjunto jerárquico de rasgos específicos que varía según la morfología de cada arritmia.

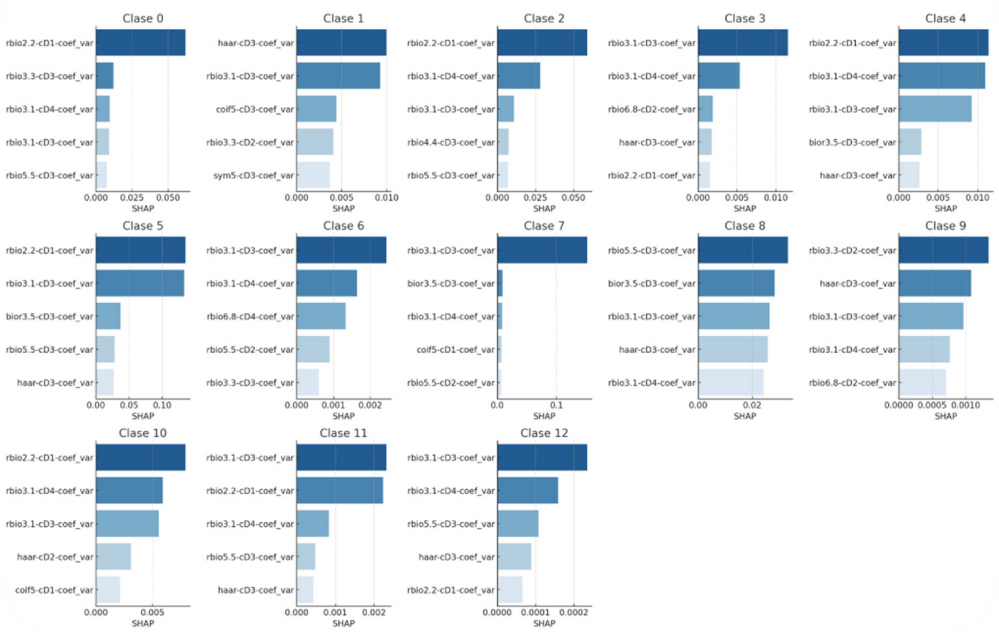

Figura 8. Características más influyentes por clase según el análisis SHAP agregado.

La consistencia de estas pautas interpretativas fortalece la validez del diseño propuesto y aporta un argumento adicional para futuras estrategias de selección automática de características y compresión dirigida del modelo.