Optimización de Selección de Características mediante Algoritmo Genético para la Clasificación de Arritmias Cardíacas con Señales ECG

Genetic Algorithm-Based Optimization of Feature Selection for Arrhythmia Classification in ECG Signals

Josue Valenzuela Perez[1]
[1] Universidad de Guanajuato. j.valenzuelaperez@ugto.mx

Resumen

La clasificación automática de arritmias cardíacas a partir de señales de electrocardiograma (ECG) suele apoyarse en conjuntos amplios de características morfológicas, temporales y espectrales que incrementan el costo computacional sin garantizar mejoras proporcionales en desempeño. Este trabajo compara tres estrategias de reducción del espacio de características para la clasificación binaria de latidos (Normal vs. Anormal) sobre la base de datos MIT-BIH Arrhythmia Database: (1) un modelo baseline con el conjunto completo de 30 características, (2) reducción mediante Análisis de Componentes Principales (PCA) reteniendo el 95% de la varianza, y (3) selección de características mediante un Algoritmo Genético (GA) con codificación binaria. Se extrajeron 109,398 latidos de 48 registros, y la validación se realizó mediante validación cruzada estratificada y agrupada por paciente (StratifiedGroupKFold) para evitar fuga de información entre pacientes. El subconjunto seleccionado por el GA (11 de 30 características) obtuvo el mejor desempeño en todas las métricas evaluadas: exactitud de 91.6% (±2.0%), F1 macro de 0.844 (±0.041), sensibilidad de 71.1% y especificidad de 96.0%, superando tanto al baseline (30 características, 90.2% exactitud) como a PCA (15 componentes, 88.5% exactitud), y con un tiempo de inferencia 22% menor que el baseline. Estos resultados muestran que la selección dirigida por GA no solo reduce la dimensionalidad, sino que mejora la generalización entre pacientes frente a técnicas de reducción no supervisada como PCA.

Palabras clave: algoritmo genético; selección de características; clasificación de arritmias; electrocardiograma; MIT-BIH.

Introducción

Las enfermedades cardiovasculares continúan siendo una de las principales causas de mortalidad a nivel mundial, y la detección temprana de arritmias mediante el análisis automatizado del electrocardiograma (ECG) es una herramienta clave para el diagnóstico oportuno. Los sistemas de clasificación automática de latidos típicamente extraen decenas de características por latido: morfológicas, temporales (intervalos RR) y espectrales (coeficientes wavelet), antes de aplicar un clasificador supervisado. Sin embargo, no todas las características aportan igual poder discriminativo, y conjuntos de alta dimensionalidad incrementan el costo computacional de entrenamiento e inferencia, un factor crítico en sistemas de monitoreo en tiempo real o dispositivos embebidos de bajo consumo.

La reducción de dimensionalidad es un paso común para mitigar este problema. El Análisis de Componentes Principales (PCA) es la técnica más utilizada, pero al ser una transformación no supervisada, no considera la relación entre las características y la variable objetivo, lo que puede descartar información discriminativa relevante para la clasificación. Como alternativa, los Algoritmos Genéticos (GA) permiten formular la selección de características como un problema de optimización combinatoria, donde la función de aptitud (fitness) puede incorporar directamente el desempeño del clasificador, guiando la búsqueda hacia subconjuntos que maximizan la capacidad discriminativa real.

Un problema frecuentemente subestimado en trabajos de clasificación de arritmias sobre MIT-BIH es la fuga de información entre pacientes: al dividir los latidos de forma aleatoria en validación cruzada, latidos del mismo paciente terminan simultáneamente en entrenamiento y prueba, y dado que la morfología del ECG es altamente específica de cada individuo, el clasificador puede memorizar patrones del paciente en lugar de

aprender el patrón de la arritmia, inflando artificialmente las métricas reportadas (de Chazal *et al.,* 2004). Este trabajo controla explícitamente este efecto mediante validación cruzada agrupada por paciente.

El objetivo de este trabajo es comparar, bajo un protocolo de validación libre de fuga entre pacientes, tres estrategias de manejo de características (conjunto completo, PCA y selección por GA) para la clasificación binaria de latidos ECG, evaluando exactitud, F1, sensibilidad, especificidad, número de características y tiempo de inferencia por latido.

Metodología Propuesta

Esta sección describe, en orden de ejecución, cada etapa del pipeline: desde la señal cruda de ECG hasta la comparación final de tres estrategias de manejo de características. Cada decisión de diseño se justifica antes de describirse técnicamente, de forma que el procedimiento completo sea replicable sin depender de conocimiento implícito. La Figura 1 resume el flujo completo del pipeline.

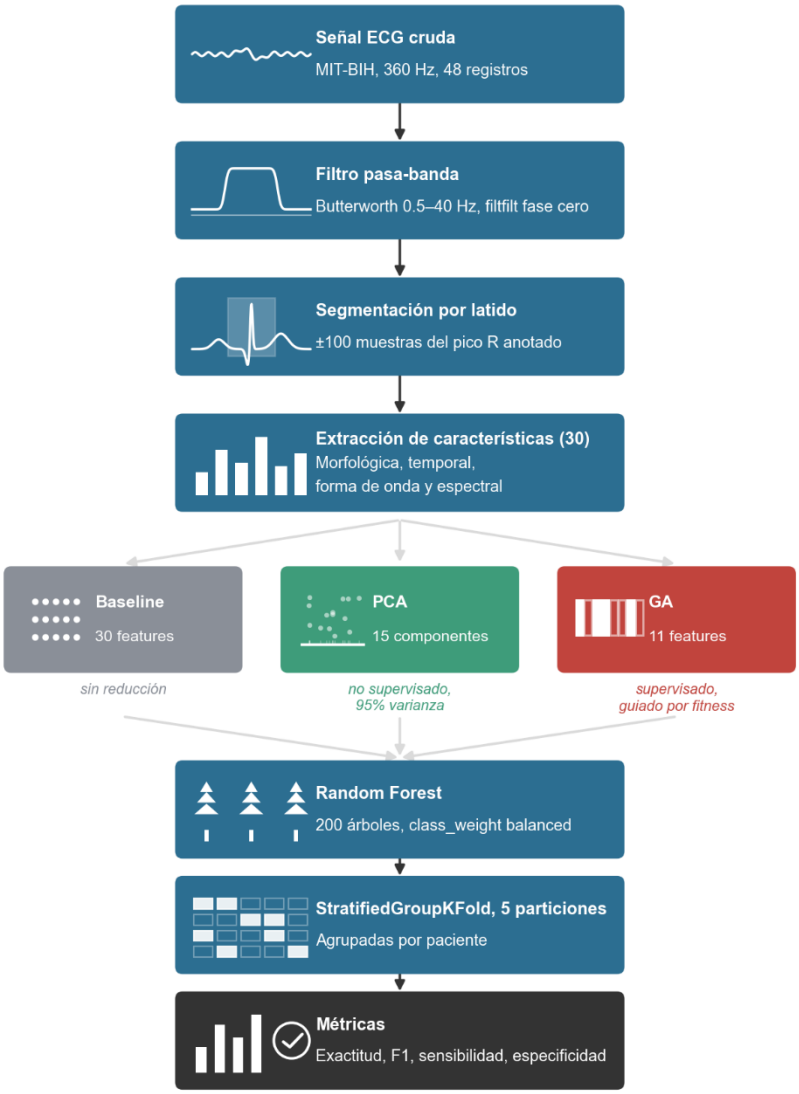


Figura 1. *Diagrama de flujo del pipeline completo: preprocesamiento, extracción de características, bifurcación en las tres estrategias comparadas (baseline, PCA, GA) y evaluación conjunta bajo el mismo protocolo de validación.*
Fuente: elaboración propia.

2.1 El latido ECG

Antes de describir el procesamiento aplicado, es necesario definir los componentes de un latido cardíaco tal como aparecen en el ECG, ya que todas las características extraídas en este trabajo (sección 2.5) están definidas en función de ellos.

Un latido normal produce tres deflexiones características, en este orden temporal (Figura 2):

- Onda P: primera deflexión, de baja amplitud, corresponde a la despolarización auricular (paso previo a la contracción). Su amplitud es varias veces menor que la del QRS.
- Complejo QRS: deflexión de mayor amplitud y pendiente del latido, corresponde a la despolarización ventricular y es la referencia estándar para ubicar un latido (el "pico R"). Duración normal corta (menor a 120 ms); un QRS anormalmente ancho indica que el impulso no siguió la vía de conducción normal, como en los latidos ventriculares ectópicos.
- Onda T: deflexión más amplia y lenta después del QRS, corresponde a la repolarización ventricular. Cambios en su forma o amplitud se asocian a isquemia cardíaca.

A esto se suman el intervalo PR (inicio de P a inicio de QRS, tiempo de conducción auriculoventricular) y el intervalo RR (entre dos picos R consecutivos, duración de un ciclo cardíaco). Este último es central para este trabajo: su irregularidad es uno de los indicadores más directos de arritmia (sección 2.5.2).

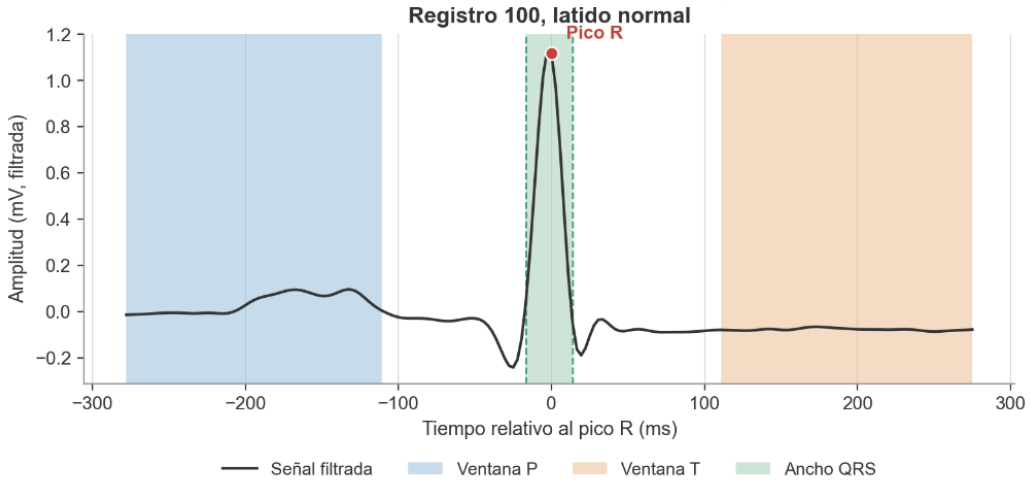


Figura 2. Anatomía de un latido normal (registro 100 de la base de datos), con las ventanas y umbrales usados por el pipeline para estimar cada componente: ventana P, ancho de QRS (umbral de amplitud) y ventana T.
Fuente: elaboración propia.

2.2. Base de datos

Se utilizó la MIT-BIH Arrhythmia Database (Moody y Mark, 2001), la base de datos de referencia más utilizada en la literatura de clasificación automática de arritmias desde su publicación. Contiene 48 registros de ECG de media hora de duración, de dos derivaciones, muestreados a 360 Hz, provenientes de 47 pacientes distintos. Cada latido fue anotado manualmente por al menos dos cardiólogos, marcando tanto la posición exacta del pico R como el tipo de latido (normal, ectópico ventricular, supraventricular, etc.). Se eligió esta base de datos precisamente porque sus anotaciones, validadas clínicamente y públicamente disponibles, permiten evaluar un clasificador contra una referencia confiable ("ground truth"), en lugar de depender de etiquetas generadas automáticamente.

2.3. Preprocesamiento de la señal

Una señal de ECG cruda contiene dos fuentes principales de ruido: la deriva de línea base (baseline wander, oscilaciones de baja frecuencia por respiración/movimiento) y el ruido muscular/de alta frecuencia (actividad electromiográfica, interferencia eléctrica). Sin eliminarlos, distorsionan la amplitud real de las ondas, afectando las características morfológicas de la sección 2.1.

Se aplicó un filtro pasa-banda Butterworth de orden 4 (0.5-40 Hz): el límite inferior elimina la deriva de línea base sin afectar el contenido bajo del ECG (como la onda T); el superior elimina ruido muscular conservando la energía del QRS, concentrada por debajo de esa frecuencia. Es una práctica estándar en procesamiento de ECG (Van Alste & Schilder, 1985).

El filtro se implementó con fase cero (scipy.signal.filtfilt): un filtro causal calcula cada punto usando solo muestras *pasadas*, lo que introduce un retraso dependiente de la frecuencia y deja la señal filtrada *corrida en el tiempo* respecto a la original. filtfilt evita esto aplicando el filtro dos veces, una vez hacia adelante y otra hacia atrás sobre el resultado: el retraso del primer pase se cancela exactamente con el segundo, y la señal queda alineada en el tiempo con la original, sin desfase.

Esto es necesario porque los picos R usados para segmentar (sección 2.4) provienen de anotaciones sobre la señal sin filtrar; si el filtrado desplazara la señal en el tiempo, esas posiciones dejarían de coincidir con el pico real, desalineando las mediciones de QRS y las ventanas P/T.

2.4. Segmentación de latidos y etiquetado

Para analizar cada latido de forma individual, la señal continua debe dividirse en segmentos centrados en cada latido. Existen dos formas de ubicar el punto central (pico R): mediante un algoritmo de detección automática, o utilizando las anotaciones ya validadas por los cardiólogos que incluye la base de datos. Se optó por la segunda opción: un detector automático introduciría su propio margen de error (falsos positivos, desplazamientos de muestras) que se mezclaría con el desempeño del clasificador, el verdadero objeto de estudio. Así, cualquier diferencia entre baseline, PCA y GA se atribuye exclusivamente a la etapa de características/clasificación.

A partir de cada posición anotada se extrajo una ventana fija de 100 muestras antes y 100 después del pico R (≈556 ms totales a 360 Hz), suficiente para contener la onda P, el complejo QRS y la onda T de un latido típico (Figura 2).

Las anotaciones de MIT-BIH usan más de una decena de símbolos, agrupados según el estándar AAMI EC57 (AAMI, 2012) en cinco superclases (Normal, Supraventricular, Ventricular, Fusión, Desconocido). Para mantener el alcance de este primer estudio comparativo, estas se colapsaron a un esquema binario Normal vs. Anormal, consistente con trabajos previos que abordan primero el problema binario antes de escalar a multiclase (de Chazal et al., 2004).

2.5. Extracción de características

Con base en la anatomía de la sección 2.1, discriminar Normal de Anormal requiere información complementaria: forma del latido (morfología), ritmo cardíaco (intervalos) y contenido frecuencial (no visible en el dominio del tiempo). Se extrajeron 30 características, en cuatro grupos.

2.5.1. Características morfológicas

Estas características describen la forma del complejo QRS y las ondas adyacentes (Figura 2), cuyas alteraciones son indicadores clínicos directos de arritmia (por ejemplo, un QRS ensanchado es característico de latidos ventriculares ectópicos, mientras que cambios en la onda T se asocian a isquemia, sección 2.1).

- Amplitud del pico R: valor de la señal filtrada en la muestra correspondiente al pico R, tomado directamente de la anotación del latido.
- Ancho del complejo QRS: el complejo QRS es, por mucho, la deflexión de mayor amplitud y pendiente del latido; las ondas P y T que lo rodean son varias veces más pequeñas (sección 2.1). Por eso, alejándose del pico R hacia ambos lados, la señal cae abruptamente al salir del QRS: ese salto de amplitud sirve como frontera, sin necesidad de un segmento dedicado.

  Concretamente, se recorre la señal muestra por muestra desde el pico R mientras la amplitud se mantenga por encima de un umbral relativo:

  $$|x[n]| > 0.1 \cdot |x[R]| \quad (1)$$

  Se usa un umbral relativo (10% del propio pico R de cada latido) y no absoluto porque la amplitud varía de paciente a paciente: un valor fijo en milivoltios no serviría igual en todos los latidos. Al normalizar contra cada pico individual, el criterio se adapta solo a la escala de cada uno.

El recorrido se detiene donde la condición deja de cumplirse a cada lado (qrs_width_samples() en preprocess_features.py); esa distancia, convertida a milisegundos con la frecuencia de muestreo (360 Hz), es el ancho del QRS (región verde, Figura 2). Es una aproximación heurística, no un segmentador clínico de precisión, pero suficiente para el propósito comparativo de este trabajo.

- Amplitud pseudo-P y pseudo-T: valor máximo de la señal en dos ventanas fijas relativas al pico R (100 a 40 muestras antes, ≈278–111 ms, región azul en Figura 2; y 40 a 100 muestras después, ≈111–278 ms, región naranja). Es una aproximación posicional, no una detección explícita de onda P/T, adecuada para el propósito comparativo de este estudio.

2.5.2. Características temporales (intervalos RR)

Como se introdujo en la sección 2.1, la irregularidad del intervalo RR es uno de los indicadores más informativos de arritmia: por ejemplo, una contracción ventricular prematura (PVC) típicamente presenta un intervalo RR corto seguido de una pausa compensatoria larga, mientras que la fibrilación auricular se caracteriza por una alta variabilidad RR sostenida. Para capturar este comportamiento se calcularon, para cada latido i:

- RR previo ($\mathrm{RR}_{\mathrm{prev}} = t_i - t_{i-1}$)y RR siguiente ($\mathrm{RR}_{\mathrm{next}} = t_{i+1} - t_i$), en milisegundos.
- Media y desviación estándar del RR sobre una ventana deslizante de los 10 latidos previos, capturando la tendencia y variabilidad local del ritmo más allá del par de intervalos inmediatos.

2.5.3. Estadísticas de forma de onda

Se incluyeron cuatro descriptores globales de la ventana completa del latido (200 muestras: 100 antes y 100 después del pico R): media, desviación estándar, mínimo y máximo de la amplitud, que resumen la distribución general de la señal sin depender de la localización exacta de picos individuales, complementando así a las características morfológicas puntuales.

2.5.4. Características espectrales (wavelet)

Las características anteriores describen el dominio del tiempo, pero muchas alteraciones patológicas se manifiestan como cambios de frecuencia (ej. componentes de alta frecuencia en QRS anormales). La Transformada Wavelet Discreta (DWT) descompone la señal en bandas de frecuencia sin perder localización temporal, una ventaja sobre Fourier para señales no estacionarias como el ECG (Addison, 2005).

Se aplicó una descomposición wavelet de 5 niveles con la familia Daubechies db4 (pywt.wavedec), cuya forma asimétrica se asemeja a la morfología del complejo QRS, lo que la hace ampliamente utilizada en el procesamiento de señales ECG. La descomposición a 5 niveles sobre una señal muestreada a 360 Hz produce 6 conjuntos de coeficientes, cada uno correspondiente a una banda de frecuencia específica (Tabla 1):

Tabla 1. *Bandas de frecuencia de la descomposición wavelet (fs = 360 Hz, db4, 5 niveles).*

Coeficiente	Tipo	Banda de frecuencia aproximada
cA5	Aproximación	0 – 5.6 Hz
cD5	Detalle	5.6 – 11.25 Hz
cD4	Detalle	11.25 – 22.5 Hz
cD3	Detalle	22.5 – 45 Hz
cD2	Detalle	45 – 90 Hz
cD1	Detalle	90 – 180 Hz

Fuente: elaboración propia, calculado por bisección diádica del ancho de banda de Nyquist (180 Hz).

Para cada uno de los 6 conjuntos de coeficientes se calcularon tres estadísticos: media, desviación estándar y energía (suma de los coeficientes al cuadrado), resultando en 18 características espectrales:

$$E_k = \sum c_k[n]^2 \quad (2)$$

donde c_k es el conjunto de coeficientes del nivel k. La energía resume la magnitud de actividad de la señal en esa banda de frecuencia, mientras que la media y desviación estándar capturan el sesgo y dispersión de los coeficientes.

La Tabla 2 resume el catálogo completo de las 30 características extraídas.

Tabla 2. *Catálogo de características extraídas por latido.*

Categoría	Característica	Descripción
Morfológica	r_amplitude	Amplitud del pico R
Morfológica	qrs_width_ms	Ancho del complejo QRS (ms)
Morfológica	p_amplitude	Amplitud máxima en ventana pre-QRS (proxy onda P)
Morfológica	t_amplitude	Amplitud máxima en ventana post-QRS (proxy onda T)
Temporal	rr_prev	Intervalo RR previo (ms)
Temporal	rr_next	Intervalo RR siguiente (ms)
Temporal	rr_mean_window	Media de RR, ventana de 10 latidos previos
Temporal	rr_std_window	Desviación estándar de RR, ventana de 10 latidos previos
Forma de onda	beat_mean, beat_std, beat_min, beat_max	Estadísticos globales de la ventana del latido
Espectral (wavelet)	wav0-wav5_mean, wav0-wav5_std, wav0-wav5_energy	Media, desviación estándar y energía de cada nivel de descomposición (Tabla 1)

Fuente: elaboración propia.

2.6. Clasificador base: Random Forest

Se eligió un Random Forest (Breiman, 2001) como clasificador para las tres estrategias comparadas (baseline, PCA, GA), por tres razones: (1) es robusto a escalas muy distintas (milisegundos, milivoltios, energía wavelet) sin normalización previa; (2) captura relaciones no lineales sin especificarlas manualmente, relevante porque ancho de QRS y tipo de arritmia no se relacionan linealmente; y (3) al ser un ensamble de árboles, es rápido de entrenar e inferir, relevante para el tiempo de inferencia que este trabajo reporta.

Se configuraron 200 árboles y class_weight="balanced", que reescala el peso de cada clase inversamente a su frecuencia: sin este ajuste, el clasificador favorecería la clase mayoritaria (Normal) a costa de no detectar la clase de interés clínico. Implementación de scikit-learn (Pedregosa et al., 2011).

2.7. Validación cruzada agrupada por paciente

La forma más común de validar es k-fold cross-validation aleatorio. Pero dividir a nivel de latido individual produce un problema documentado en clasificación de ECG (de Chazal *et al.,* 2004): el ECG de un paciente es muy característico de él (morfología, ritmo, ruido de fondo), así que si dos latidos del mismo paciente caen uno en entrenamiento y otro en prueba, el clasificador puede reconocer al "paciente" en vez de la arritmia.

Esto infla las métricas artificialmente, sin sostenerse frente a un paciente nunca visto, el escenario real de uso clínico.

Para evitarlo, se usó StratifiedGroupKFold (5 particiones), agrupando por paciente: todos los latidos de un paciente quedan en una sola partición, mientras se mantiene balanceada la proporción de clases entre particiones. Se aplicó igual a los tres métodos, para que las diferencias reportadas reflejen la calidad real de cada estrategia y no un artefacto de la partición.

2.8. Reducción con Análisis de Componentes Principales

Como primer método de comparación se aplicó PCA, que no elige un subconjunto sino que combina linealmente las características en "componentes" ordenados por varianza capturada. Antes, las 30 características se estandarizaron a media cero y varianza unitaria: sin esto, una característica de valores grandes (energía wavelet) dominaría artificialmente el cálculo frente a una de valores pequeños (amplitud en mV), aunque igual de informativas. Se retuvieron los componentes mínimos que explican 95% de la varianza: 15 de 30 originales.

PCA es no supervisada: no considera la etiqueta (Normal/Anormal), solo la varianza estadística. Puede retener componentes de alta varianza pero poco discriminativos, y descartar variación pequeña pero muy útil. Esta limitación motiva compararla contra un método que sí usa la etiqueta directamente: el Algoritmo Genético.

2.9. Selección de características con Algoritmo Genético

A diferencia de PCA, un GA formula la selección como búsqueda guiada directamente por el desempeño del clasificador, no por criterios estadísticos indirectos: evoluciona una población de subconjuntos a lo largo de generaciones, favoreciendo a los que clasifican mejor, hasta converger en un subconjunto de alta calidad. Se implementó con la librería deap (Fortin *et al.,* 2012).

Representación: cada solución candidata ("individuo") es un cromosoma binario de 30 genes, uno por característica original, donde 1 significa que esa característica está incluida en el subconjunto y 0 que está excluida.

Función de aptitud (fitness): se entrena el clasificador solo con las características activas del individuo y se mide su F1 macro con CV rápido (3 particiones, agrupado por paciente para no reintroducir el problema de 2.7). A ese valor se resta una penalización proporcional al número de características activas:

$$\text{fitness} = \text{F1}_{\text{macro}}(CV) - \lambda \cdot n_{\text{características}} \quad (3)$$

Sin esta penalización, el GA no tendría incentivo para reducir características (dos subconjuntos con igual desempeño obtendrían el mismo fitness aunque uno tenga 10 y otro 25), y el objetivo es encontrar subconjuntos *compactos* de alto desempeño. El valor de $\lambda = 0.001$ se eligió para que la penalización sea marginal frente a las diferencias típicas de F1, pero suficiente para romper empates a favor de subconjuntos más pequeños.

Operadores evolutivos: en cada generación se seleccionan "padres" por torneo (grupos de 3 individuos al azar, gana el de mejor fitness), favoreciendo a los mejores sin perder diversidad. Con ellos se generan nuevos individuos por cruza de dos puntos (probabilidad 0.6): se intercambian genes entre dos posiciones al azar del cromosoma. Por último, mutación bit-flip (probabilidad 0.2 por individuo, 0.02 por gen) introduce variación aleatoria para evitar un óptimo local.

Hiperparámetros: población de 40 individuos durante 40 generaciones, balance entre calidad de búsqueda y tiempo de cómputo (cada evaluación requiere entrenar un Random Forest). El mejor individuo se conserva (elitismo) y se reevalúa al final con el protocolo completo (5 particiones, sección 2.7), igual que baseline y PCA, para una comparación estrictamente equivalente.

Resultados

3.1. Comparación global de métodos

La Tabla 3 resume el desempeño de los tres métodos evaluados, conjunto completo de características (baseline), reducción por PCA y selección por Algoritmo Genético (GA), bajo el protocolo de validación

cruzada agrupada por paciente descrito en la sección 2.7, sobre los 109,398 latidos extraídos de los 48 registros de la base de datos.

Tabla 3. *Comparación de desempeño entre baseline, PCA y GA (media ± desviación estándar sobre 5 particiones agrupadas por paciente).*

Método	N° Caract.	Exactitud	F1 macro	Sensibilidad	Especificidad	Precisión	Inferencia (ms/latido)
Baseline	30	0.902 ±0.032	0.821 ±0.060	0.689 ±0.137	0.946 ±0.038	0.846 ±0.072	0.00392 ±0.00012
PCA (95% var)	15	0.885 ±0.034	0.798 ±0.043	0.689 ±0.191	0.927 ±0.064	0.833 ±0.066	0.00438 ±0.00017
GA	**11**	**0.916 ±0.020**	**0.844 ±0.041**	**0.711 ±0.152**	**0.960 ±0.030**	**0.878 ±0.049**	**0.00305 ±0.00008**

Fuente: elaboración propia.

El subconjunto del GA (11 de 30 características, 63% de reducción) superó a baseline y PCA en las siete métricas simultáneamente: al ser el único método que usa la etiqueta directamente en su búsqueda (ecuación 3, sección 2.9), optimiza explícitamente lo que los otros logran solo indirectamente. PCA, pese a menor dimensión que el baseline (15 vs. 30), obtuvo el peor desempeño de los tres, consistente con la sección 2.8: al priorizar varianza sin garantía de relevancia para distinguir Normal de Anormal, sacrificó información discriminativa que sí conservan las características crudas.

Un segundo hallazgo es la desviación estándar: el GA tuvo también la menor variabilidad entre particiones (accuracy_std = 0.020 vs. 0.032–0.034). Como cada partición es un grupo distinto de pacientes (sección 2.7), esto indica que el subconjunto del GA generaliza de forma más consistente, no solo que "acierta más en promedio".

3.2. Efecto de la validación entre pacientes

Antes de fijar el protocolo final (sección 2.7), se evaluó el mismo modelo baseline con una validación cruzada estratificada sin agrupar por paciente (StratifiedKFold estándar, división aleatoria a nivel de latido). Los resultados de esa validación incorrecta fueron:

- Exactitud: 99.1% (± 0.01%)
- F1 macro: 0.984 (± 0.0003)

frente a los valores reales, ya reportados en la Tabla 3, de 90.2% exactitud y 0.821 F1 macro con StratifiedGroupKFold. La diferencia, casi 9 puntos porcentuales de exactitud y una caída de desviación estándar prácticamente nula a un rango realista, es la evidencia empírica directa del efecto descrito en la sección 2.7, ya documentado en la literatura para MIT-BIH (de Chazal *et al.,* 2004). Todos los resultados reportados en este trabajo (Tabla 3) corresponden exclusivamente al protocolo correcto, agrupado por paciente.

3.3. Matriz de confusión y análisis de errores

La Figura 3 muestra la matriz de confusión agregada (predicciones fuera-de-muestra de las 5 particiones, sección 2.7) para el método GA, el de mejor desempeño global, normalizada por fila (porcentaje sobre la clase real) para que el fuerte desbalance de clases (Normal supera ~5:1 a Anormal) no oculte visualmente el desempeño relativo de la clase minoritaria.

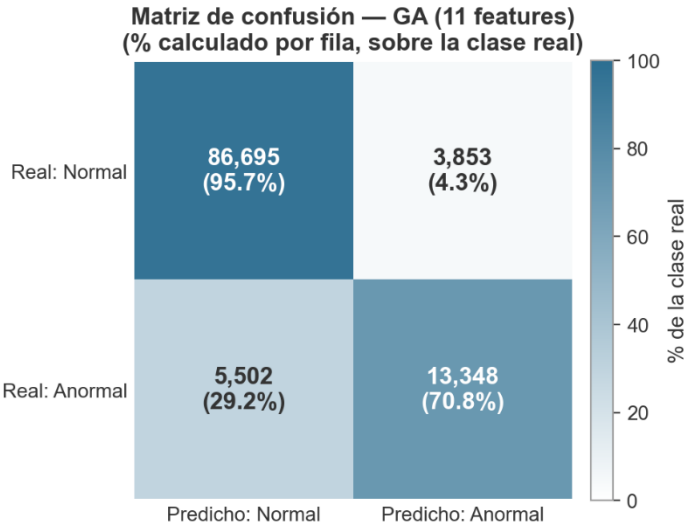


Figura 3. Matriz de confusión agregada, normalizada por fila, método GA (n = 109,398 latidos).
Fuente: elaboración propia.

De los 18,850 latidos anormales reales, el modelo detectó 13,348 (sensibilidad 71.1%) y no detectó 5,502. Esta es la principal limitación actual: un falso negativo (arritmia real clasificada como normal) es clínicamente más grave que un falso positivo (revisión manual innecesaria en el peor caso). Aunque el GA mejoró la sensibilidad frente al baseline (71.1% vs. 68.9%) y redujo los falsos positivos (3,853 vs. 5,023), el nivel de falsos negativos remanente indicaría, para uso clínico real, ajustar el punto de operación (ej. un umbral que priorice sensibilidad), fuera del alcance de este trabajo.

3.4. Convergencia del Algoritmo Genético

La Figura 4 muestra la evolución del fitness (ecuación 3, sección 2.9) a lo largo de las 40 generaciones evaluadas, para el mejor individuo de cada generación (línea sólida) y el promedio de la población completa (línea discontinua).

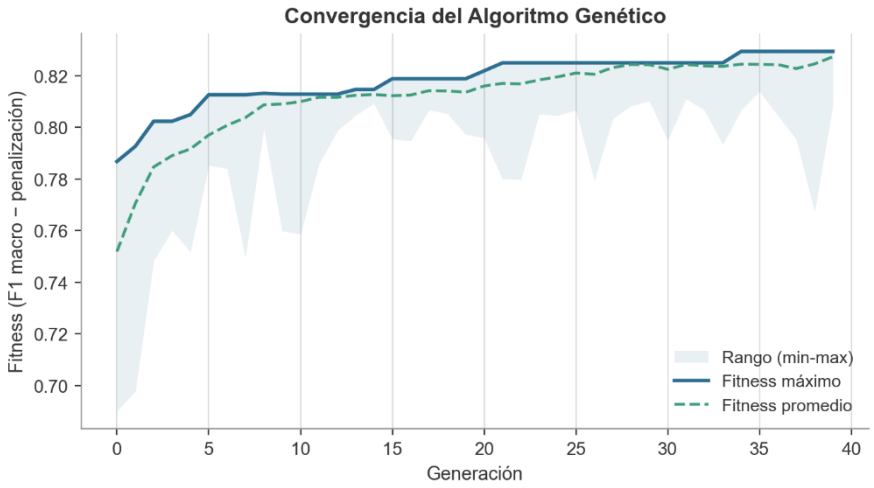


Figura 4. Curva de convergencia del Algoritmo Genético.
Fuente: elaboración propia.

El fitness máximo pasó de 0.786 en la generación 0 a 0.830 en la generación 34, donde se estabiliza (34–39 sin mejora adicional), indicando que 40 generaciones bastaron para converger. El fitness promedio sigue una tendencia similar pero con oscilaciones esperables, dado que la mutación (0.02 por gen) reintroduce variabilidad incluso tras estabilizarse el mejor individuo, y se mantiene por debajo del máximo: el algoritmo conserva diversidad en vez de colapsar prematuramente.

3.5. Características seleccionadas por el Algoritmo Genético

El subconjunto final de 11 características elegido por el GA fue: qrs_width_ms, rr_prev, rr_next, rr_mean_window, rr_std_window, beat_min, wav0_std, wav0_energy, wav2_mean, wav3_mean y wav5_energy (ver Tabla 2, sección 2.5, para la definición de cada una).

El subconjunto no se concentra en una sola categoría (sección 2.5), sino que combina las cuatro: morfológica (ancho de QRS), temporal (los cuatro descriptores de RR), forma de onda (amplitud mínima) y espectral (energía/desviación de cA5, medias de cD4 y cD3, energía de cD1, Tabla 1). Esto sugiere que la información discriminativa está distribuida entre representaciones, y que un método que descarte una categoría a priori, como PCA al proyectar sobre varianza global, arriesga perder señal que el GA sí retiene al evaluarla contra el desempeño real.

Es esperable que las cuatro características de RR se hayan conservado casi por completo: su irregularidad es uno de los indicadores más directos de arritmia (sección 2.5.2), y su selección consistente por el GA es coherente con ese fundamento fisiológico.

3.6. Tiempo de inferencia

El GA fue el más rápido en inferencia (0.00305 ms/latido, Tabla 3): 22% más rápido que baseline (0.00392) y 30% más que PCA (0.00438). PCA, pese a menor dimensión que baseline, resultó más lento porque el tiempo incluye transformar las 30 características al espacio de componentes (multiplicación por la matriz de proyección), costo que el GA no tiene al solo descartar características no seleccionadas. Para monitoreo en tiempo real o dispositivos embebidos, esto hace del GA la opción más eficiente además de la más precisa.

Conclusiones y Trabajo futuro

Este trabajo comparó tres estrategias (conjunto completo, PCA, GA) para clasificación binaria de arritmias sobre MIT-BIH, con validación cruzada agrupada por paciente (sección 2.7). El subconjunto del GA (11 características, 63% de reducción) superó a conjunto completo y PCA en las siete métricas: exactitud (91.6% vs. 90.2% y 88.5%), F1 macro (0.844 vs. 0.821 y 0.798), sensibilidad, especificidad, precisión e inferencia (22% más rápido que baseline). La selección dirigida por desempeño, en vez de por varianza, produce subconjuntos más compactos y generalizables entre pacientes.

Un segundo aporte es la evidencia empírica del efecto de fuga de información entre pacientes (sección 3.2): el mismo baseline pasó de 99.1% de exactitud con validación aleatoria a nivel de latido, a 90.2% con el protocolo correcto agrupado por paciente. Esta diferencia de casi 9 puntos confirma, con datos propios, un problema documentado desde de Chazal et al. (2004) y subraya la importancia de este protocolo para trabajo futuro en esta base de datos.

Como limitación principal, la sensibilidad del mejor modelo (71.1%) permanece moderada para uso clínico: ~3 de cada 10 latidos anormales no fueron detectados (sección 3.3). Como trabajo futuro: extender a 5 clases AAMI en vez de binario; explorar funciones de aptitud que penalicen explícitamente falsos negativos, no solo F1 macro; y validar el subconjunto seleccionado sobre bases de datos adicionales de ECG.

Referencias

Association for the Advancement of Medical Instrumentation (AAMI). (2012). ANSI/AAMI EC57:2012 — Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. Association for the Advancement of Medical Instrumentation.

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

Breiman, L. (2001). Random forests. *Machine Learning,* 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

de Chazal, P., O'Dwyer, M., y Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering,* 51(7), 1196–1206. 10.1109/TBME.2004.827359

Fortin, F. A., De Rainville, F. M., Gardner, M. A., Parizeau, M., y Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research,* 13, 2171–2175. https://www.jmlr.org/papers/volume13/fortin12a/fortin12a.pdf

Moody, G. B., y Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine,* 20(3), 45–50. 10.1109/51.932724

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research,* 12, 2825–2830. https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Van Alste, J. A., & Schilder, T. S. (1985). Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. *IEEE Transactions on Biomedical Engineering,* BME-32(12), 1052–1060. 10.1109/TBME.1985.325514