



ANALISIS DE LA REDUCCIÓN DE DIMENSIONALIDAD DE SEÑALES DE ECG PARA EL ANÁLISIS MORFOLÓGICO DEL RITMO CARDÍACO





PROBLEMA

El problema radica no solo en identificar la presencia de una anomalía, sino en determinar qué segmento de la señal la origina, cómo puede explicarse desde un punto de vista clínico y si es posible reconstruir la señal a partir de su representación comprimida.





DATASET MIT – BIH





Índice

01 Contexto

02 Metadatos

03 Columnas

Contexto

El dataset en cuestión es una colección de 48 extractos de media hora de grabaciones de ECG ambulatorio de dos canales, obtenidos de 47 sujetos. Estas grabaciones fueron recolectadas por el Laboratorio de Arritmias BIH entre 1975 y 1979.

23 grabaciones fueron escogidas al azar de un conjunto más amplio de 4000 registros, representando una población mixta de pacientes hospitalizados 60% y ambulatorios 40%.

Datos en Crudo

mit-bih

> mitdbdir

> x_mitdb

100.atr

100.dat

100.he

100.xws

101.atr

101.dat

101.he

101.xws

102-0.atr

102.atr

102.dat

102.he

102.xws

103.atr

103.dat

EXPLORADOR

PARCIAL1

> .qodo

> AED_images

> dashboard

> DataWranglin

mit-bih

> mitdbdir

> x_mitdb

100.atr

100.dat

100 hea

100.he

mit-bih > 100.he

You, hace 11 minutos | 1 author (You)

1 100 2 360 650000

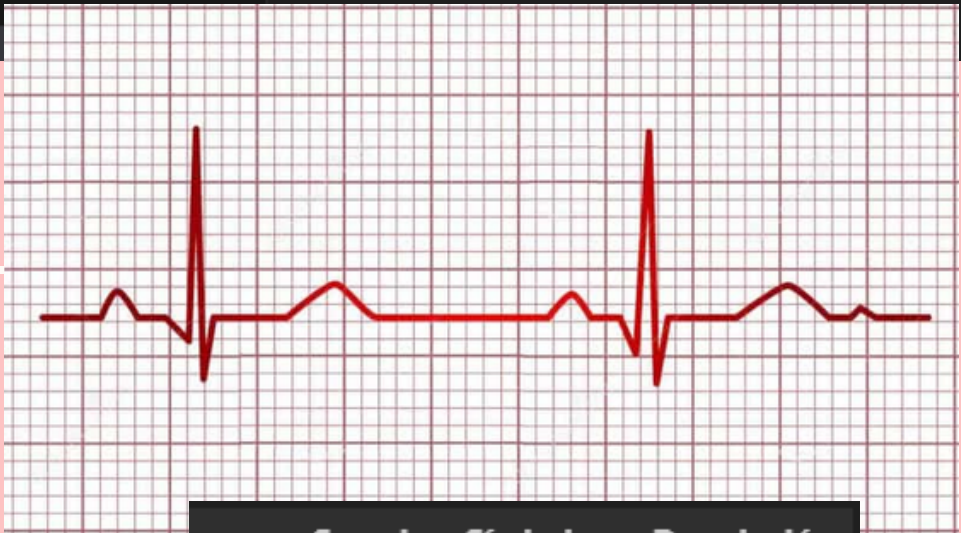
2 100.dat 212 200 11 1024 995 -22131 0 MLII

3 100.dat 212 200 11 1024 1011 20052 0 V5

4 # 69 M 1085 1629 x1

5 # Aldomet, Inderal

6



	Sample	Símbolo	Descripción
0	18	+	Desconocido
1	77	N	Latido normal
2	370	N	Latido normal
3	662	N	Latido normal
4	946	N	Latido normal
...
2269	648978	N	Latido normal
2270	649232	N	Latido normal
2271	649484	N	Latido normal
2272	649734	N	Latido normal
2273	649991	N	Latido normal

```
Registros encontrados: ['100', '101', '102']
Procesando registro: 100
Procesando registro: 101
Procesando registro: 102
Procesando registro: 103
Procesando registro: 104
Procesando registro: 105
Procesando registro: 106
Procesando registro: 107
Procesando registro: 108
Procesando registro: 109
Procesando registro: 111
Procesando registro: 112
Procesando registro: 113
Procesando registro: 114
Procesando registro: 115
Procesando registro: 116
Procesando registro: 117
Procesando registro: 118
Procesando registro: 119
Procesando registro: 121
Procesando registro: 122
Procesando registro: 123
Procesando registro: 124
Procesando registro: 200
...
Procesando registro: 232
Procesando registro: 233
Procesando registro: 234
¡Dataset final creado con éxito!
```

Datos

df_final

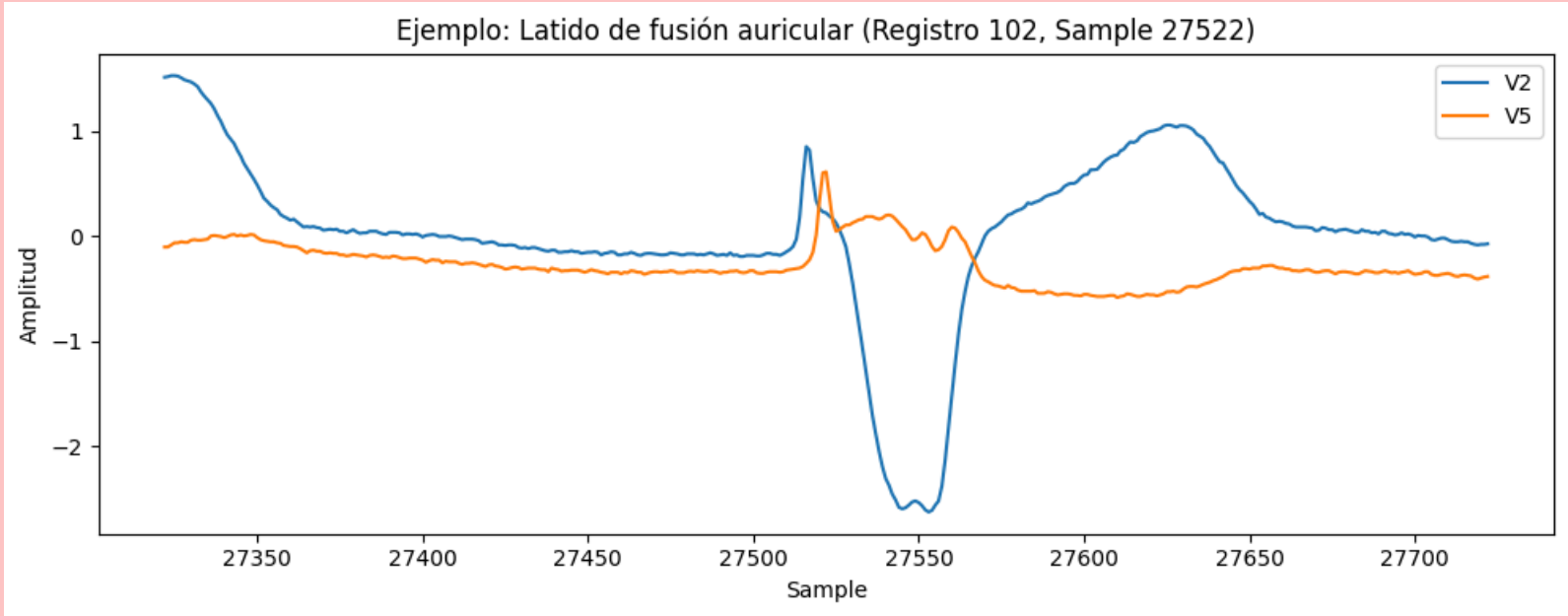
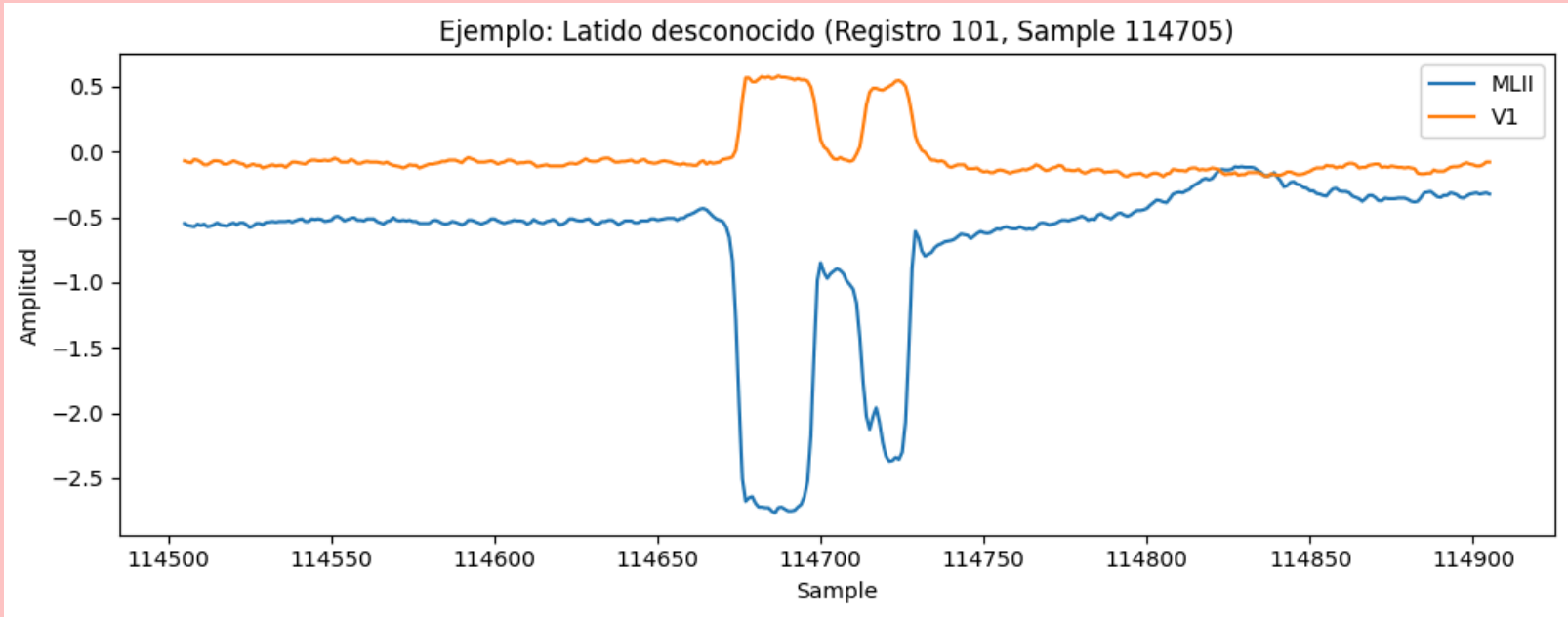
	MLII	V5	Sample	Símbolo	Descripción	Registro	V1	V2	V4
0	-0.145	-0.065	0	NaN	NaN	100	NaN	NaN	NaN
1	-0.145	-0.065	1	NaN	NaN	100	NaN	NaN	NaN
2	-0.145	-0.065	2	NaN	NaN	100	NaN	NaN	NaN
3	-0.145	-0.065	3	NaN	NaN	100	NaN	NaN	NaN
4	-0.145	-0.065	4	NaN	NaN	100	NaN	NaN	NaN
...
31199995	-0.390	NaN	649995	NaN	NaN	234	0.070	NaN	NaN
31199996	-0.390	NaN	649996	NaN	NaN	234	0.070	NaN	NaN
31199997	-0.395	NaN	649997	NaN	NaN	234	0.075	NaN	NaN
31199998	-0.380	NaN	649998	NaN	NaN	234	0.080	NaN	NaN
31199999	0.000	NaN	649999	NaN	NaN	234	0.000	NaN	NaN

31200000 rows × 9 columns

Columna	Tipo de Dato	Significado
MLII / V1 / V2 / V4 / V5	Numérico (float)	Valor de voltaje (mV) de la señal ECG en un instante específico para la derivación correspondiente.
Sample	Numérico (entero/float)	Índice o número secuencial de la muestra de ECG, indicando la posición temporal. No tiene unidad física directa, sino que se convierte a tiempo usando la frecuencia de muestreo.
Símbolo	Categorico (cadena de texto)	Código de anotación para el tipo de latido cardíaco o evento en ese punto de la señal (ej., 'N' para normal, 'V' para ventricular prematuro). Puede ser nulo si no hay anotación.
Descripción	Categorico (cadena de texto)	Descripción textual detallada del 'Símbolo' de anotación. Puede ser nulo.
Registro	Numérico / Categorico (entero/cadena)	Identificador único del sujeto o de la grabación de ECG a la que pertenece la muestra.

Tabla de Símbolos

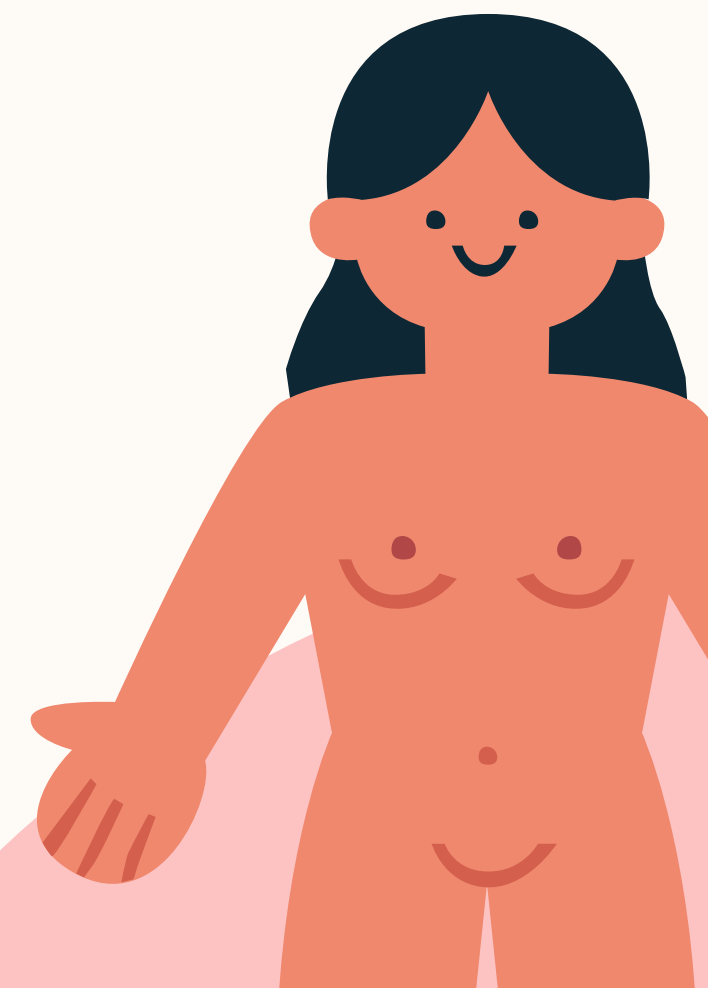
Símbolo	Descripción
N	Latido normal.
L	Latido de bloqueo de rama izquierda.
R	Latido de bloqueo de rama derecha.
V	Latido ventricular prematuro.
A	Latido auricular prematuro.
F	Latido de fusión.
/	Latido de fusión ventricular.
f	Latido de fusión auricular.
j	Latido de escape nodal.
E	Latido de escape ventricular.
a	Latido auricular aberrante.
J	Latido de escape de la unión.
S	Latido de marcapasos.
e	Latido ventricular aberrante.
Q	Latido desconocido.
+	Latido desconocido.





ECGLENS: INTERACTIVE VISUAL EXPLORATION OF LARGE SCALE ECG DATA FOR ARRHYTHMIA DETECTION

Ke Xu , Shunan Guo





Índice

01 Introducción

02 Dataset

03 Metodologia

04 Evauación

01

Problema

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed semper consectetur porta. Curabitur malesuada sodales urna at molestie. Nulla faucibus, odio at porta mollis, tortor leo mollis lacus, at imperdiet diam leo sed velit. Nulla malesuada sit amet ex eget venenatis. Maecenas blandit ante mauris, sed hendrerit nibh scelerisque ut. Sed tincidunt est eu lobortis convallis.

02 Dataset

MIT-BIH Arrhythmia Database

El estudio utiliza el conjunto de datos MIT-BIH Arrhythmia Database, un repositorio público ampliamente reconocido que contiene registros de electrocardiogramas (ECG) anotados por cardiólogos. Este dataset incluye una gran cantidad de latidos cardíacos etiquetados, lo que permite entrenar y validar modelos de clasificación automática. Los registros de ECG son de larga duración, típicamente de 24 horas, lo que genera alrededor de 86,000 latidos por paciente. Esto proporciona una base sólida para el análisis de arritmias, incluyendo latidos normales (Sinus) y anómalos (VEB, SVEB, Fusion).

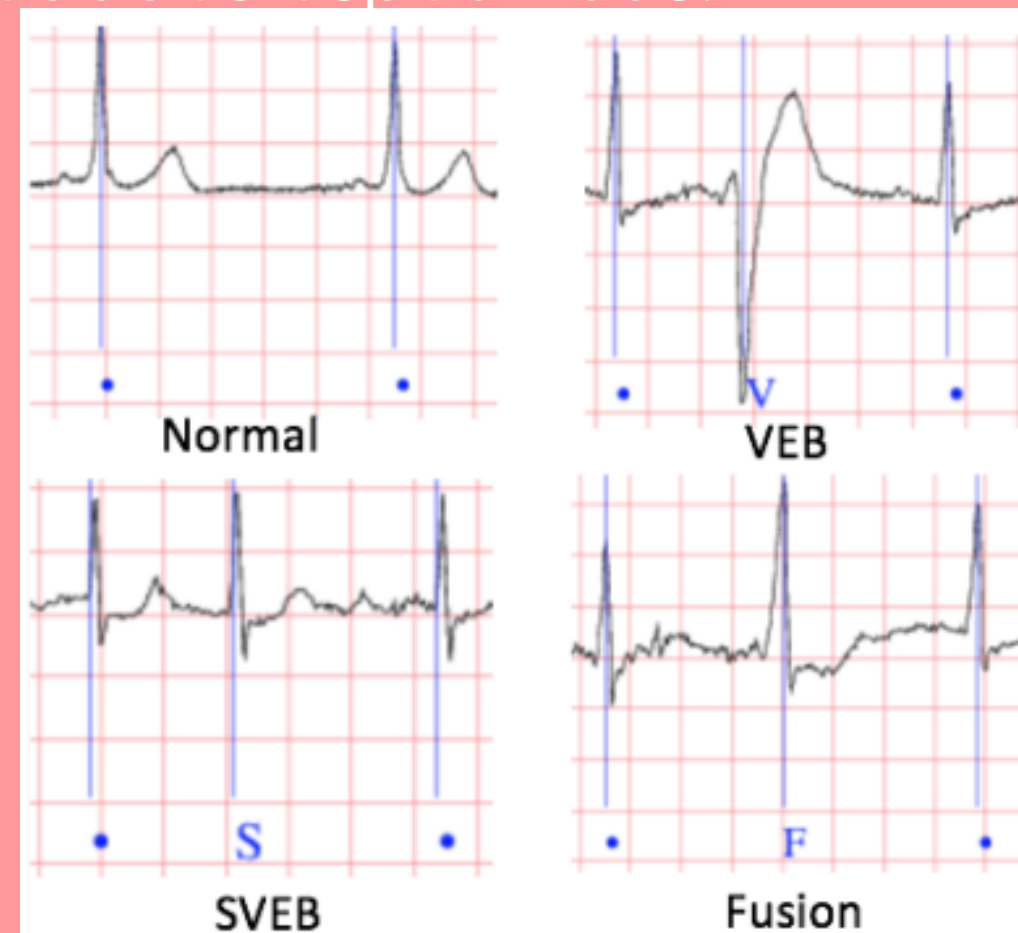


03 Metodología

- Se emplea una Red Neuronal Convolutiva (CNN) multicapa para clasificar los latidos en cuatro categorías clínicas (Sinus, VEB, SVEB, Fusion).
- El modelo consta de tres etapas:
 - a. Transformación: Incluye mapeo de identidad, transformación espectral (dominio temporal) y submuestreo para reducir ruido.
 - b. Capa de conexión: Utiliza convoluciones locales con capas ReLU y max-pooling.
 - c. Capa final: Combina convoluciones completas con una capa softmax para la clasificación.
- El modelo alcanza una precisión del 91.59%.

Detección de Anomalías

- Se aplica el algoritmo Local Outlier Factor (LOF) para identificar latidos mal clasificados o potencialmente patológicos.
- También se utiliza un algoritmo de detección de categorías raras para resaltar anomalías correlacionadas con enfermedades específicas.



Visualización e Interacción

- A-glyph: Un glifo novedoso que representa los componentes clave de un latido (onda P, complejo QRS, segmento ST, onda T) mediante elementos visuales intuitivos (círculos, rectángulos y líneas) que indican desviaciones en amplitud y duración.
- Vistas múltiples: Incluyen una línea de tiempo circular para valores QTc, proyección MDS para distribución de latidos, y vistas temporales y estadísticas para análisis detallado.
- Interacciones: Permiten corrección manual de etiquetas, comparación contextual de latidos y exploración guiada por anomalías detectadas.



REQUERIMIENTOS DE DISEÑO

- **R1. Soporte para datos a gran escala.**

El sistema debe estar diseñado para permitir un análisis y exploración efectivos de datos ECG a gran escala, permitiendo a los usuarios examinar resultados capturados durante varias horas.

- **R2. Clasificación precisa de latidos cardíacos.**

Los algoritmos de clasificación de latidos deben integrarse en el sistema para apoyar la detección eficiente de arritmias en grandes volúmenes de datos. El sistema debe proporcionar métodos que ayuden a los usuarios a identificar rápidamente latidos anómalos que puedan requerir una investigación clínica y un eventual diagnóstico, así como posibles errores de clasificación.

- **R3. Reconocimiento y comparación mejorados de latidos.**

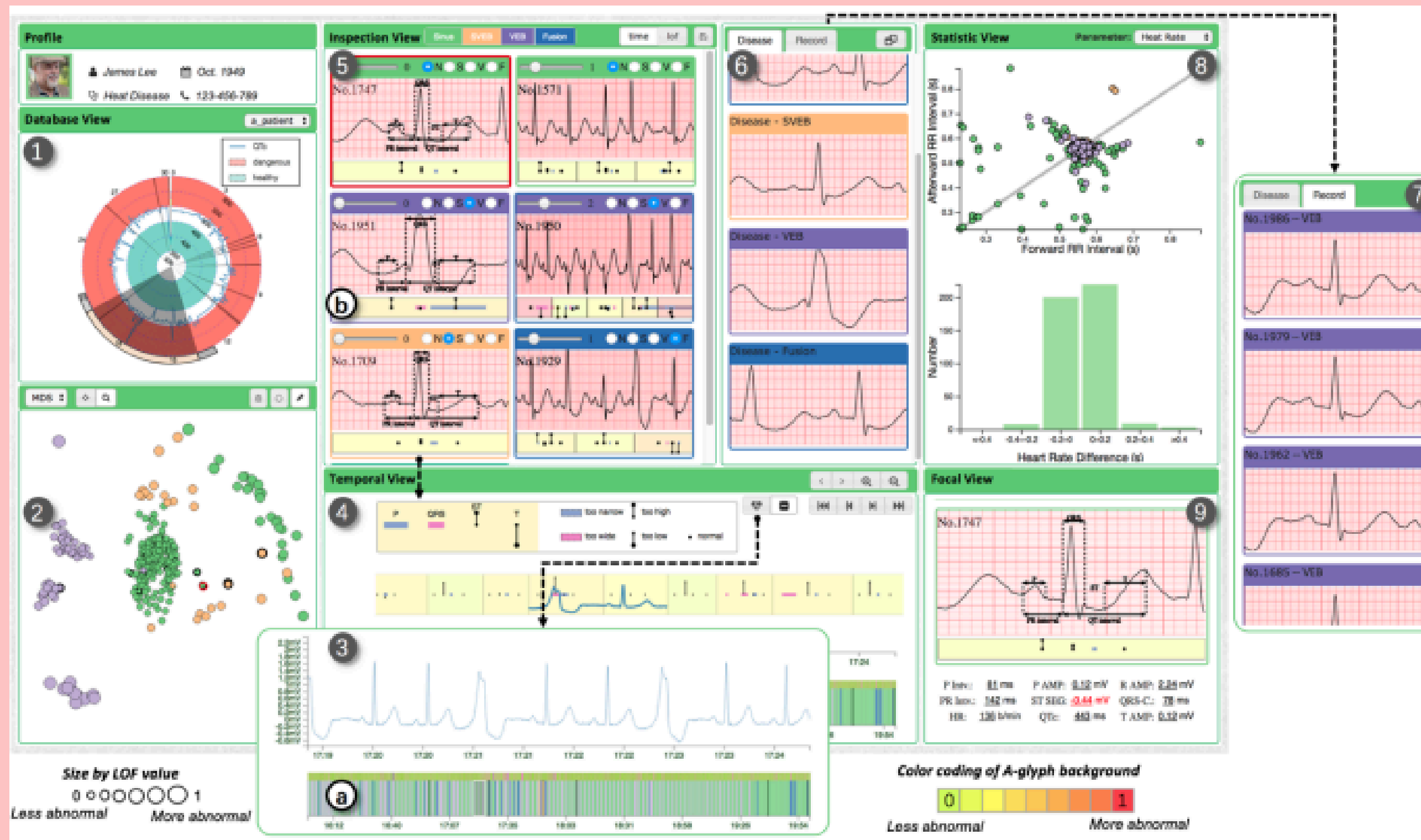
El sistema debe resaltar los parámetros clave de cada latido y facilitar la comparación entre ellos, para ayudar a los cardiólogos a leer más rápidamente las ondas ECG y encontrar patrones que puedan corresponder a enfermedades.

- **R4. Exploración interactiva de datos.**

Para apoyar la exploración y comparación eficiente de grandes cantidades de datos ECG, es necesario incorporar interacciones flexibles que ayuden a los usuarios a navegar rápidamente entre una gran cantidad de latidos y características asociadas.



04 Evaluación



04 Evaluación

Estudio con usuarios: 18 cardiólogos evaluaron la eficacia de A-glyph frente a representaciones tradicionales. Los resultados mostraron que A-glyph reduce el tiempo de detección de arritmias y mejora la precisión en comparación con glifos convencionales.

Entrevistas con expertos: Dos cardiólogos validaron la utilidad del sistema en casos reales, destacando su capacidad para identificar patrones complejos y reducir el tiempo de análisis.

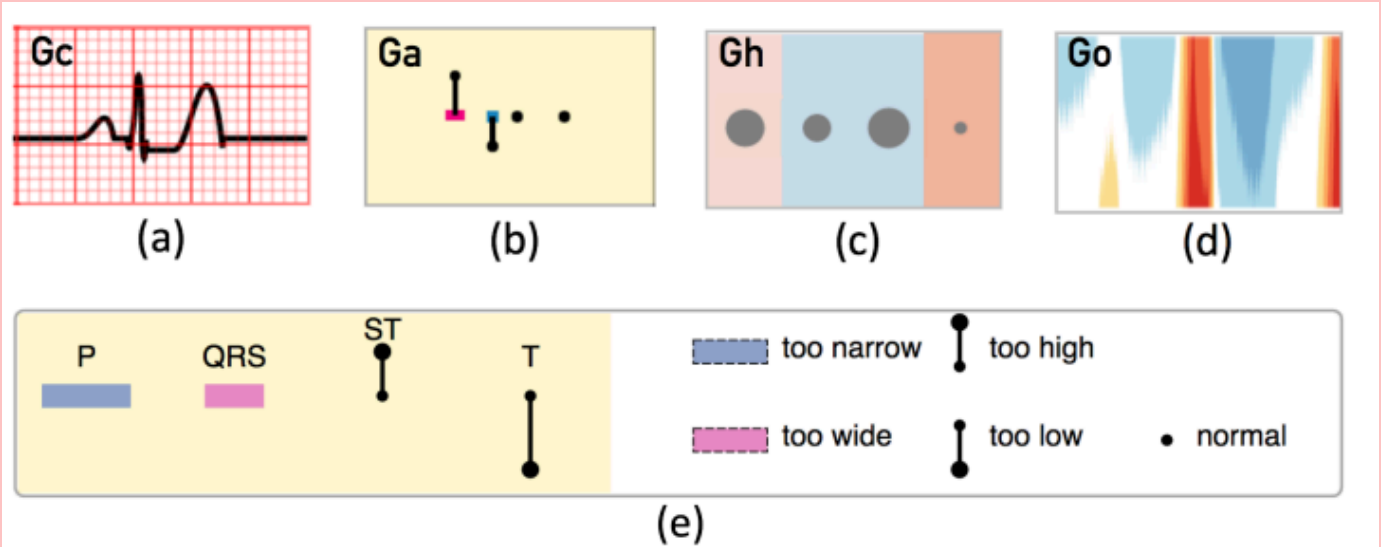
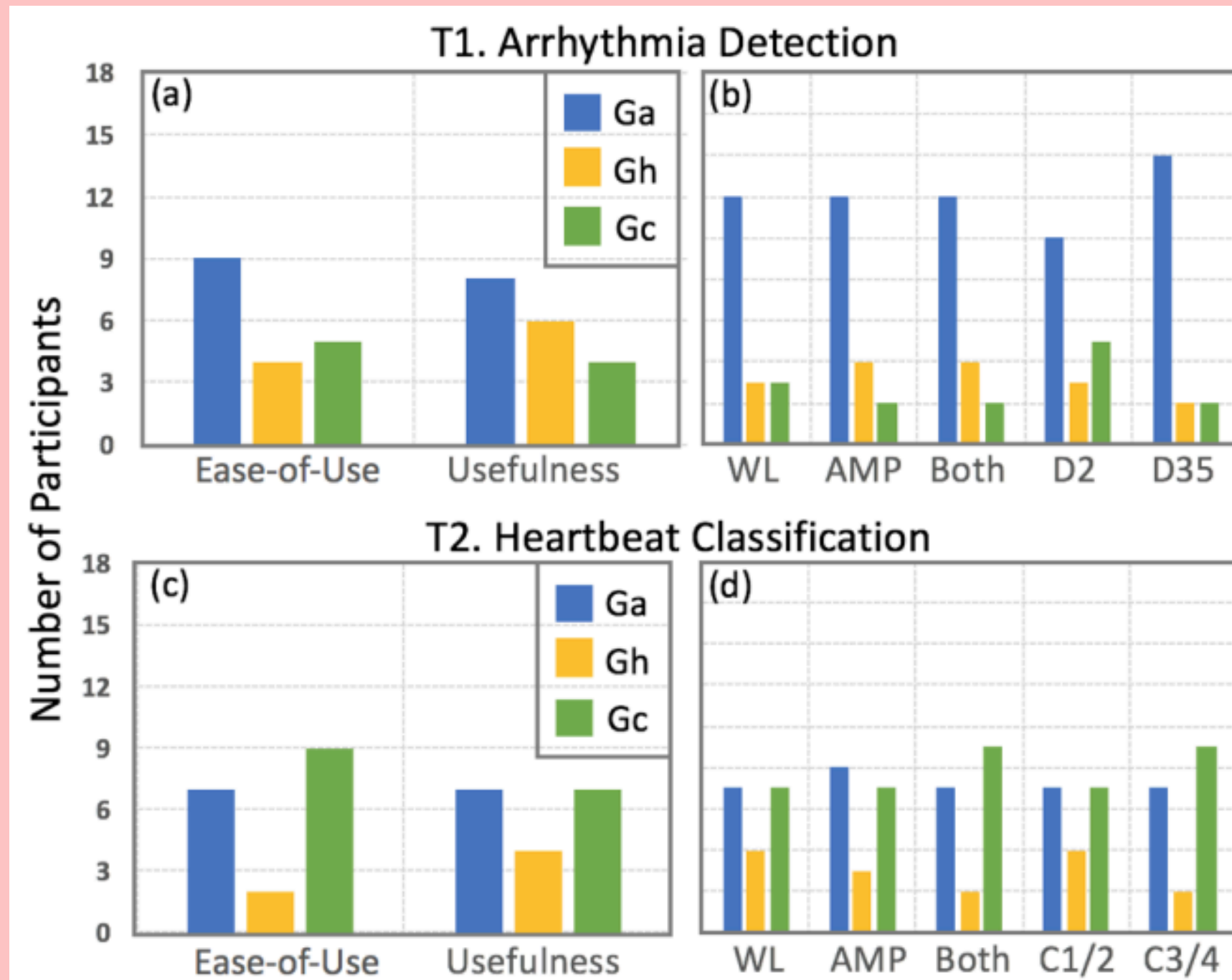


Figure 4. ECG visualization designs: (a) conventional ECG (G_c), (b) A-glyph (G_a), (c) heatmap-based glyph (G_h) and (d) horizon chart (G_o). The encoding principles of A-glyph are illustrated in (e).

	18	Users
×	3	Glyph designs
×	2	Successive Tasks (T1 & T2)
×	2	Data sizes (small(2), large(35))
×	4	Number of Abnormal components (1, 2, 3, 4)
×	3	Anomaly Types (wavelength, amplitude, both)
×	3	Repetitions
7776		Trials

Table 1. Testing Conditions.

04 Evaluación



Calificaciones de los usuarios sobre diferentes glifos en cuanto a su usabilidad en (a) T1 y (c) T2, y su eficiencia bajo diferentes condiciones en (b) T1 y (d) T2.

A cada usuario se le pidió calificar un glifo a la vez. En la figura, el eje y indica la cantidad de calificaciones. WL significa Wavelength (longitud de onda), AMP significa Amplitude (amplitud), D2 y D35 representan diferentes tamaños de datos, y C1/2/3/4 indica el número de componentes anómalos.



**XAI-REDUCT: ACCURACY PRESERVATION DESPITE
DIMENSIONALITY REDUCTION FOR HEART DISEASE
CLASSIFICATION USING EXPLAINABLE AI**

Surajit Das, Mahamuda Sultana
(2023)





Índice

01 Problema

02 Dataset

03 Metodologia

01 Problemática

El objetivo principal del estudio es mejorar la clasificación de enfermedades cardíacas utilizando modelos de aprendizaje automático (ML) interpretables, preservando la precisión incluso después de reducir la dimensionalidad del conjunto de características.

Los autores utilizan técnicas de XAI para explicar las decisiones del modelo y seleccionar un subconjunto de características más relevantes.

02 Dataset

Heart Disease

Este conjunto de datos data de 1988 y consta de cuatro bases de datos: Cleveland, Hungría, Suiza y Long Beach V.

Contiene 76 atributos, incluido el atributo predicho, pero todos los experimentos publicados utilizan un subconjunto de 14 de ellos.

El campo "objetivo" se refiere a la presencia de cardiopatía en el paciente.

Su valor es entero: 0 = sin cardiopatía y 1 = con cardiopatía.

1. age
2. sex
3. chest pain type (4 values)
4. resting blood pressure
5. serum cholestoral in mg/dl
6. fasting blood sugar > 120 mg/dl
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9. exercise induced angina
10. oldpeak = ST depression induced by exercise relative to rest
11. the slope of the peak exercise ST segment
12. number of major vessels (0-3) colored by flourosopy
13. thal: 0 = normal: 1 = fixed defect: 2 = reversable defect

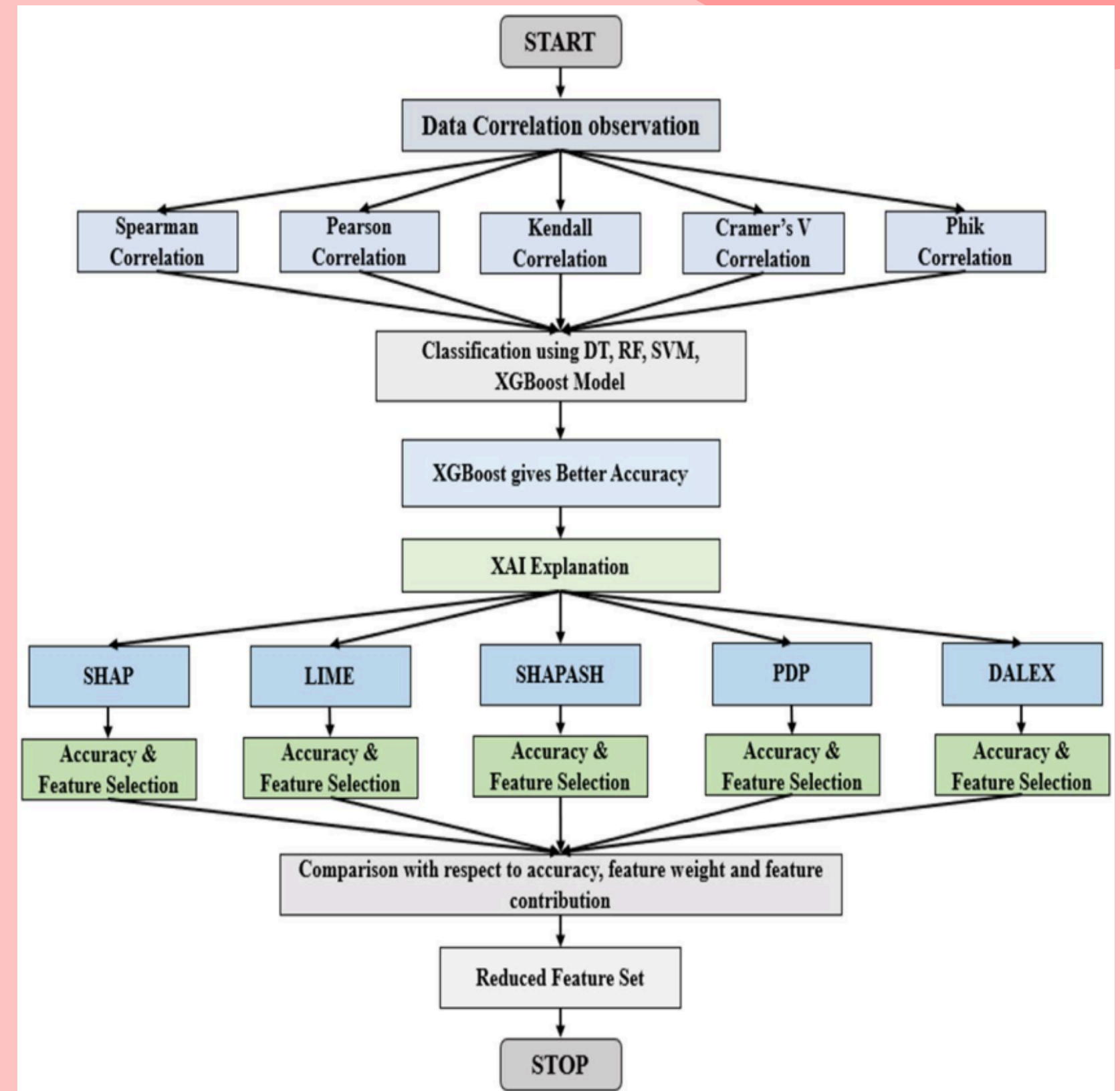
03 Metodología

Uso de XGBoost como modelo de clasificación, que supera en precisión a otros modelos como árboles de decisión, bosques aleatorios y máquinas de vectores de soporte (SVM).

Reducción de dimensionalidad mediante técnicas de XAI (SHAP, LIME, SHAPASH, PDP, DALEX) para seleccionar las características más relevantes sin perder precisión.

Identificación de las cuatro características principales que influyen en el diagnóstico de enfermedades cardíacas:

- **chest_pain_type** (tipo de dolor en el pecho).
- **st_depression** (depresión del segmento ST inducida por ejercicio).
- **thalassemia** (talasemia).
- **num_major_vessels** (número de vasos principales afectados).

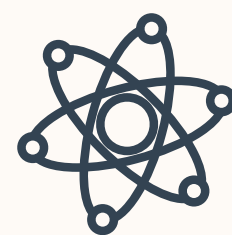
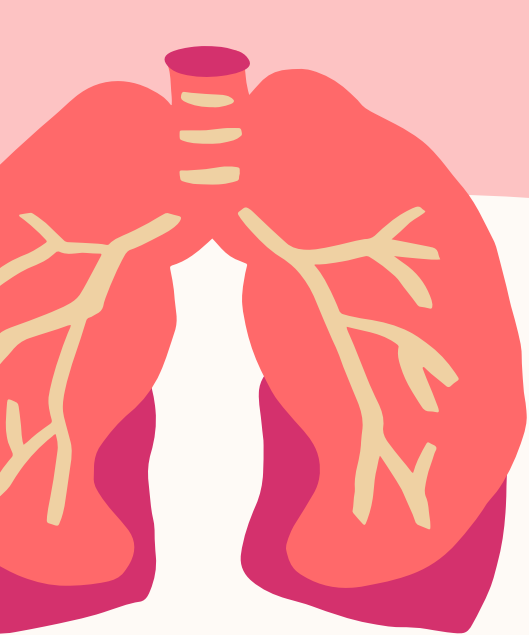


04 Resultados

Table 5 Feature and relevant contributions for the four models post explanation using SHAP							
Model Name	Contribution	Features	Value	Model name	Contribution	Features	Value
Decision tree	+0.529	<BIAS>	1.000	Random forest	+0.530	<BIAS>	1.000
	+0.409	Age	− 1.449		+0.219	Age	− 2.248
	−0.053	Cp	−0.481		+0.040	Fbs	−0.244
					+0.038	Trestbps	0.809
					+0.036	Restecg	0.187
					+0.022	Thalac	−0.781
					+0.022	Ca	−0.042
					+0.021	Oldpeak	0.324
					+0.020	Slope	0.042
					+0.015	Exang	−0.499
					+0.014	Sex	−0.551
					+0.012	Chol	0.275
					+0.007	Thal	−0.090
					+0.716	Cp	2.000
					+0.356	Age	50.000
					+0.272	Thal	2.000
					+0.156	Oldpeak	0.000
Support vec- tor machine	+2.560	Age	− 2.248	XGBoost	+0.094	<BIAS>	1.000
	+0.198	Trestbps	0.809		+0.077	Ca	0.000
	+0.160	Exang	−0.499		+0.073	Thalac	163.000
	+0.081	<BIAS>	1.000		+0.037	Trestbps	129.000
	+0.009	Chol	0.275		+0.024	Chol	196.000
	+0.009	Slope	0.042		+0.017	Slope	2.000
	+0.007	Thalac	−0.781		+0.011	Fbs	0.000
	+0.003	Ca	−0.042				
	−0.014	Restecg	0.187				
	−0.019	Thal	−0.090				
	−0.025	Fbs	−0.244				
	−0.032	Oldpeak	0.324				
	−0.097	Cp	−0.190				

Table 6 Feature, weight, and contribution for XGBoost classifier using reduced feature set

XGBoost model				
Weights	Features	Contribution	Features	Value
0.1779 ± 0.0234	Oldpeak	+0.674	Thal	2.000
0.1435 ± 0.0181	Ca	+0.448	Ca	0.000
0.1247 ± 0.0314	Thal	+0.206	Chol	163.000
0.1065 ± 0.0126	Age	+0.183	Oldpeak	0.000
0.0526 ± 0.0190	Chol	+0.111	Age	196.000
0.0461 ± 0.0104	Slope	+0.096	<BIAS>	1.000
0.0331 ± 0.0199	Cp	+0.095	Thalac	0.000
0.0221 ± 0.0119	Thalac			
0.0065 ± 0.0109	Sex			



Instituto Borcelle



¡Gracias por la atención!

Daniel Casanova

