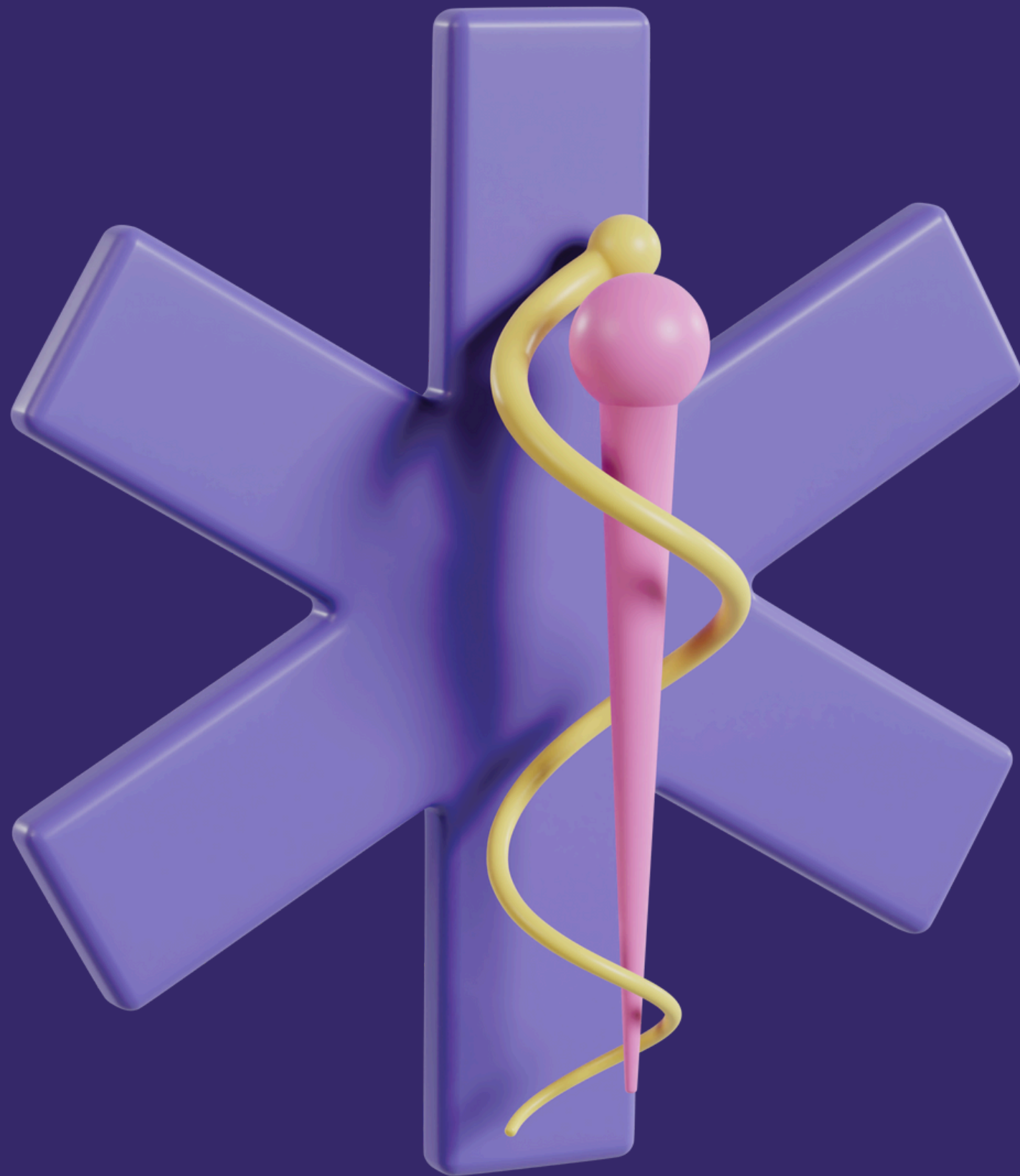


ECG DATASET



Alma Paulina González Sandoval A01631256
Paola Félix Torres A00227869
Herbert Eduardo Euroza Hernández A01635715
Jared Andres Silva Villa A01068244

INTRODUCCIÓN



Las enfermedades cardiovasculares (ECV) son la principal causa de muerte a nivel mundial, con 16.7 millones de fallecimientos anuales.

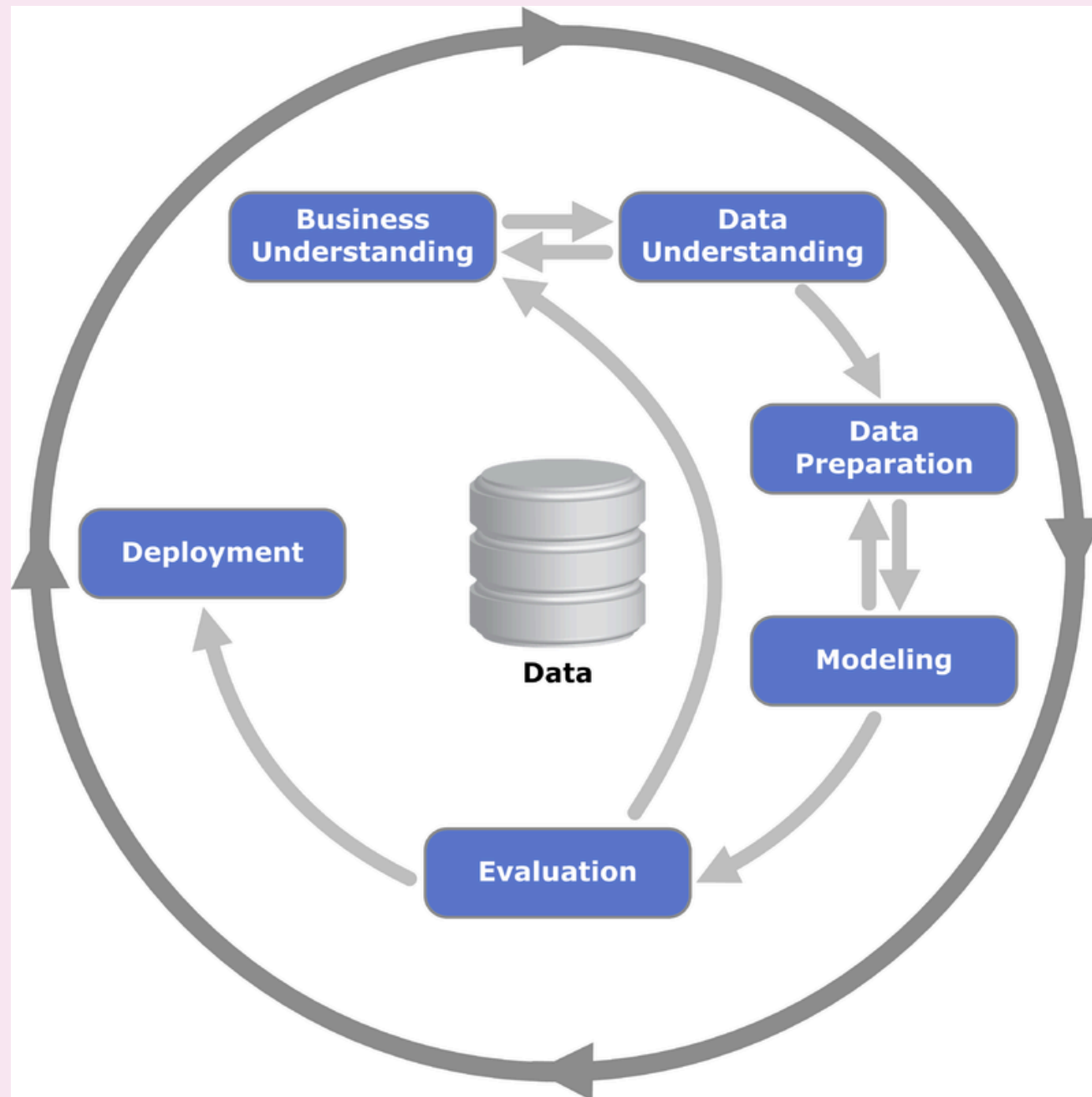
El Electrocardiograma (ECG) es una herramienta clave para detectar anomalías cardíacas y enfrenta retos en su interpretación debido a ruido y limitaciones clínicas.

Machine learning ha revolucionado el análisis de ECG, mejorando la clasificación de latidos y detección de anomalías.

Modelos avanzados como redes neuronales han demostrado alta precisión, sensibilidad y especificidad, optimizando el diagnóstico y la toma de decisiones en la salud cardíaca.

METODOLOGÍA

CRISP-DM



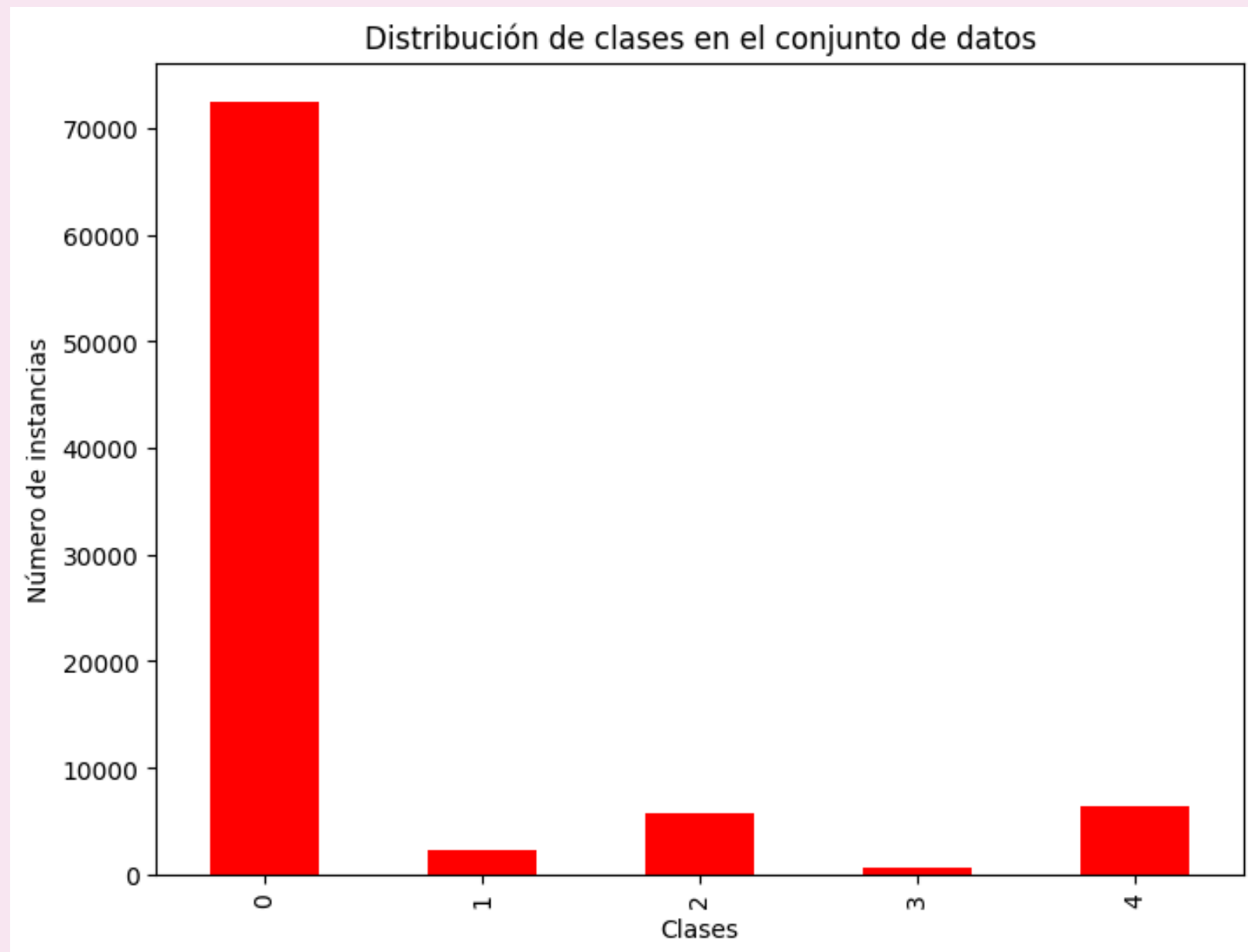
- A** Se cargan los datos en el código.
- B** Se limpian los datos de manera correcta.
- C** Se transforman los datos.
- D** Se selecciona el modelo más adecuado a la problemática.
- E** Se configura el modelo.
- F** Se entrena el modelo.
- G** Se separan los datos y se comparan.
- H** Se seleccionan las métricas adecuadas.

B

Exploración y limpieza de datos.

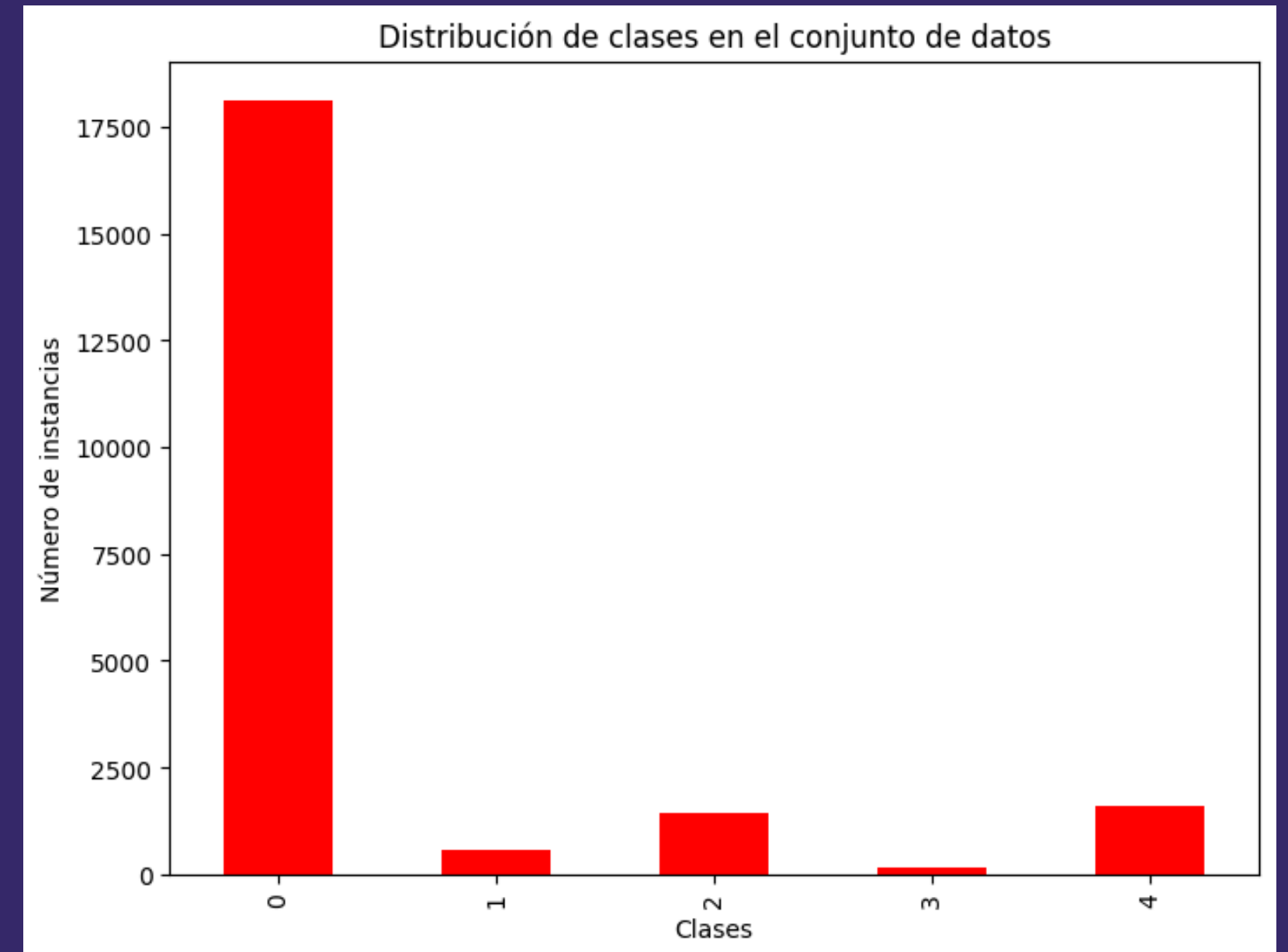


MIT DATABASE



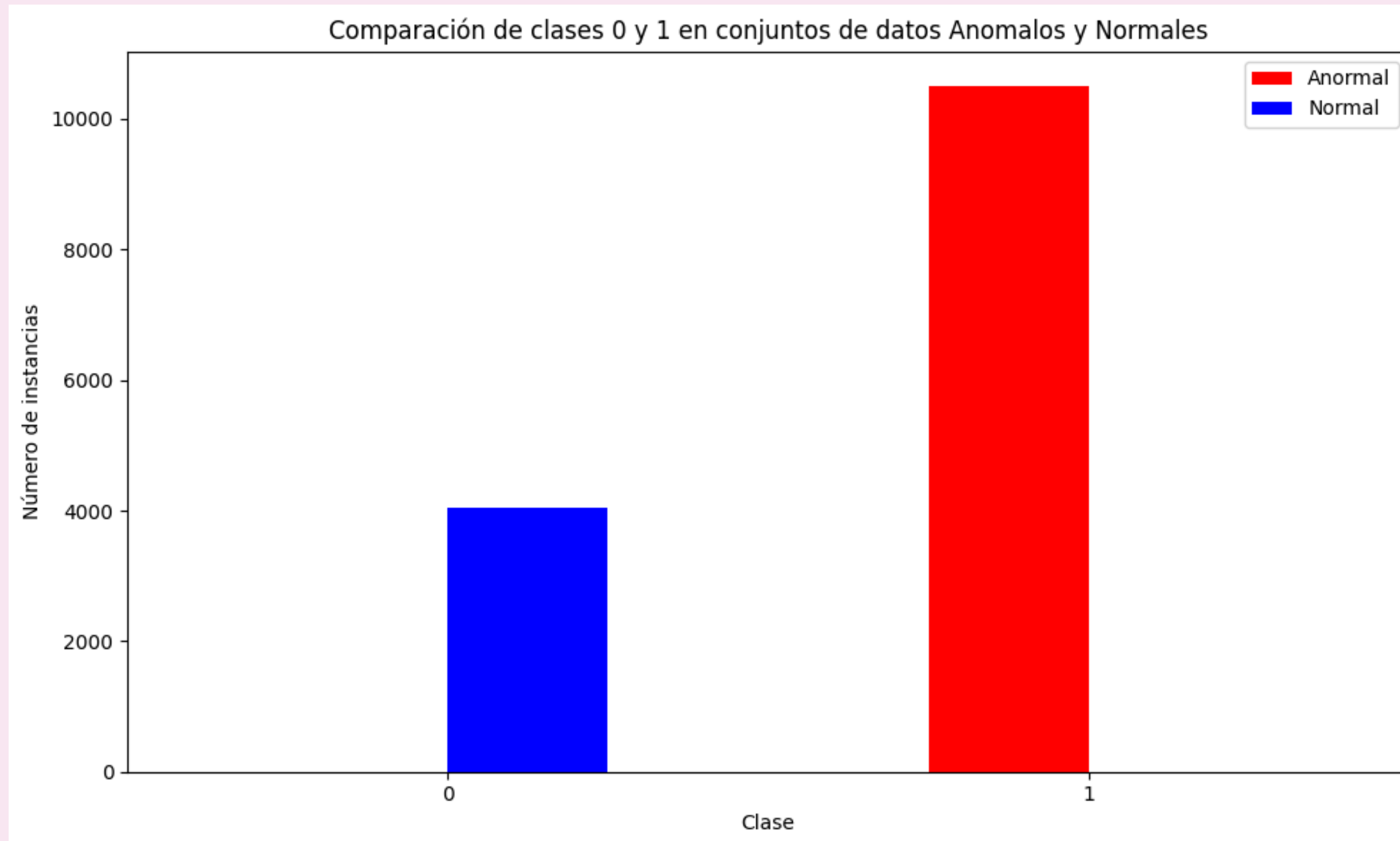
TRAIN

TEST

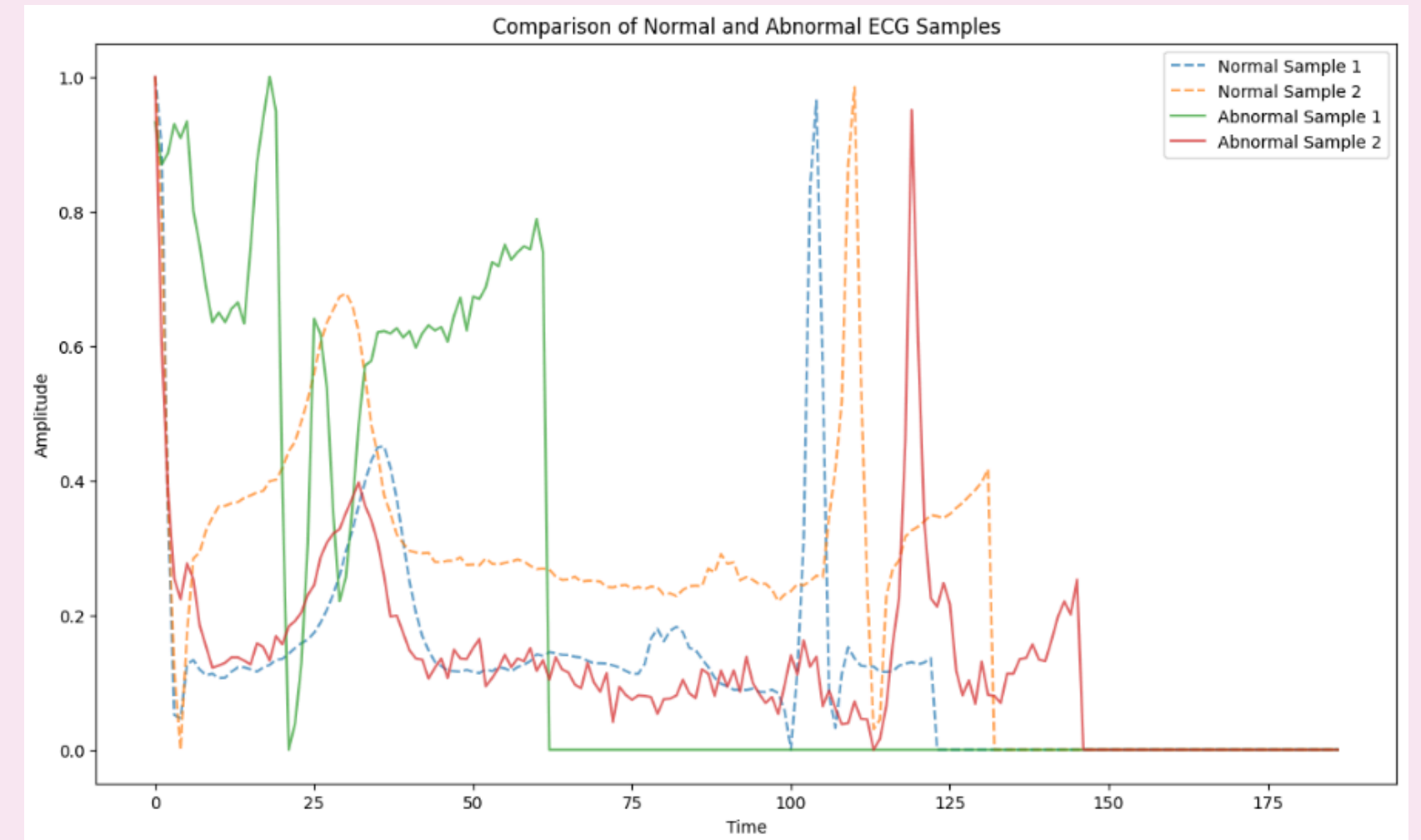
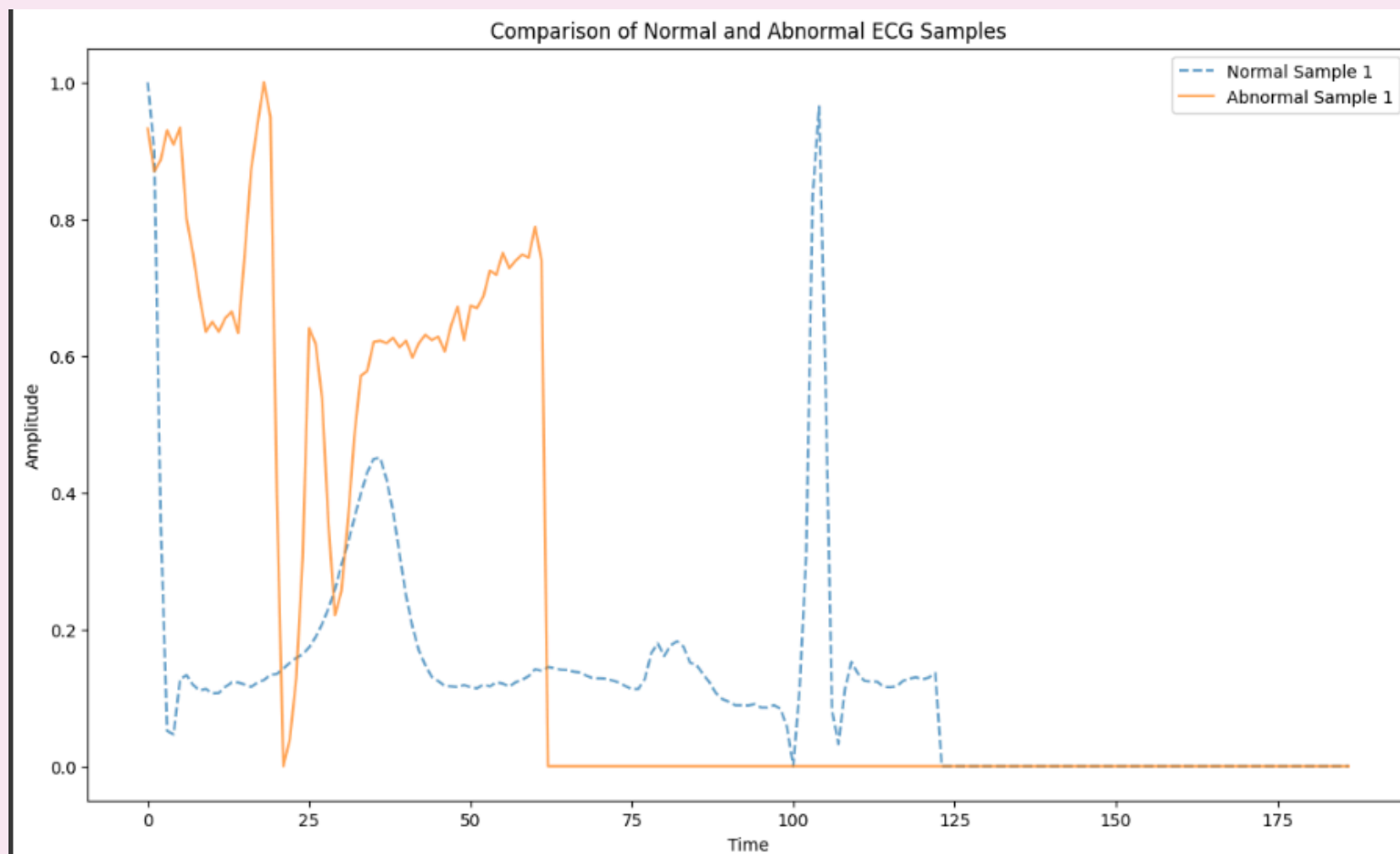


- $N(0)$ = Ritmo Normal
- $S(1)$ = Arritmias supraventriculares.
- $Q(2)$ = Desconocido o sin clasificar.
- $V(3)$ = Arritmias ventriculares.
- $F(4)$ = Latidos de fusión.

PTB DATABASE



VISUALIZACIÓN



LIMPIEZA

DATOS NORMALIZADOS

DATOS NULOS

0	1	2	3	4	\
21892.000000	21892.000000	21892.000000	21892.000000	21892.000000	
0.894410	0.761902	0.426627	0.221596	0.201676	
0.234560	0.218659	0.228572	0.208711	0.177727	
0.000000	0.000000	0.000000	0.000000	0.000000	
0.924260	0.683366	0.251197	0.050505	0.082873	
0.990431	0.828996	0.432777	0.167630	0.147642	
1.000000	0.912319	0.583991	0.347092	0.259211	
1.000000	1.000000	1.000000	1.000000	1.000000	
5	6	7	8	9	\
21892.000000	21892.000000	21892.000000	21892.000000	21892.000000	
0.209891	0.204805	0.200992	0.197634	0.196022	
0.172194	0.177946	0.176142	0.170228	0.166707	
0.000000	0.000000	0.000000	0.000000	0.000000	
0.087912	0.072663	0.065997	0.064516	0.068493	
0.158111	0.144068	0.144509	0.150422	0.149029	
0.287356	0.298453	0.294563	0.289907	0.282956	
1.000000	1.000000	1.000000	1.000000	0.991429	
...	178	179	180	181	\
...	21892.000000	21892.000000	21892.000000	21892.000000	
...	0.004588	0.004327	0.004020	0.003789	
...	0.043128	0.042187	0.040255	0.039397	
...	0.000000	0.000000	0.000000	0.000000	
...	0.000000	0.000000	0.000000	0.000000	
...	0.000000	0.000000	0.000000	0.000000	
...	0.000000	0.000000	0.000000	0.000000	
...	0.000000	0.000000	0.000000	0.000000	
...	0.980392	1.000000	0.966102	1.000000	
182	183	184	185	186	\
21892.000000	21892.000000	21892.000000	21892.000000	21892.000000	
0.003638	0.003459	0.003166	0.003000	0.002946	
0.038535	0.037717	0.035903	0.035522	0.035266	
0.000000	0.000000	0.000000	0.000000	0.000000	
0.000000	0.000000	0.000000	0.000000	0.000000	
0.000000	0.000000	0.000000	0.000000	0.000000	
0.000000	0.000000	0.000000	0.000000	0.000000	
1.000000	1.000000	1.000000	0.996053	1.000000	

```
Total de valores nulos para ptbdb_abnormal: 0
Total de valores nulos para ptbdb_normal: 0
Total de valores nulos para mitbih_train: 0
Total de valores nulos para mitbih_test: 0
```


D

Prueba de modelos



Modelos utilizados	Exactitud
Regresión logística sin balanceo	0.91
Random Forest (RandomUnderSample)	0.85
SVM (RandomUnderSample)	0.80
K-Nearest Neighbors (RandomUnderSample)	0.75

Modelos utilizados	Exactitud
Decision Tree (RandomUnderSample)	0.71
Naive Bayes (RandomUnderSample)	0.17
Gradient Boosting (RandomUnderSample)	0.79
AdaBoost (RandomUnderSample)	0.48

REPORTES DE CLASIFICACIÓN

Regresión logística
sin balanceo

Reporte de clasificación:				
	precision	recall	f1-score	support
0.0	0.92	0.98	0.95	21828
1.0	0.85	0.39	0.53	627
2.0	0.64	0.33	0.43	1704
3.0	0.63	0.20	0.30	200
4.0	0.95	0.88	0.92	1908
accuracy			0.91	26267
macro avg	0.80	0.56	0.63	26267
weighted avg	0.90	0.91	0.90	26267

K-Nearest
Neighbors

Reporte de clasificación:				
	precision	recall	f1-score	support
0.0	0.98	0.74	0.84	14452
1.0	0.19	0.81	0.31	465
2.0	0.51	0.84	0.63	1183
3.0	0.09	0.88	0.16	127
4.0	0.77	0.93	0.85	1283
accuracy			0.76	17510
macro avg	0.51	0.84	0.56	17510
weighted avg	0.91	0.76	0.81	17510

Gradient
Boosting

Reporte de clasificación:				
	precision	recall	f1-score	support
0.0	0.98	0.78	0.87	14452
1.0	0.24	0.80	0.37	465
2.0	0.56	0.89	0.69	1183
3.0	0.09	0.87	0.17	127
4.0	0.85	0.93	0.89	1283
accuracy			0.80	17510
macro avg	0.55	0.86	0.60	17510
weighted avg	0.92	0.80	0.84	17510

REPORTES DE CLASIFICACIÓN

SVM

Reporte de clasificación:				
	precision	recall	f1-score	support
0.0	0.98	0.80	0.88	14452
1.0	0.33	0.75	0.46	465
2.0	0.59	0.86	0.70	1183
3.0	0.08	0.87	0.14	127
4.0	0.83	0.91	0.87	1283
accuracy			0.81	17510
macro avg	0.56	0.84	0.61	17510
weighted avg	0.92	0.81	0.85	17510

AdaBoost

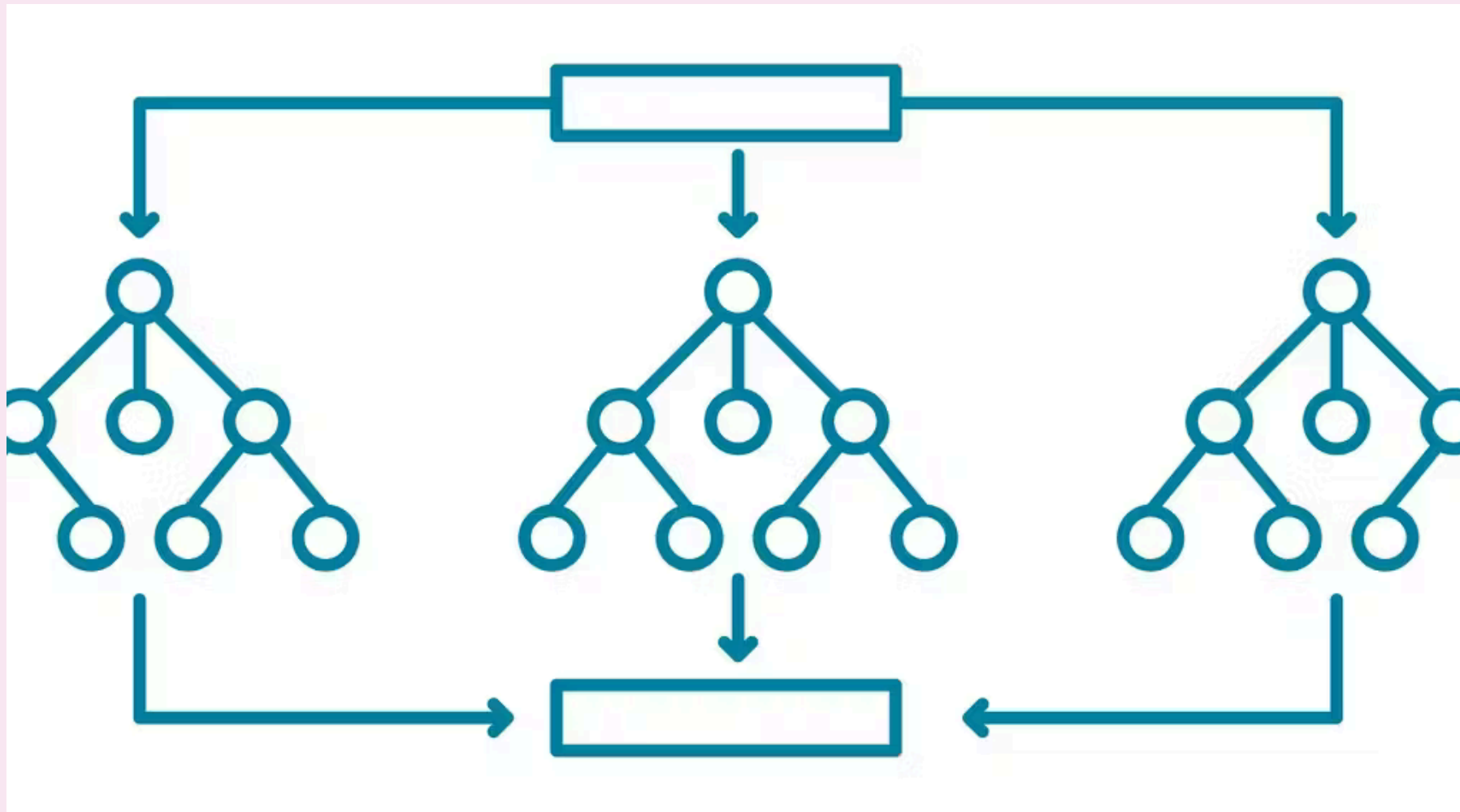
Reporte de clasificación:				
	precision	recall	f1-score	support
0.0	0.93	0.36	0.51	14452
1.0	0.05	0.55	0.09	465
2.0	0.25	0.62	0.35	1183
3.0	0.06	0.83	0.11	127
4.0	0.65	0.90	0.75	1283
accuracy			0.42	17510
macro avg	0.39	0.65	0.36	17510
weighted avg	0.83	0.42	0.51	17510

Random Forest

Reporte de clasificación:				
	precision	recall	f1-score	support
0.0	0.99	0.85	0.91	14452
1.0	0.34	0.80	0.47	465
2.0	0.62	0.92	0.74	1183
3.0	0.13	0.88	0.22	127
4.0	0.91	0.95	0.93	1283
accuracy			0.86	17510
macro avg	0.59	0.88	0.65	17510
weighted avg	0.93	0.86	0.88	17510

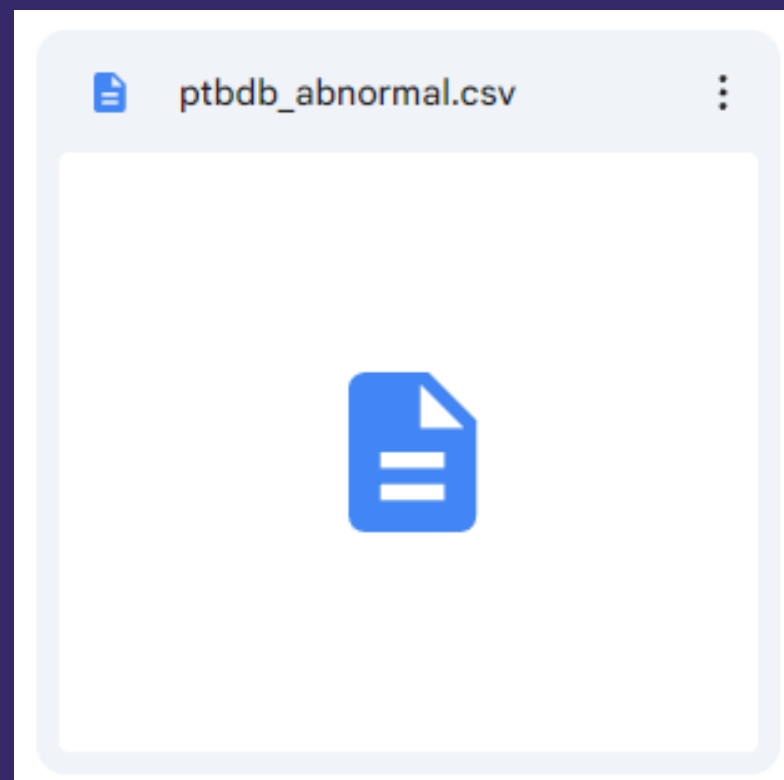
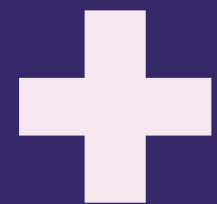
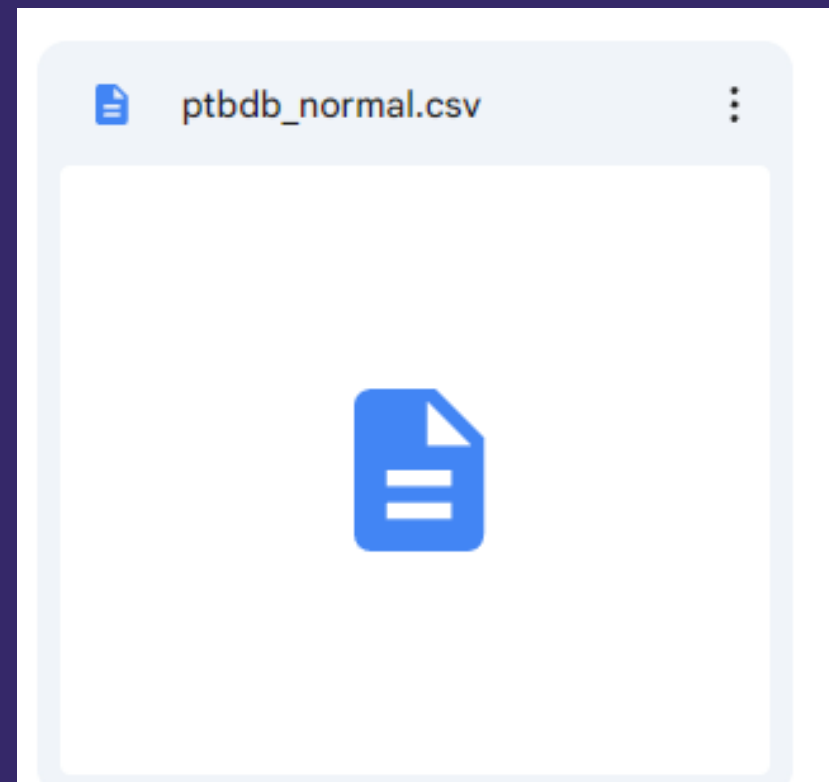
SOLUCIÓN

RANDOM FOREST



Random Forest es un modelo de aprendizaje automático, este se basa en combinaciones de múltiples árboles de decisión para realizar una predicción.

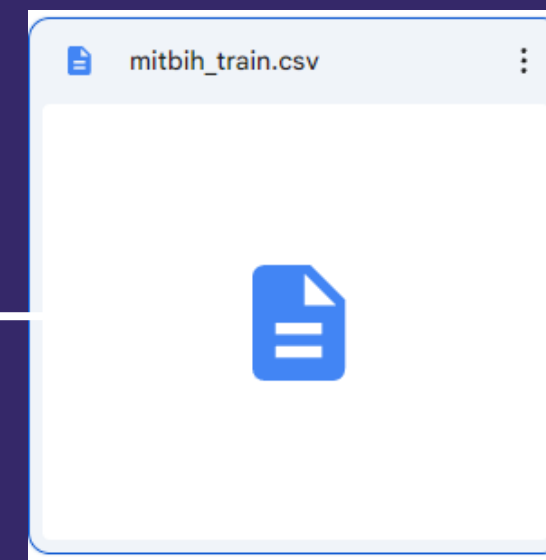
La idea principal es construir un “bosque” de árboles de decisión, que hacen predicciones mediante el voto mayoritario de todos los árboles.



Modelo 1

1

0



Modelo 2

S

Q

V

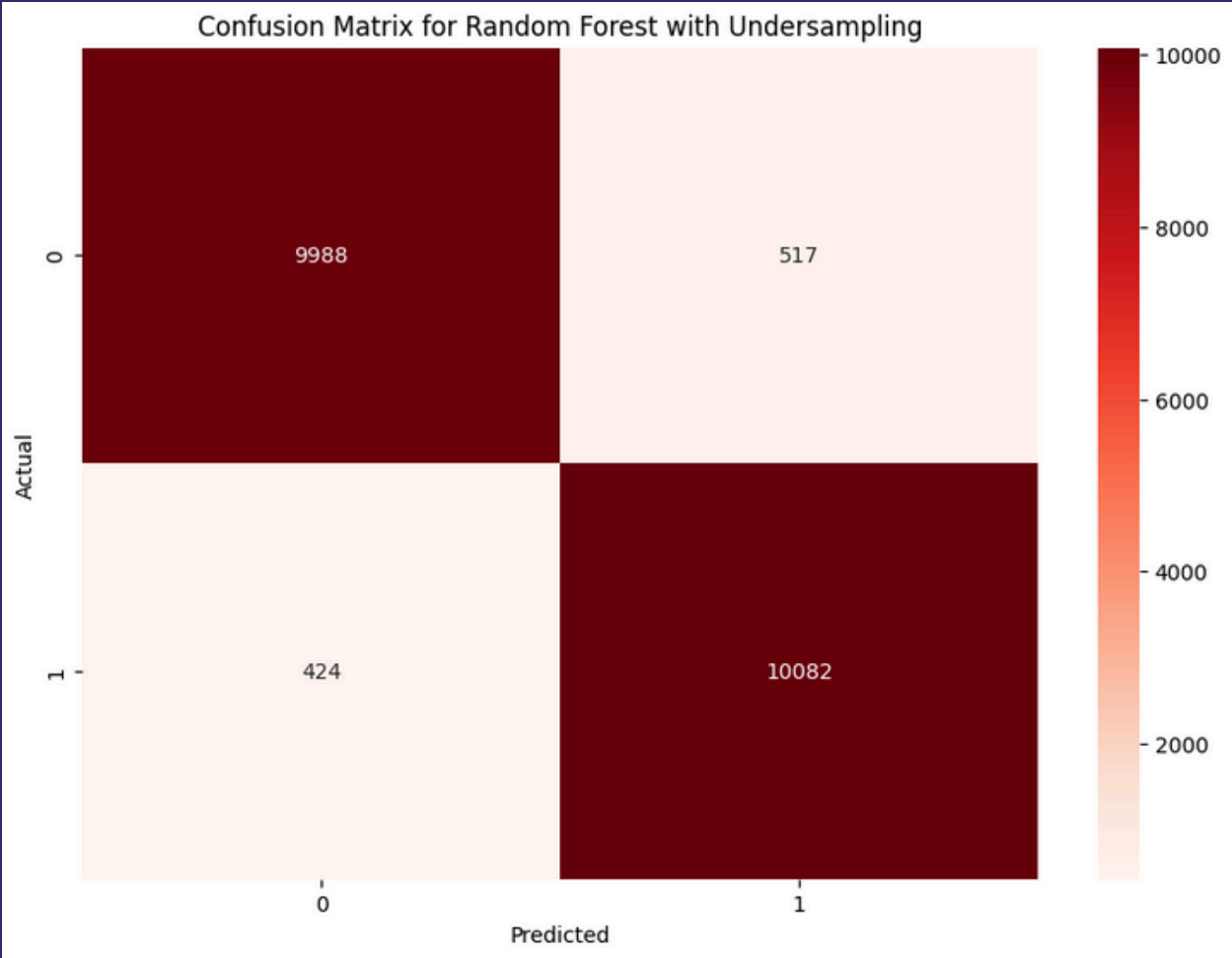
F

N

PRIMER MODELO

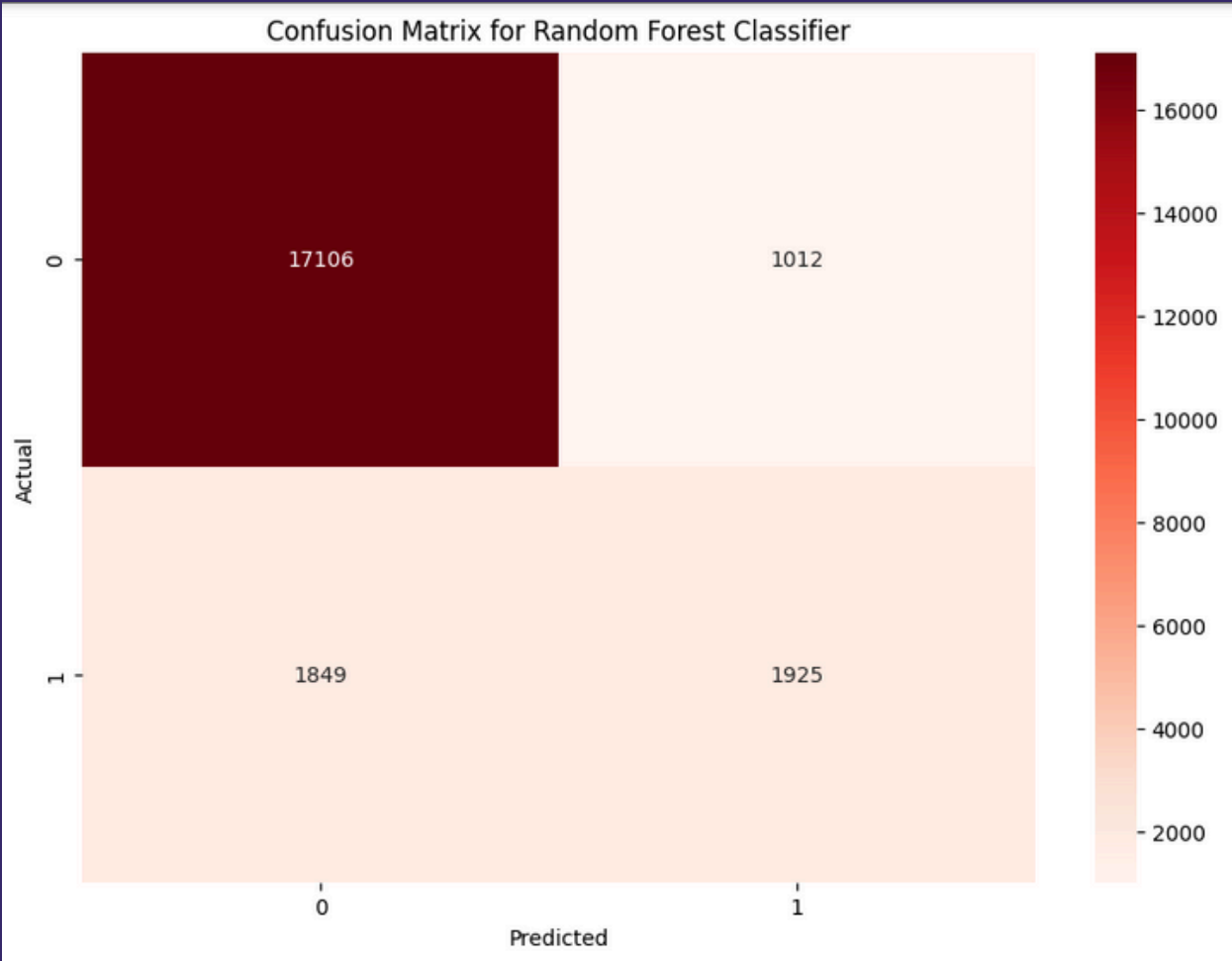
Entrenamiento

➡ Número de ejemplos por clase después de balancear:
Clase 1.0: 10506 ejemplos
Clase 0.0: 10505 ejemplos



Reporte de Clasificación - Random Forest con Undersampling:				
	precision	recall	f1-score	support
0.0	0.96	0.95	0.96	10505
1.0	0.95	0.96	0.96	10506
accuracy			0.96	21011
macro avg	0.96	0.96	0.96	21011
weighted avg	0.96	0.96	0.96	21011

Evaluación con mitbih_test



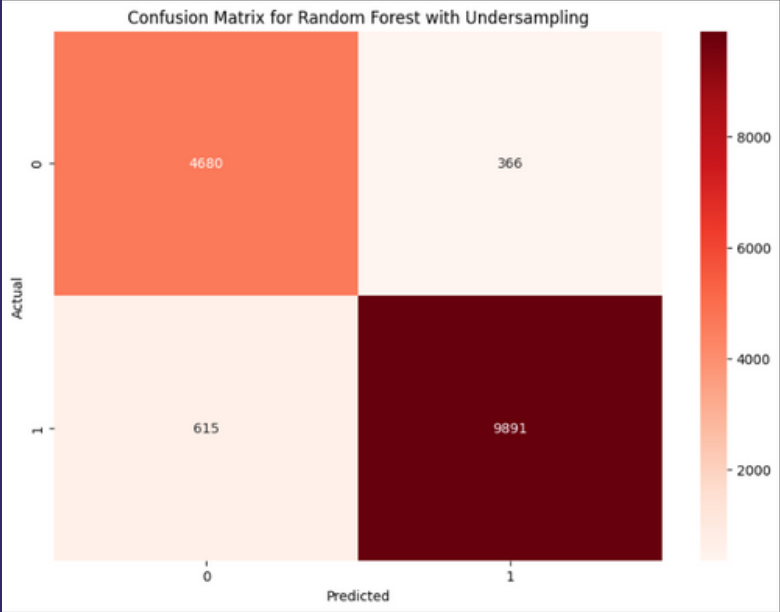
Reporte de Clasificación mitbih_test - Random Forest:				
	precision	recall	f1-score	support
0.0	0.90	0.94	0.92	18118
1.0	0.66	0.51	0.57	3774
accuracy			0.87	21892
macro avg	0.78	0.73	0.75	21892
weighted avg	0.86	0.87	0.86	21892

1

➡ Número de ejemplos por clase después de balancear:
Clase 1.0: 10506 ejemplos
Clase 0.0: 5046 ejemplos

Reporte de Clasificación - Random Forest con Undersampling:

	precision	recall	f1-score	support
0.0	0.88	0.93	0.91	5046
1.0	0.96	0.94	0.95	10506
accuracy			0.94	15552
macro avg	0.92	0.93	0.93	15552
weighted avg	0.94	0.94	0.94	15552

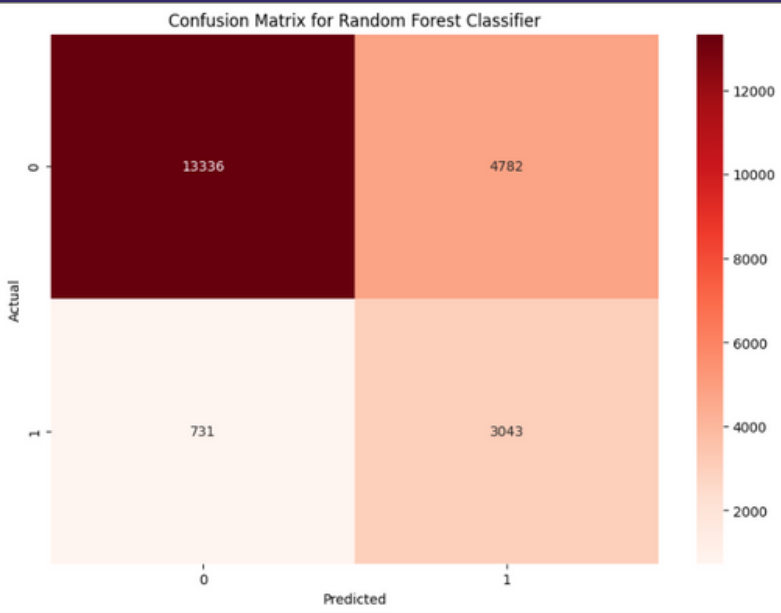


2

➡ Número de ejemplos por clase después de balancear:
Clase 1.0: 10506 ejemplos
Clase 0.0: 5546 ejemplos

Reporte de Clasificación mitbih_test - Random Forest:

	precision	recall	f1-score	support
0.0	0.95	0.74	0.83	18118
1.0	0.39	0.81	0.52	3774
accuracy			0.75	21892
macro avg	0.67	0.77	0.68	21892
weighted avg	0.85	0.75	0.78	21892

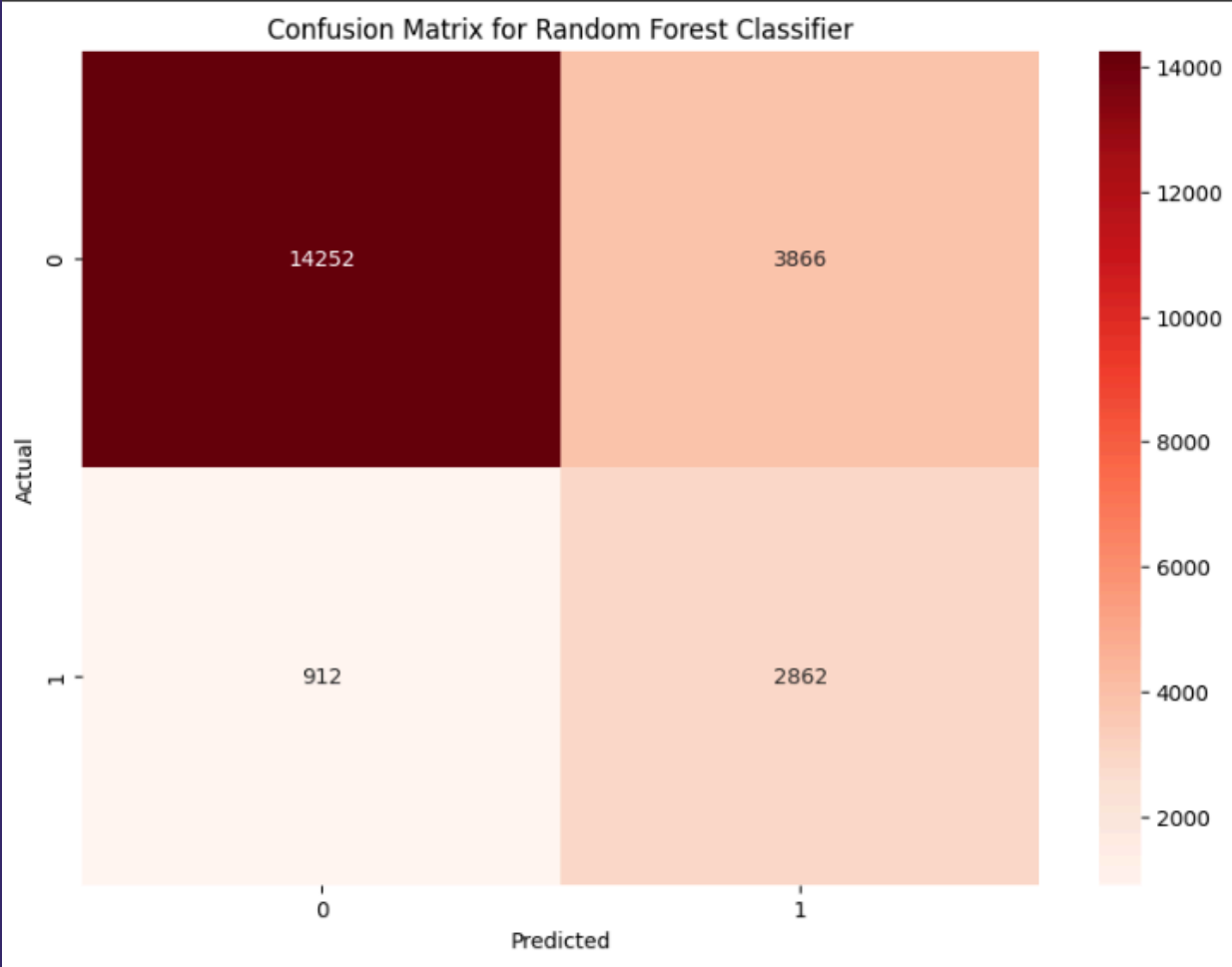


REPORTE DEL PRIMER MODELO

Reporte de Clasificación - Random Forest con Undersampling:				
	precision	recall	f1-score	support
0.0	0.92	0.94	0.93	6546
1.0	0.96	0.95	0.95	10506
accuracy			0.94	17052
macro avg	0.94	0.94	0.94	17052
weighted avg	0.94	0.94	0.94	17052

Evaluación con mitbih_test

Reporte de Clasificación mitbih_test - Random Forest:				
	precision	recall	f1-score	support
0.0	0.94	0.79	0.86	18118
1.0	0.43	0.76	0.55	3774
accuracy			0.78	21892
macro avg	0.68	0.77	0.70	21892
weighted avg	0.85	0.78	0.80	21892



SEGUNDO MODELO

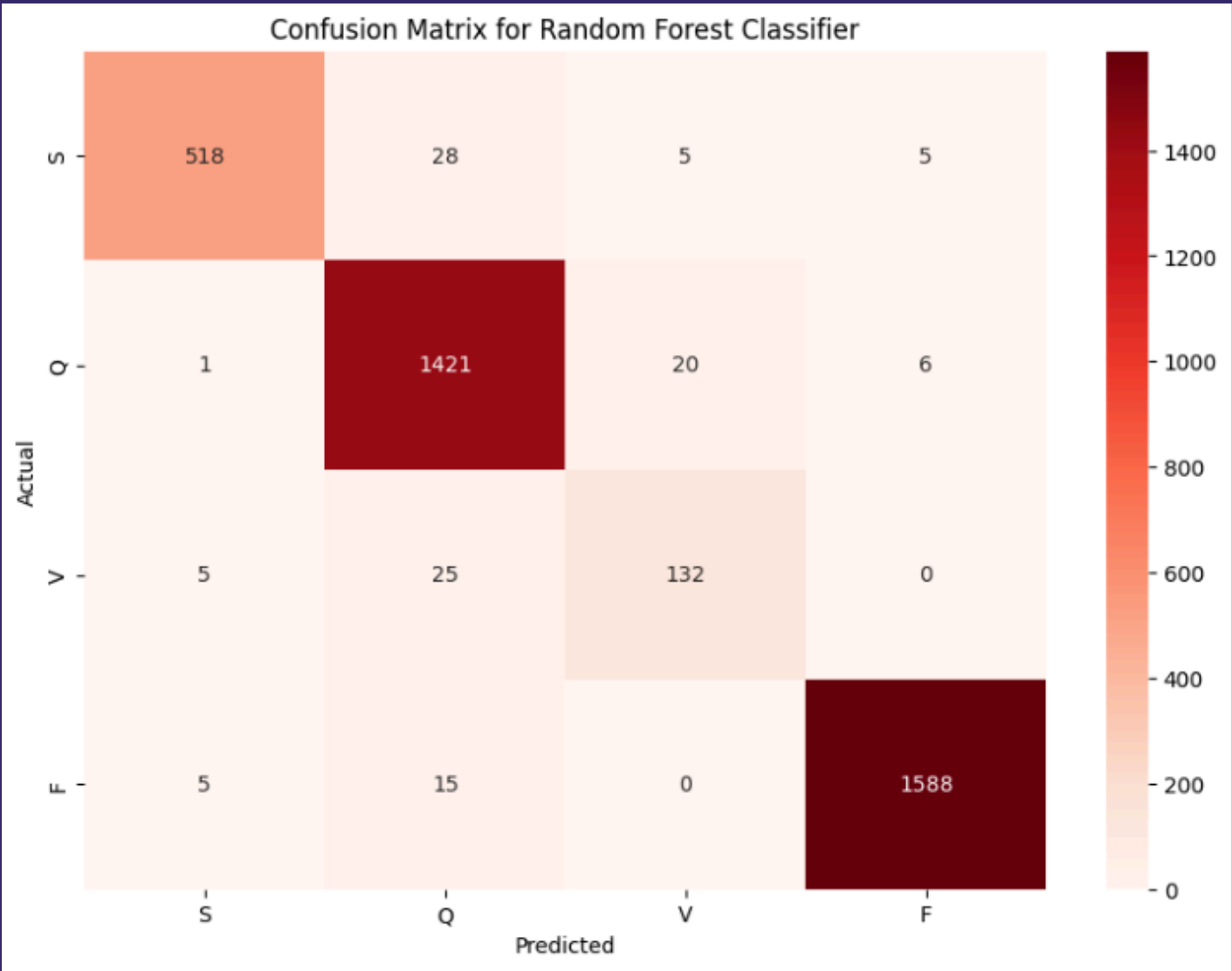
REPORTE DEL SEGUNDO MODELO

Reporte de Clasificación - Random Forest con Undersampling:				
	precision	recall	f1-score	support
1.0	0.89	0.94	0.91	2223
2.0	0.96	0.92	0.94	5788
3.0	0.66	0.93	0.77	641
4.0	0.99	0.97	0.98	6431
accuracy			0.94	15083
macro avg	0.87	0.94	0.90	15083
weighted avg	0.95	0.94	0.94	15083

- S = Arritmias supraventriculares.
- Q = Desconocido o sin clasificar.
- V= Arritmias ventriculares.
- F = Latidos de fusión.

Evaluación con mitbih_test

Reporte de Clasificación mitbih_test - Random Forest:				
	precision	recall	f1-score	support
1.0	0.98	0.93	0.95	556
2.0	0.95	0.98	0.97	1448
3.0	0.84	0.81	0.83	162
4.0	0.99	0.99	0.99	1608
accuracy			0.97	3774
macro avg	0.94	0.93	0.94	3774
weighted avg	0.97	0.97	0.97	3774



¿COMO PODEMOS MEJORAR NUESTRO MODELO?

OPTIMIZACIÓN DE HIPERPARÁMETROS

- Exploración más profunda de los parámetros del modelo

En ejecución (2 h 28 min 5 s) <cell line: 19> >

En ejecución (3 h 19 min 31 s) <cell line: 19> >

En ejecución (4 h 33 min 19 s) <cell line: 19> >

MODELOS ENSAMBLADOS

- Combinación de modelos para mejorar la precisión.

DATOS ADICIONALES

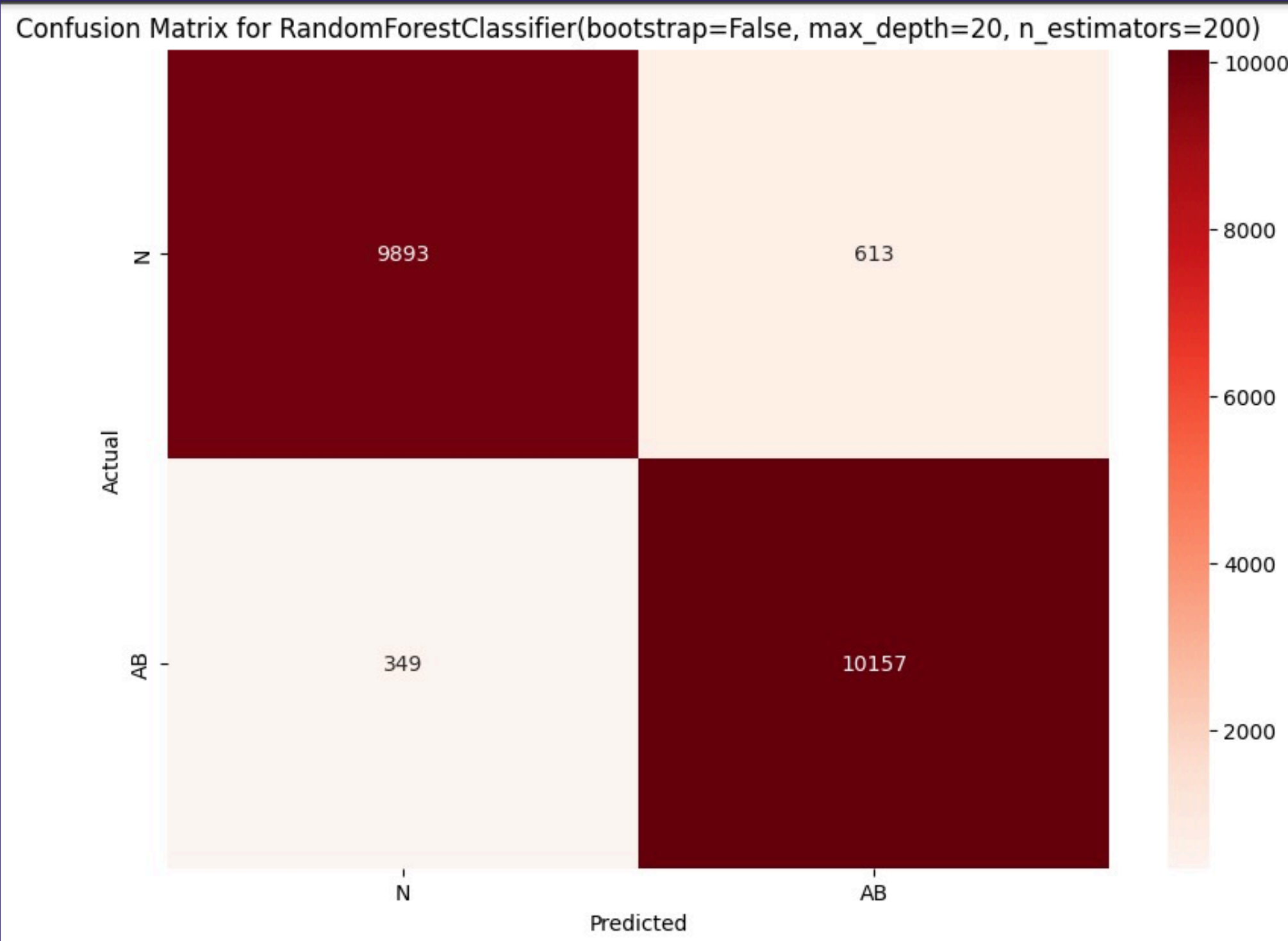
- Incorporación de datos adicionales relevantes para mejorar el modelo

CONCLUSIÓN

En problemas de clasificación, como lo es el análisis de electrocardiogramas (ECG), la capacidad del clasificador para detectar correctamente los casos anormales es de suma importancia. La mayoría de los casos anormales suelen ser raros y críticos, por lo que se requiere de una atención inmediata.

La continua mejora de los modelos nos ayudarán para que en un futuro estos problemas de malas detecciones se reduzcan considerablemente.

EVIDENCIA DE HIPERPARAMETROS



Reporte de Clasificación - Random Forest con Mejor Parámetro (Randomized Search)				
	precision	recall	f1-score	support
0.0	0.97	0.94	0.95	10506
1.0	0.94	0.97	0.95	10506
accuracy			0.95	21012
macro avg	0.95	0.95	0.95	21012
weighted avg	0.95	0.95	0.95	21012
Exactitud (Accuracy): 0.95				

GRACIAS!

