



5th International Symposium on Master Engineering *Booklets*



RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - Google Scholar DOI - REDIB - Mendeley - DIALNET - ROAD - ORCID - V|LEX - EBSCO

Title: Implementation of a Multilayer Neural Network based on MFCC, Delta, and Delta-Delta features for COPD detection

Author: Villafuerte-Lucio, Diego Ángel

Editorial label MARVID: 607-8695
BMARVID Control Number: 2025-01
BMARVID Classification (2025): 021025-0001

Pages: 32

RNA: 03-2010-032610115700-14

MARVID-México

Park Pedregal Business. 3580,
Anillo Perif., San Jerónimo
Aculco, Álvaro Obregón,
01900 Ciudad de México, CDMX,
Phone: +52 1 55 6159 2296
Skype: MARVID-México S.C.
E-mail: contact@marvid.org
Facebook: MARVID-México S. C.
X: [@Marvid_México](https://twitter.com/Marvid_México)

www.marvid.org

Holdings

Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic Republic
Spain	El Salvador	of Congo
Ecuador	Taiwan	
Peru	Paraguay	Nicaragua



Figura 1. Inteligencia artificial

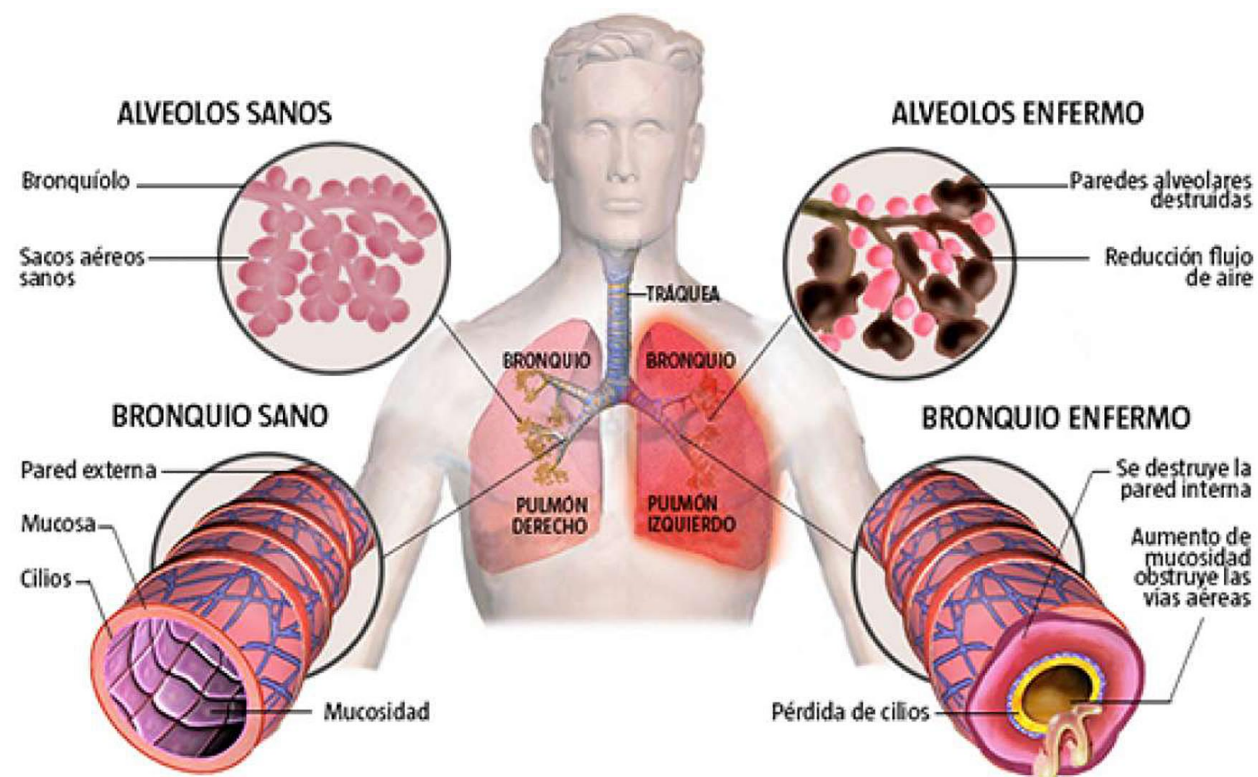
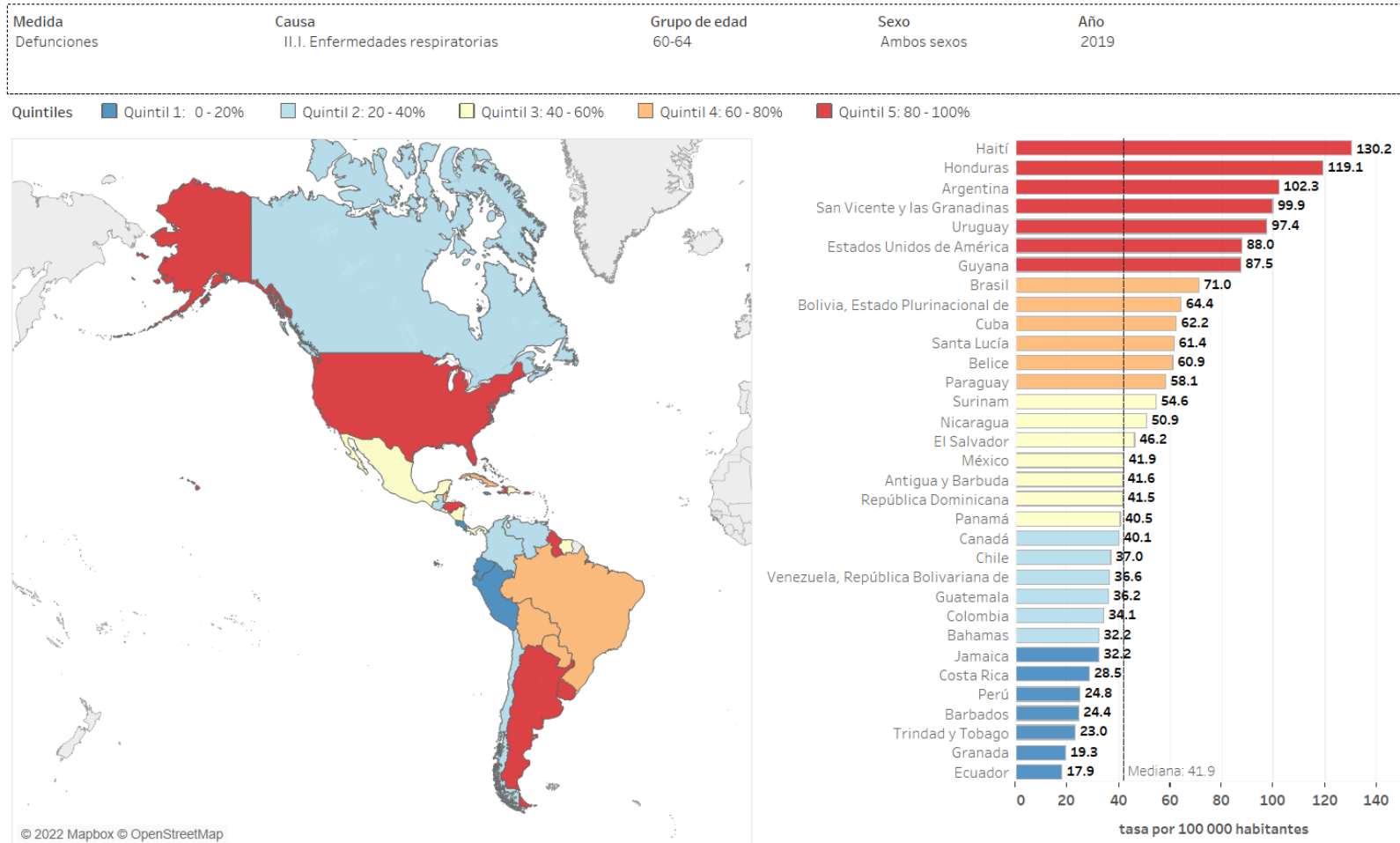


Figura 2. Afectaciones por el EPOC

Fuente: <https://www.cuidate-murcia.com/epoc-enfermedad-pulmonar-obstruccion-cronica/>



Fuente: <https://www.cuidate-murcia.com/epoc-enfermedad-pulmonar-obstructiva-cronica/>

Figura 3. Representación de las afecciones por el EPOC en el continente Americano.



¿Qué son las características MFCC?

Las características MFCC representan la forma en que varían las características espectrales en el tiempo, capturan información relevante sobre la forma y el contenido espectral de la señal de audio.

MFCC 1	Representa la energía total del espectro de frecuencia del audio.	MFCC 5	Captura características más finas o detalles del espectro de frecuencia.
MFCC 2	Se relaciona con la frecuencia dominante o el tono fundamental del sonido.	MFCC 6-9	Se asocian con armónicos y componentes espectrales más complejos.
MFCC 3	Está relacionado con el equilibrio espectral y puede revelar características armónicas o resonantes del audio.	MFCC 10-13	Están relacionados con características más sutiles del espectro de frecuencia
MFCC 4	Representa información sobre la envolvente espectral y puede capturar cambios o transiciones suaves en la forma de onda a lo largo del tiempo.		

Figura 4. Representación de los coeficientes MFCC



¿Qué son las características Delta?

Las características **Delta** representan la velocidad de cambio de los MFCC con el tiempo.

- Derivada de las características MFCC.
- Añaden información dinámica.
- Complementan información que las características MFCC no capturan.

“Mientras que los MFCC nos dicen qué frecuencias predominan, las deltas nos dicen cómo cambian esas frecuencias en el tiempo es decir, su velocidad.”

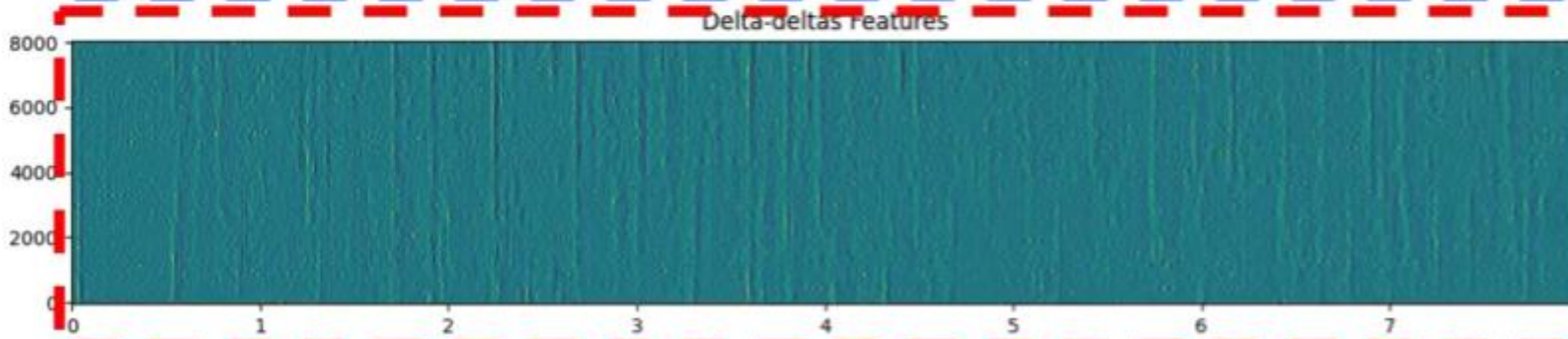
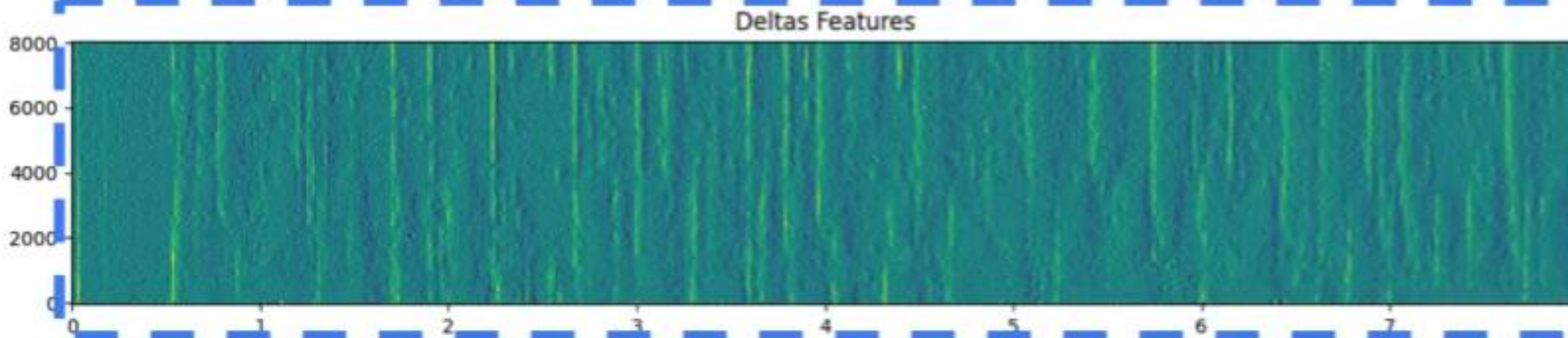
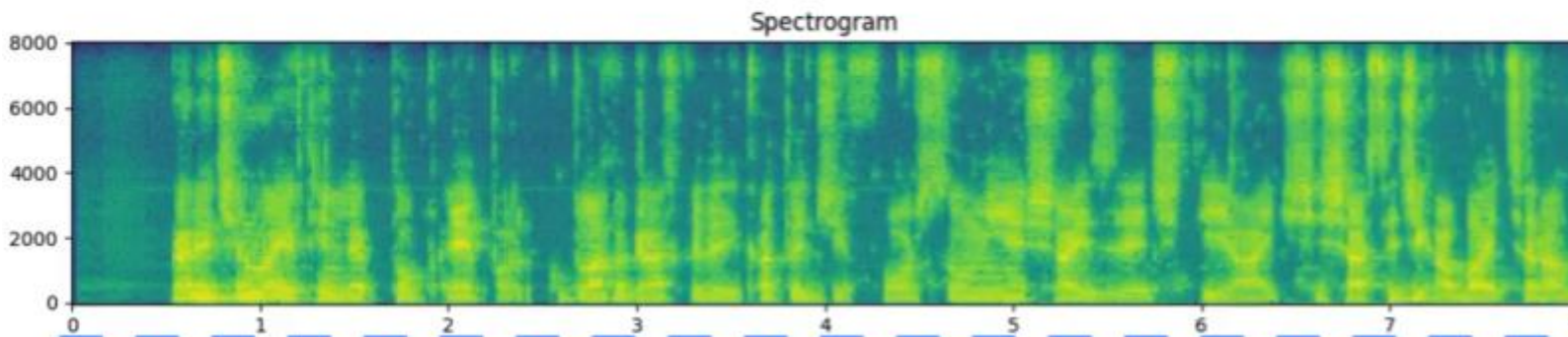
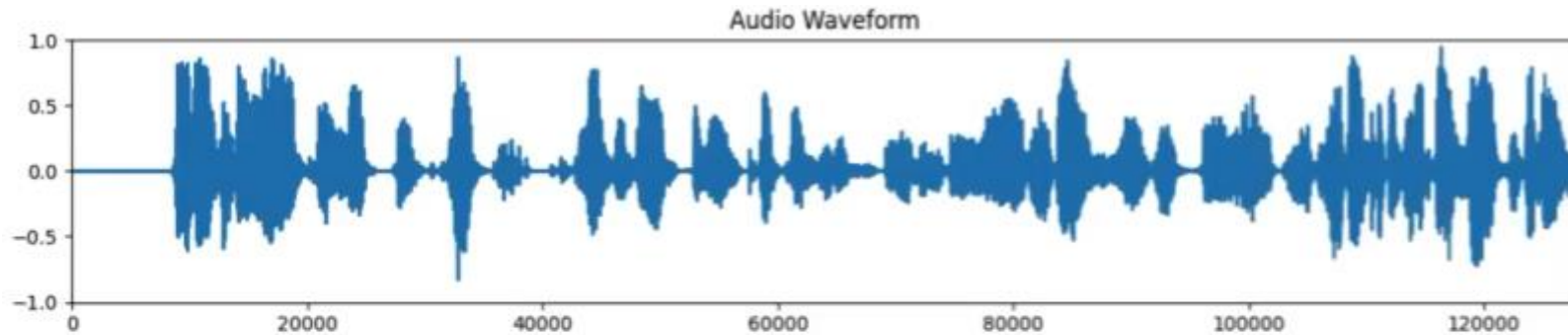


¿Qué son las características Delta-Delta?

Las características **Delta-Delta** son la **derivada** de la **derivada** de los MFCC.

- Al ser la derivada de la derivada esto nos da la aceleración.
- Capturan cambios rápidos o micro variaciones.

MFCC = posición, **Delta** = velocidad, **Delta-Delta** = aceleración



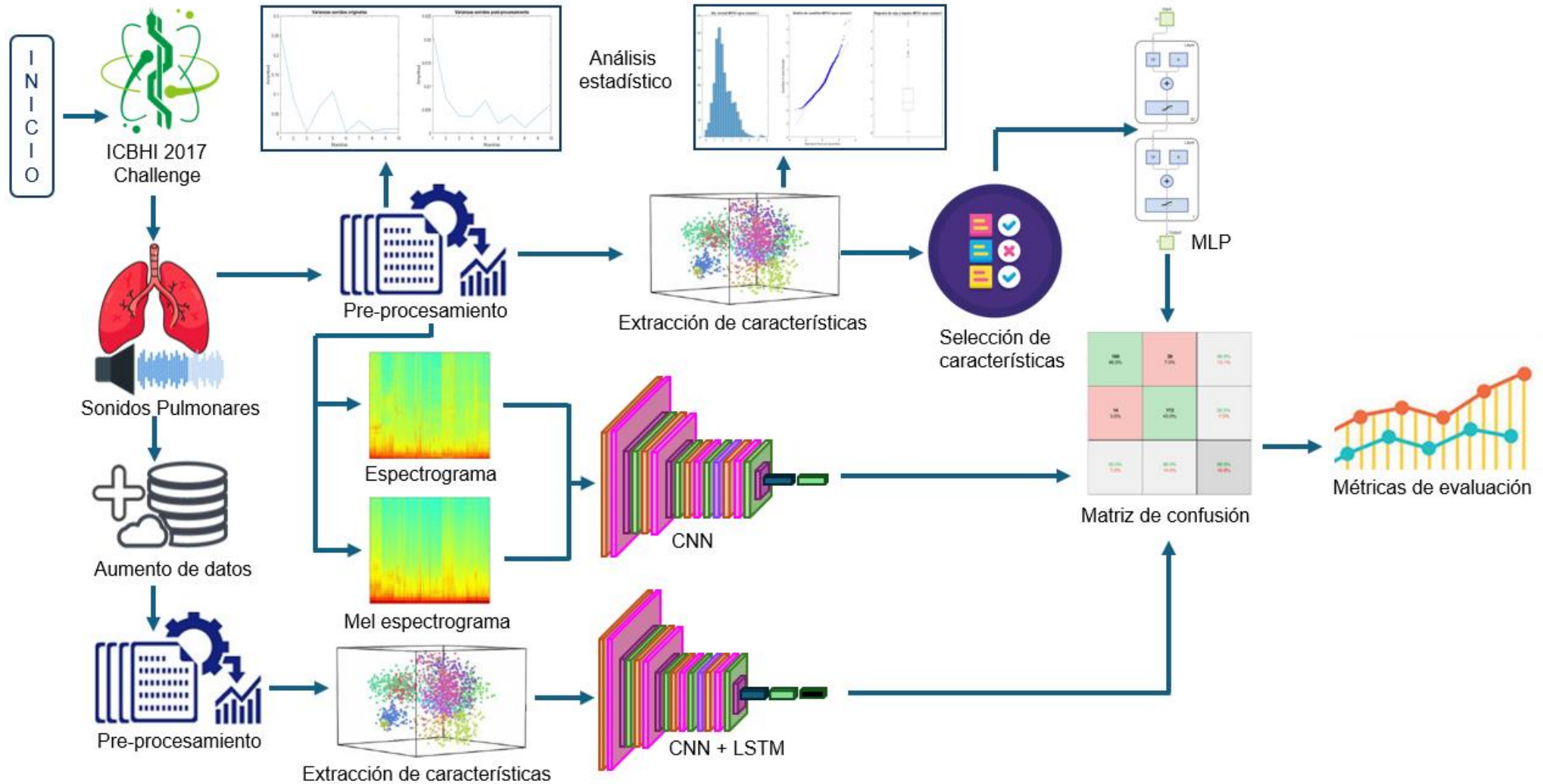


Figura 5. Metodología propuesta para la clasificación de señales de audio (Segmentación conjunto de datos ICBHI).

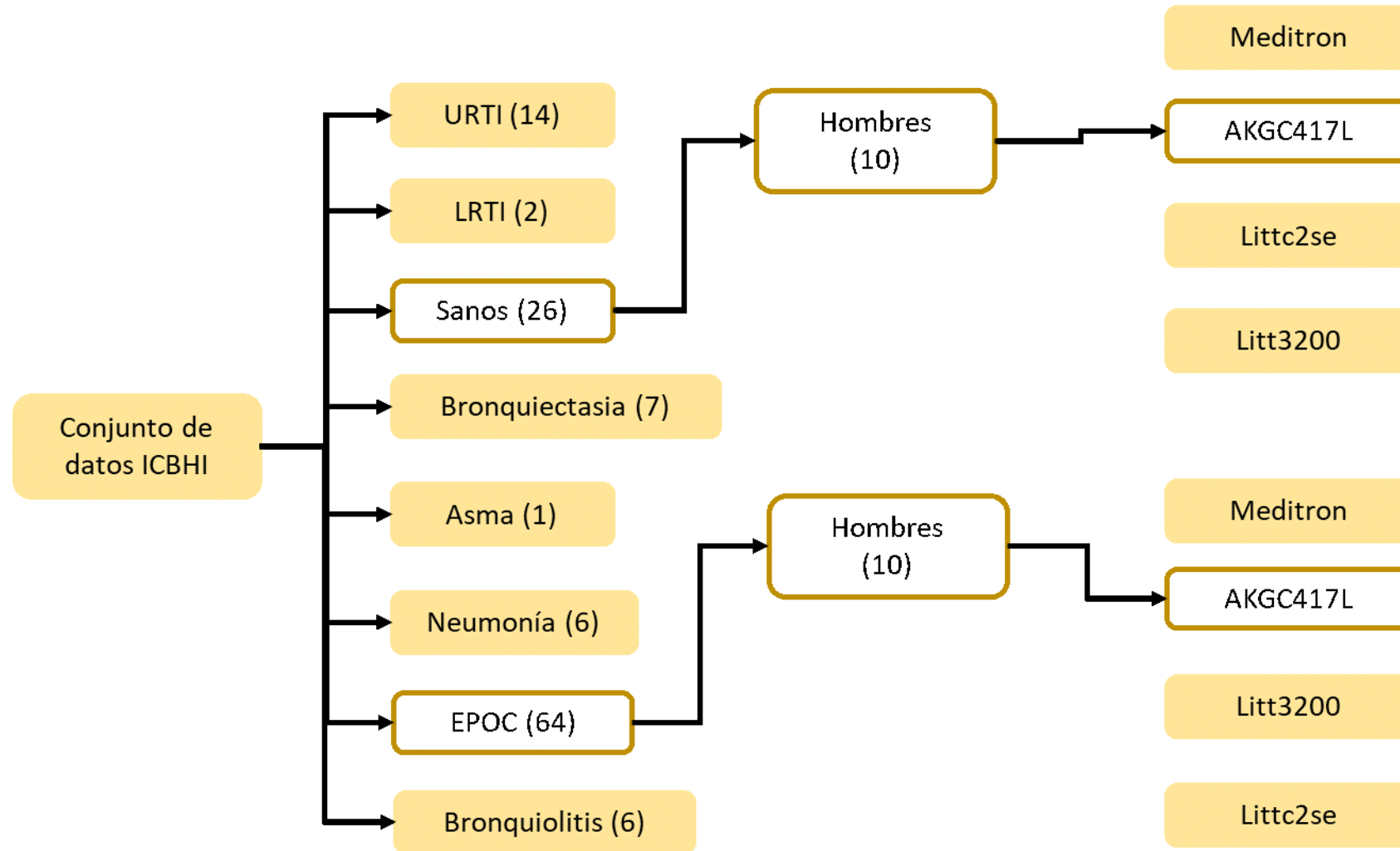


Figura 6. Segmentación de los sonidos respiratorios

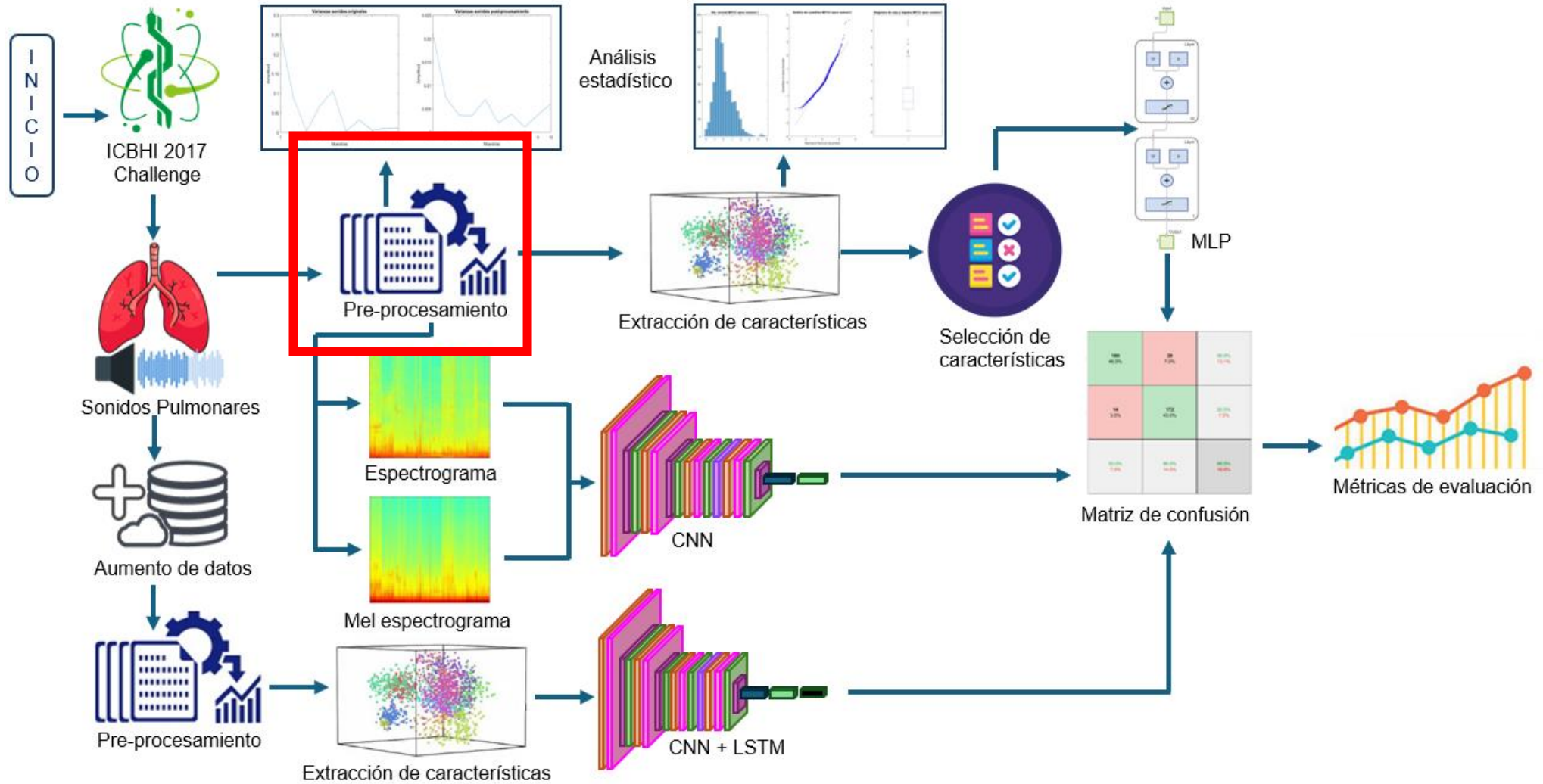


Figura 7. Metodología propuesta para la clasificación de señales de audio (**Preprocesamiento**).

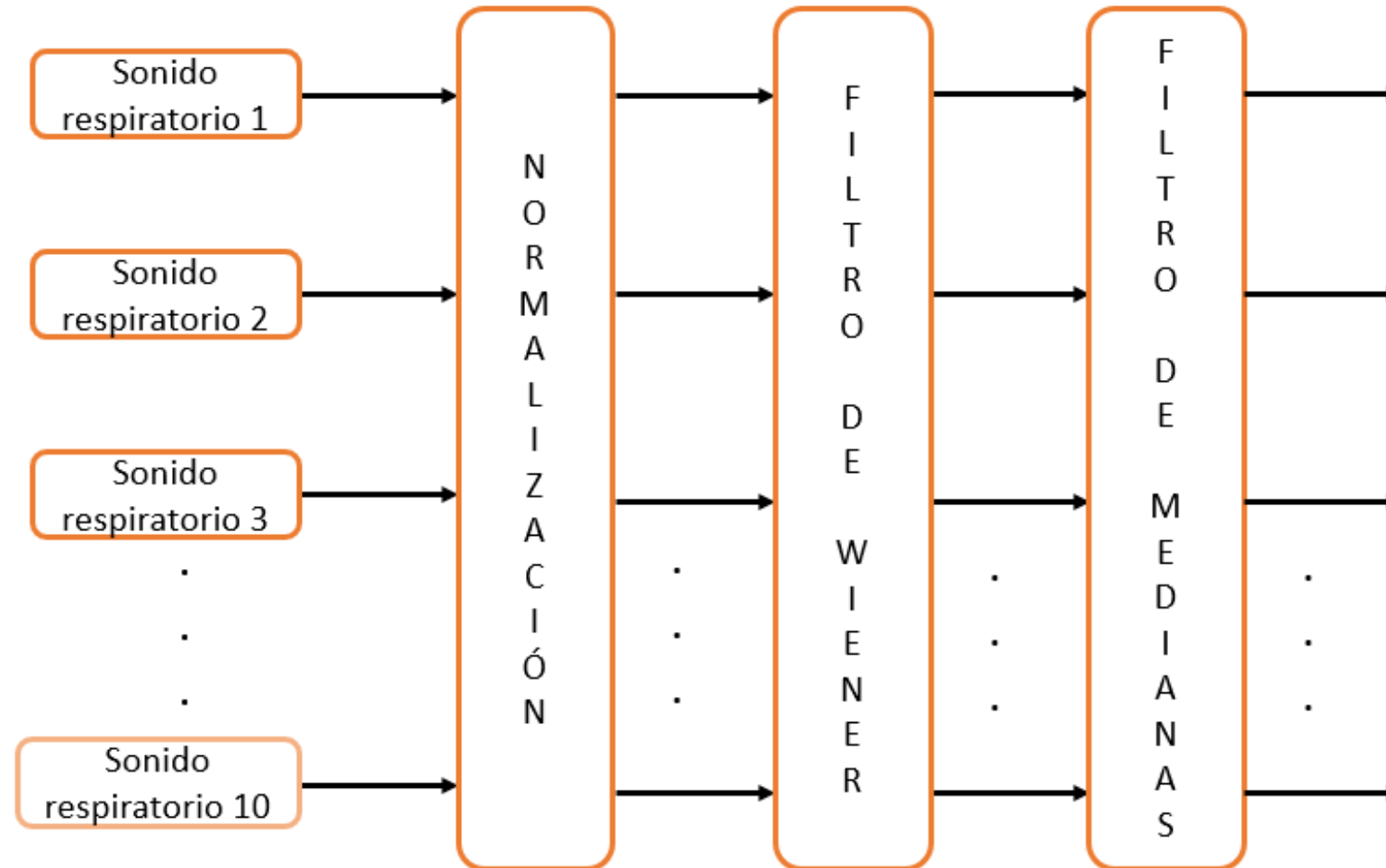


Figura 8. Etapas del preprocesamiento.

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \in [0,1]$$

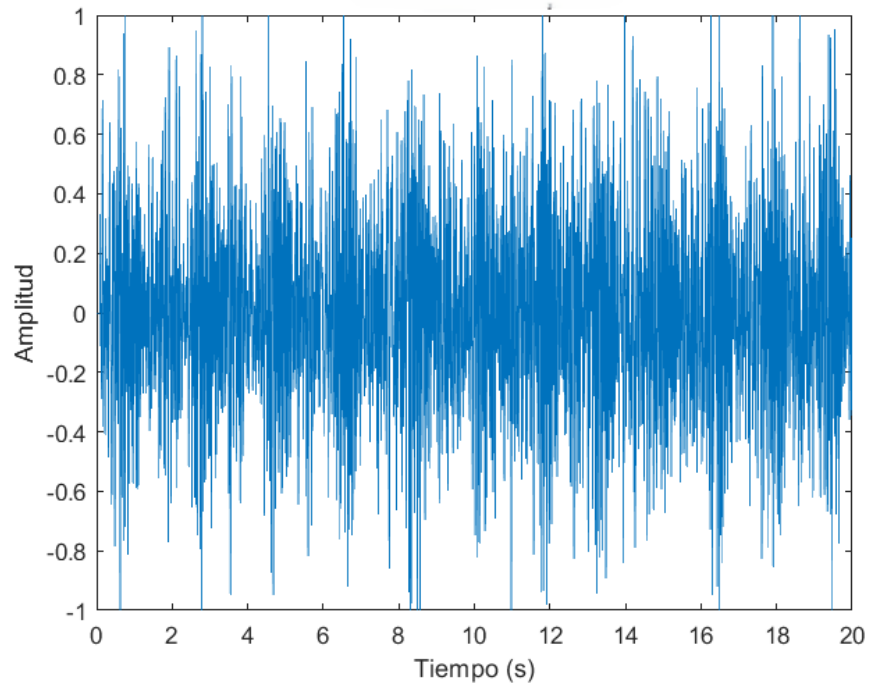


Figura 9. Sonido respiratorio sin normalizar

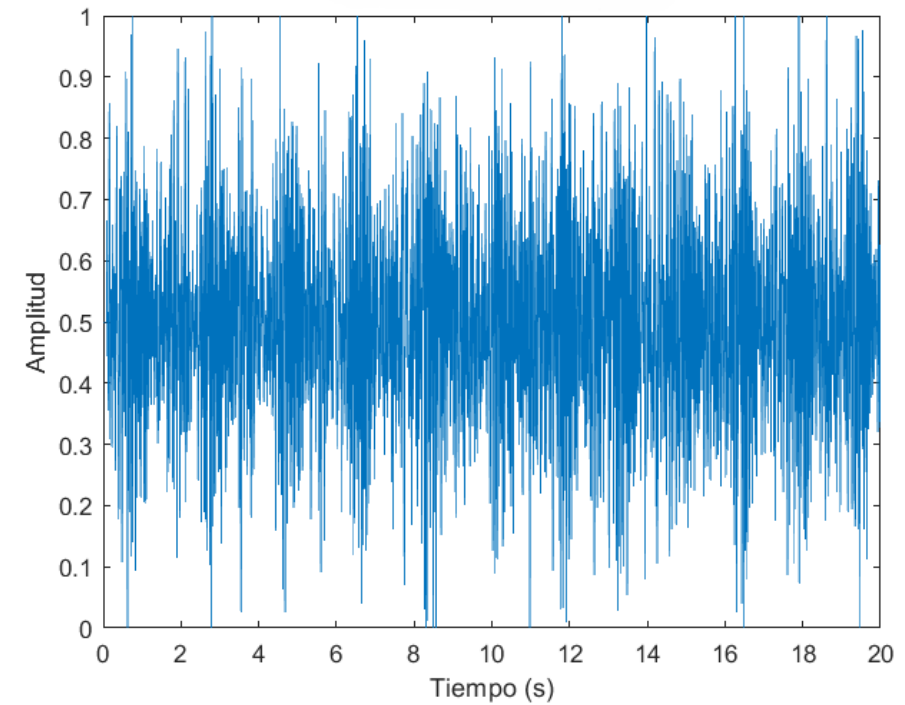


Figura 10. Sonido respiratorio normalizado

$$SNR_{post}(p, k) = \frac{|X(p, k)|^2}{\sum[|N(p, k)|^2]}$$

$$SNR_{prio}(p, k) = \frac{\sum[|S(p, k)|^2]}{\sum[|N(p, k)|^2]}$$

$$SNR_{inst}(p, k) = \frac{|X(p, k)|^2 - \sum[|N(p, k)|^2]}{\sum[|N(p, k)|^2]}$$

$$G(p, k) = g(SNR_{prio}(p, k), SNR_{post}(p, k))$$

$$S(p, k) = G(p, k)X(p, k)$$

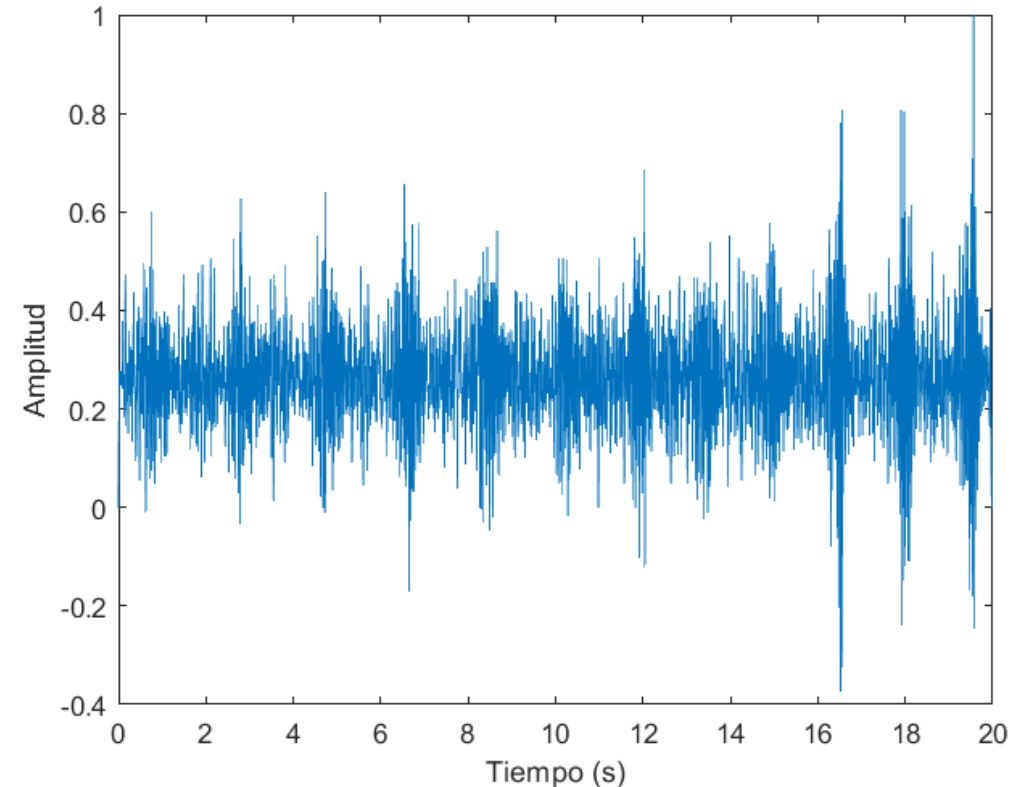


Figura 11. Sonido respiratorio una vez aplicado el filtro de Wiener.



Figura. Señal de audio antes del preprocesamiento



Figura. Señal de audio después del preprocesamiento

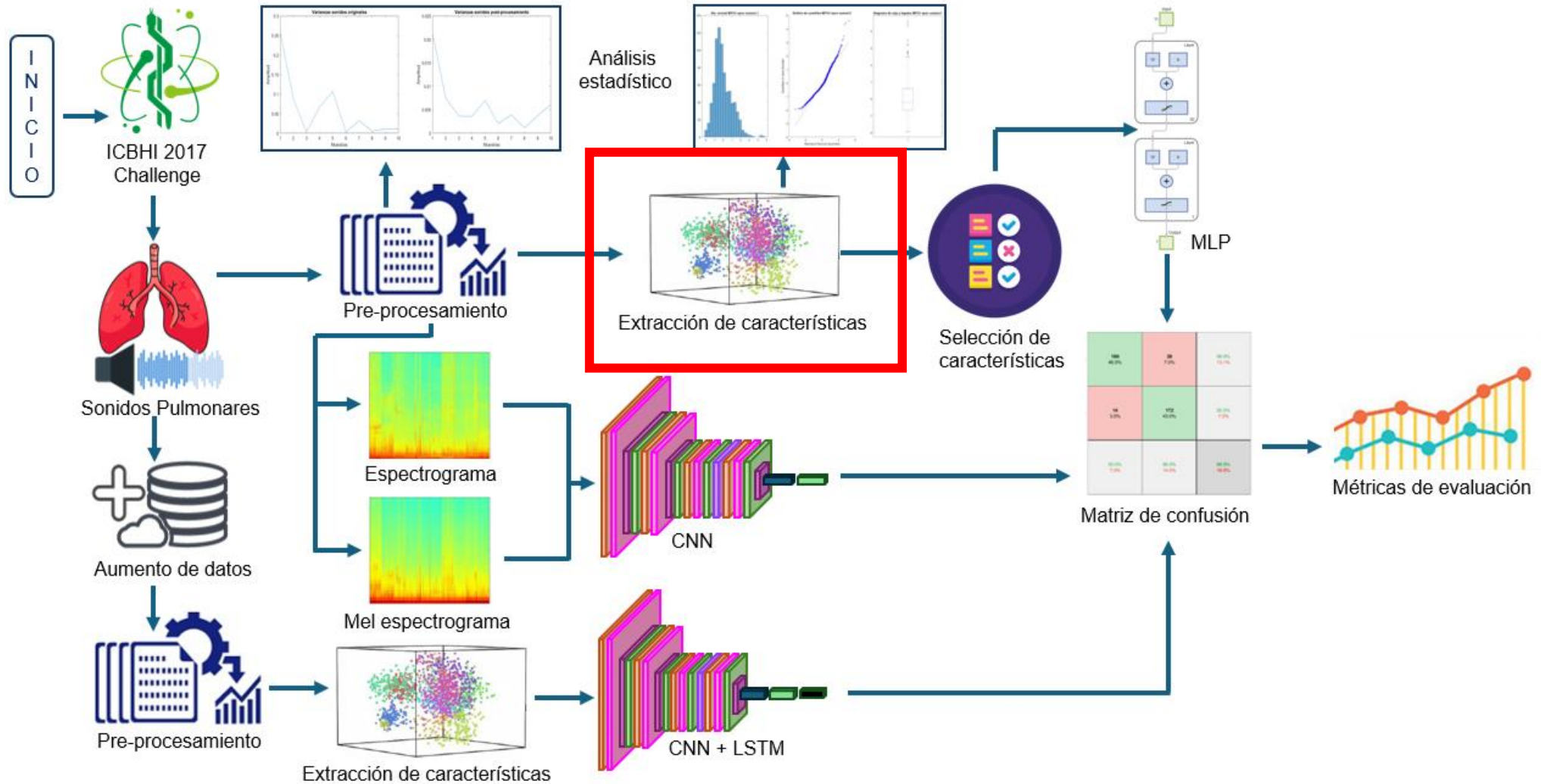


Figura 12. Metodología propuesta para la clasificación de señales de audio (**Extracción de características**).

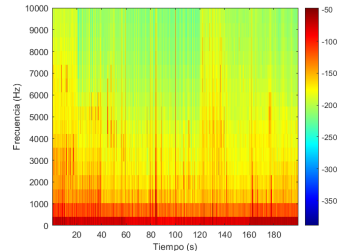
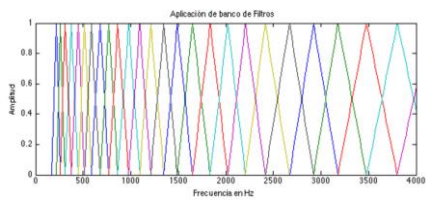
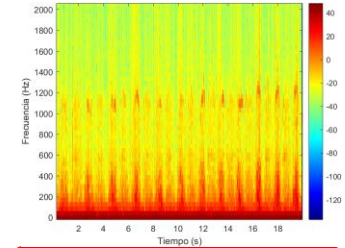
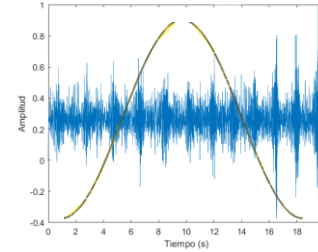
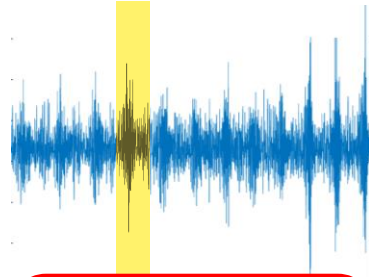
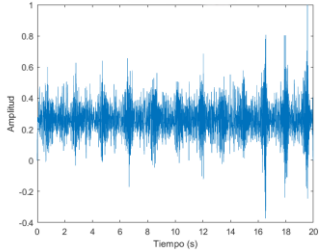
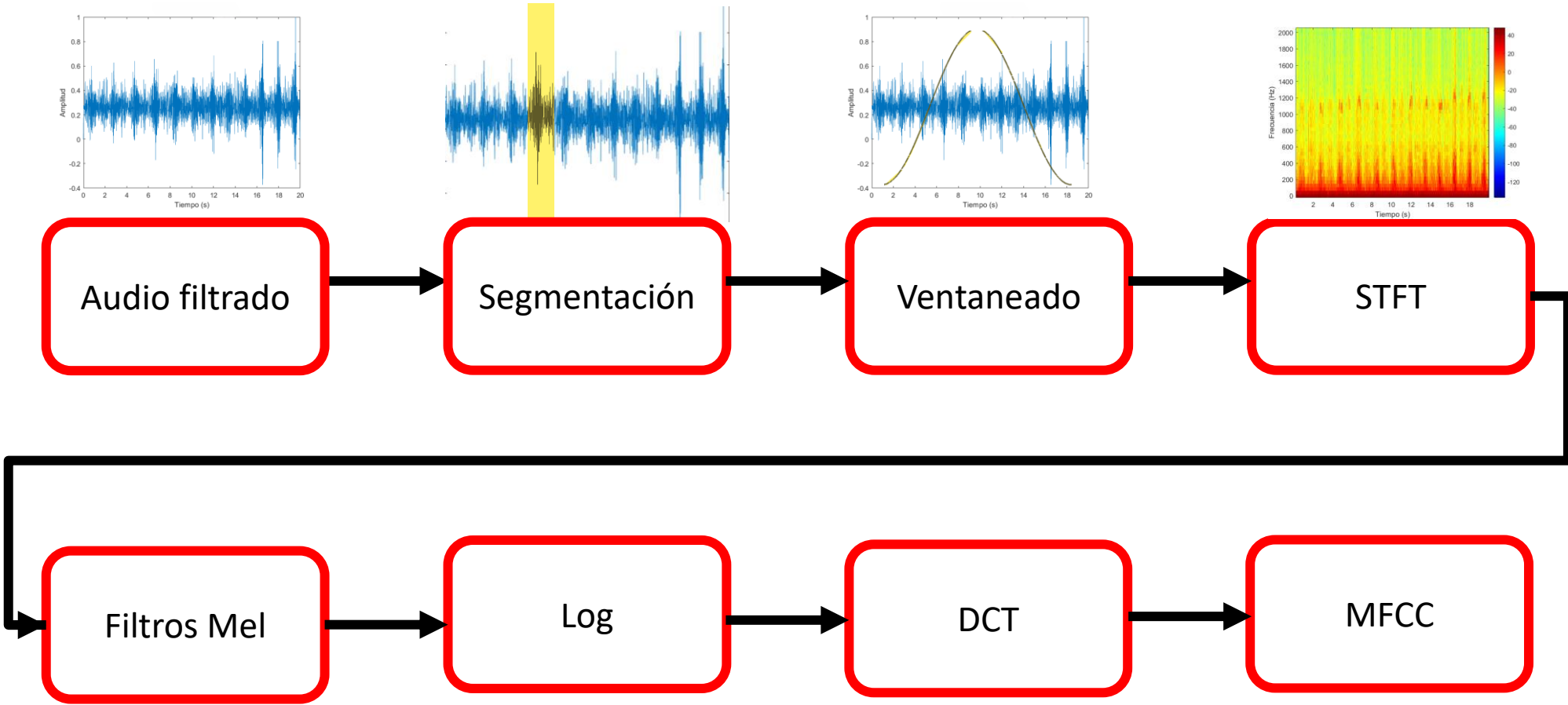
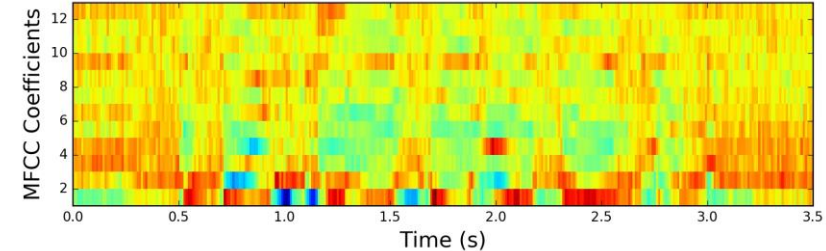


Figura 13. Proceso de extracción de características



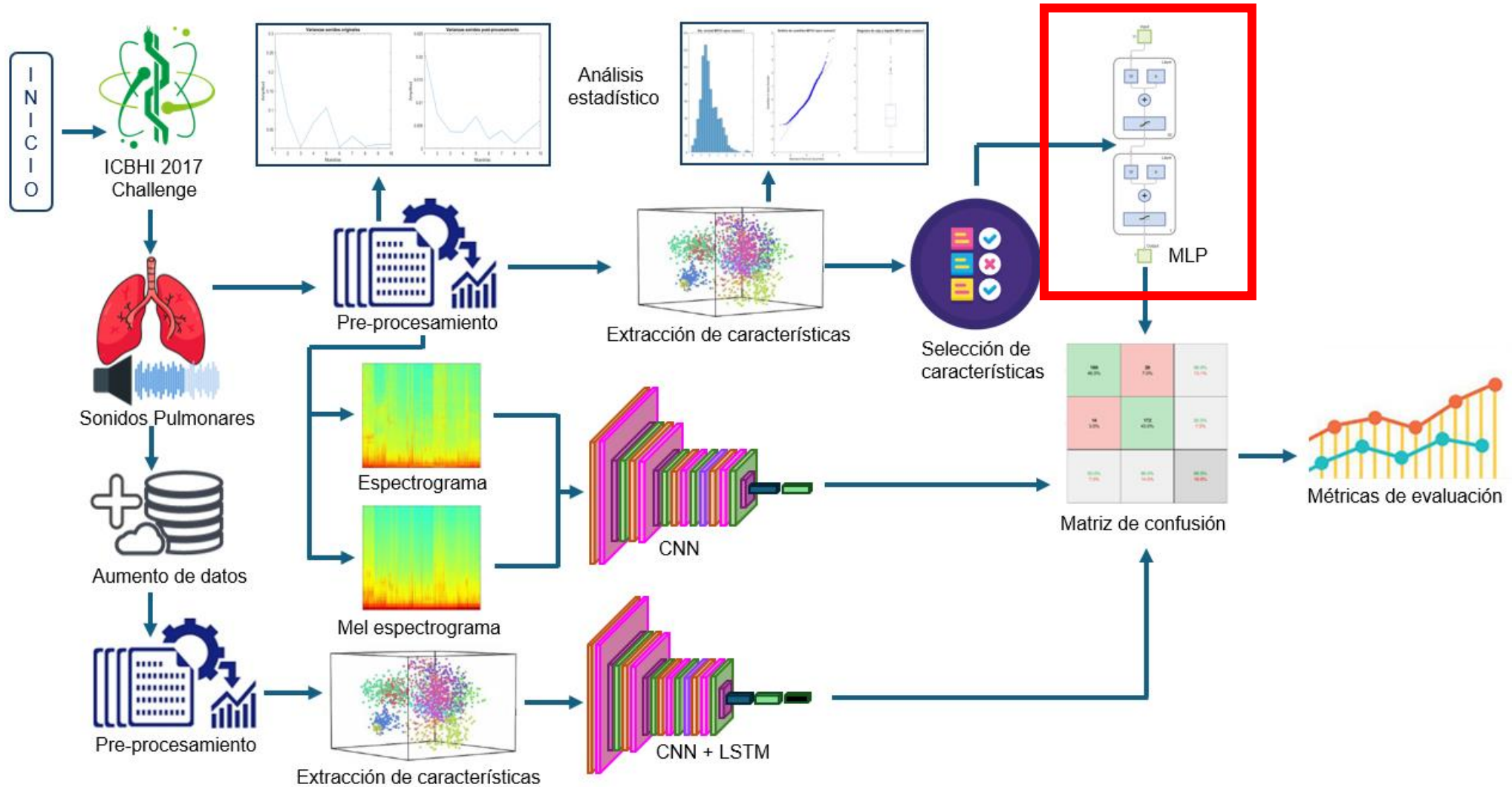


Figura 14. Metodología propuesta para la clasificación de señales de audio (Modelo MLP).

Arquitectura Perceptrón Multicapa (MLP)

Capa de entrada	Capa oculta	Capa de salida	Funciones de activación	Función de aprendizaje	Épocas	Tasa de aprendizaje
MFCC ($m \times n$)	30 neuronas	1 neurona	tansig, logsig	trainlm	300	0.001

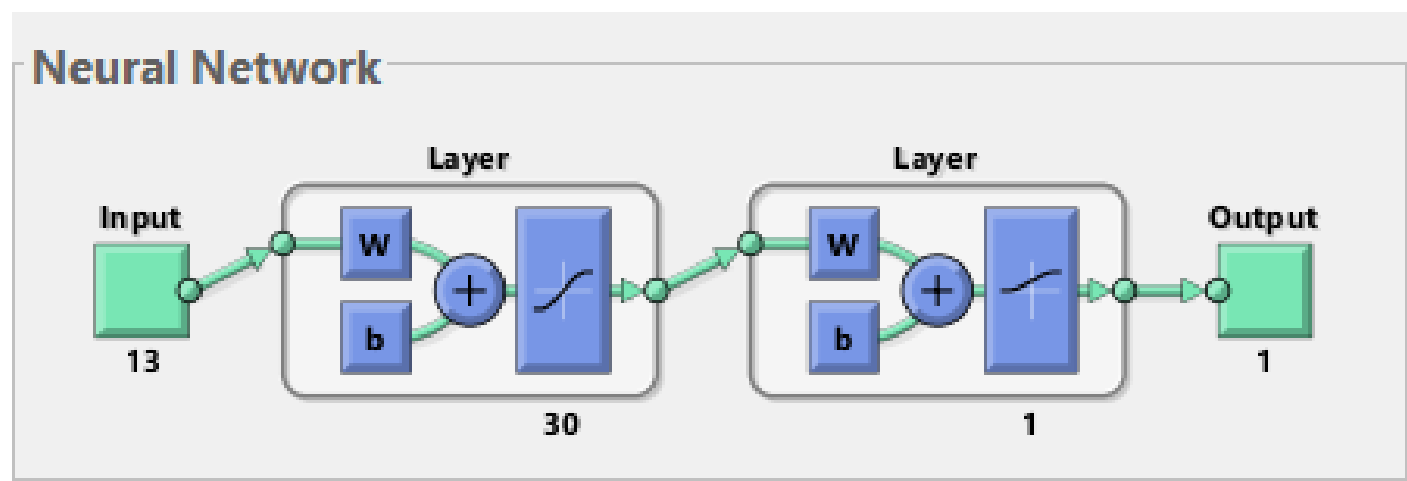


Figura 15. Arquitectura para el modelo MLP.

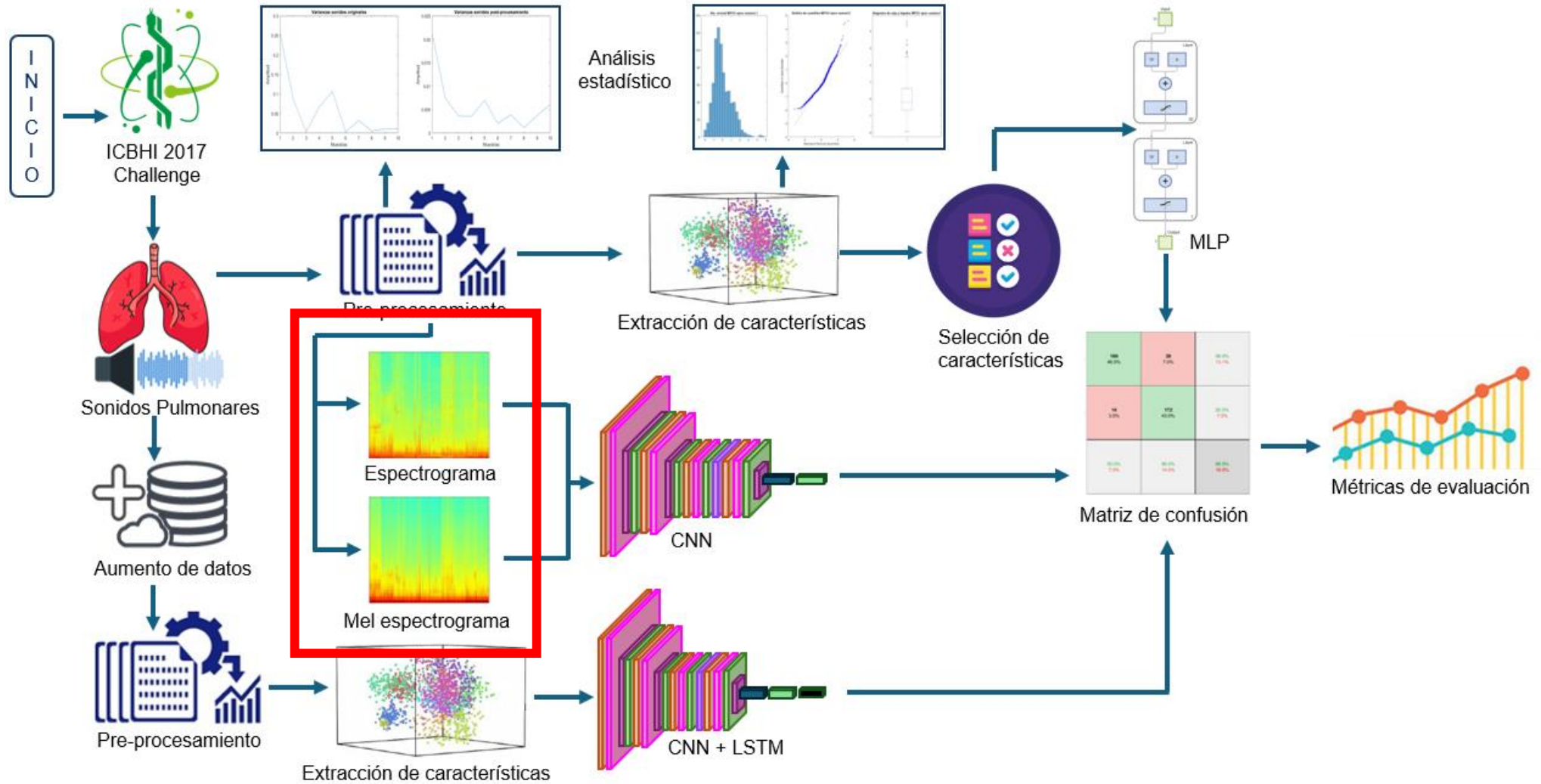


Figura 16. Metodología propuesta para la clasificación de señales de audio (**Espectrogramas**).

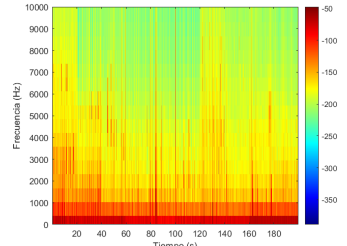
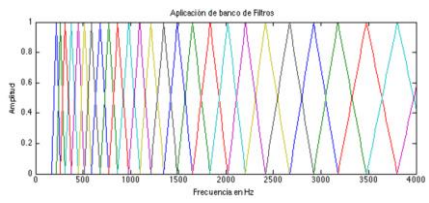
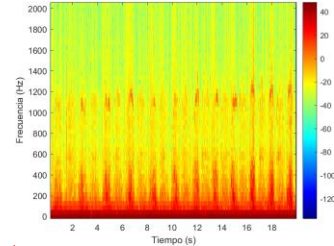
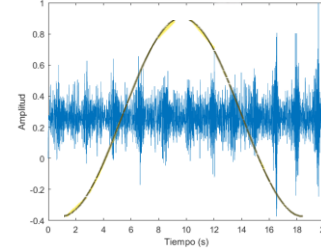
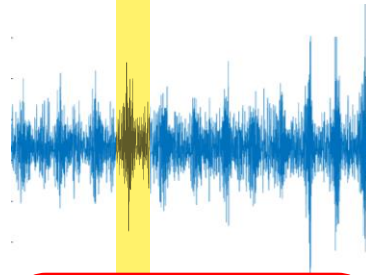
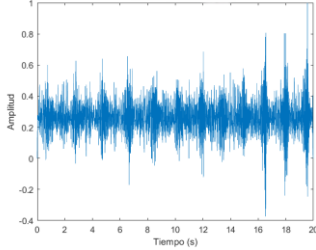
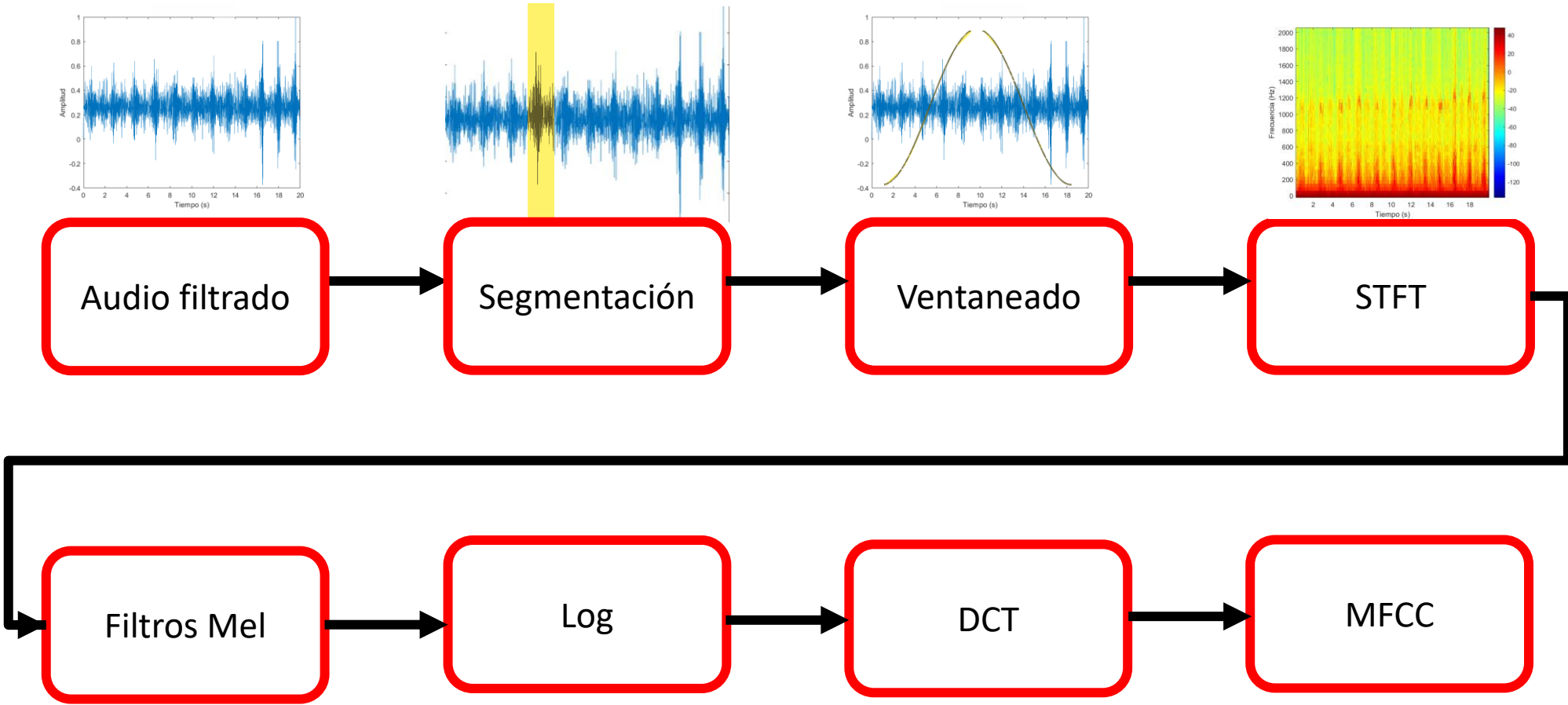
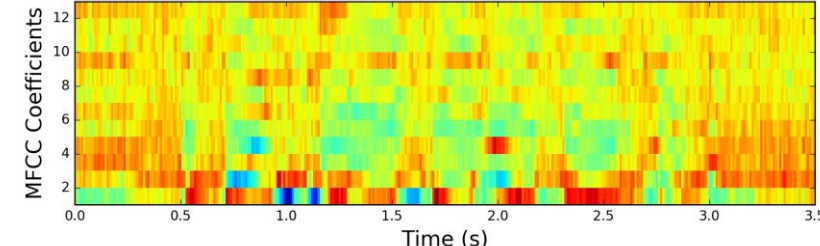


Figura 17. Proceso de extracción de características



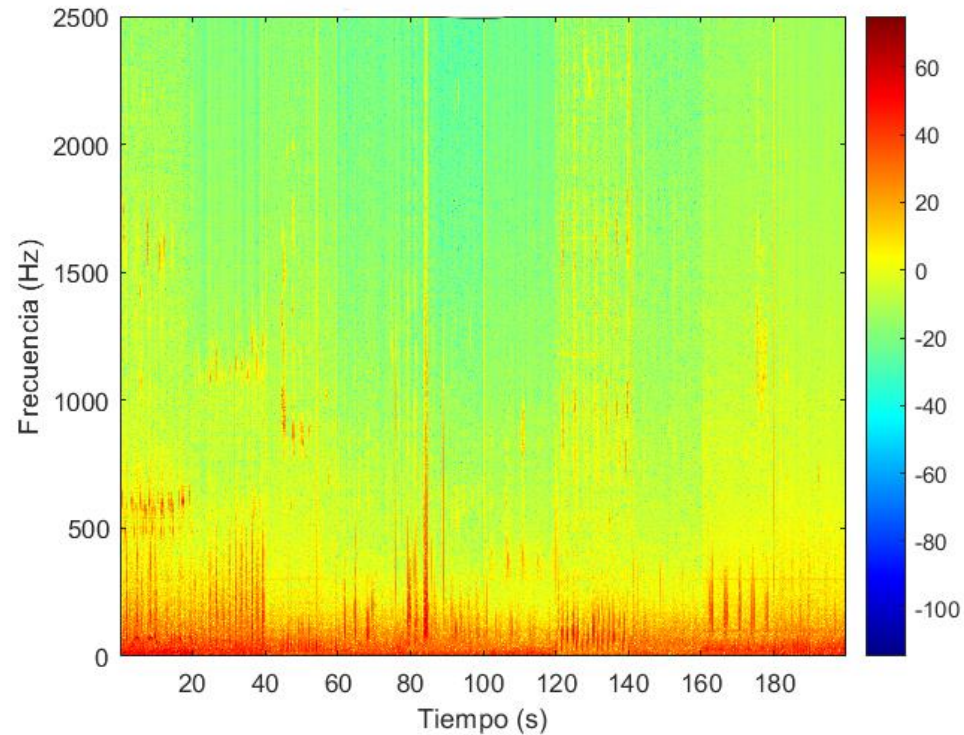


Figura 18. Espectrograma completo de los sonidos respiratorios sanos

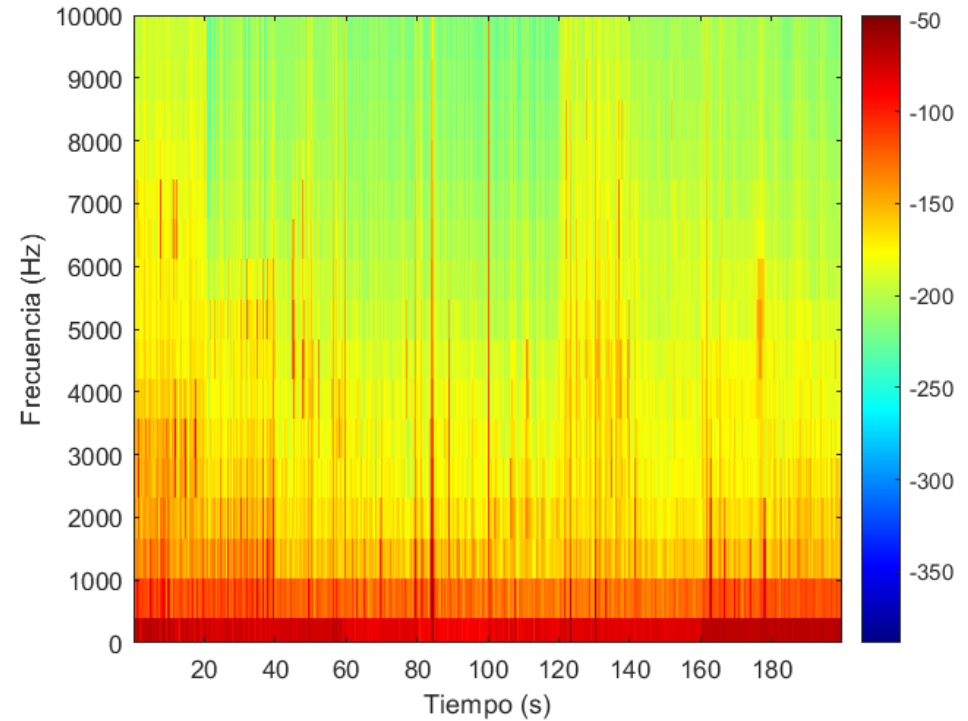


Figura 19. Mel-espectrograma completo de los sonidos respiratorios sanos

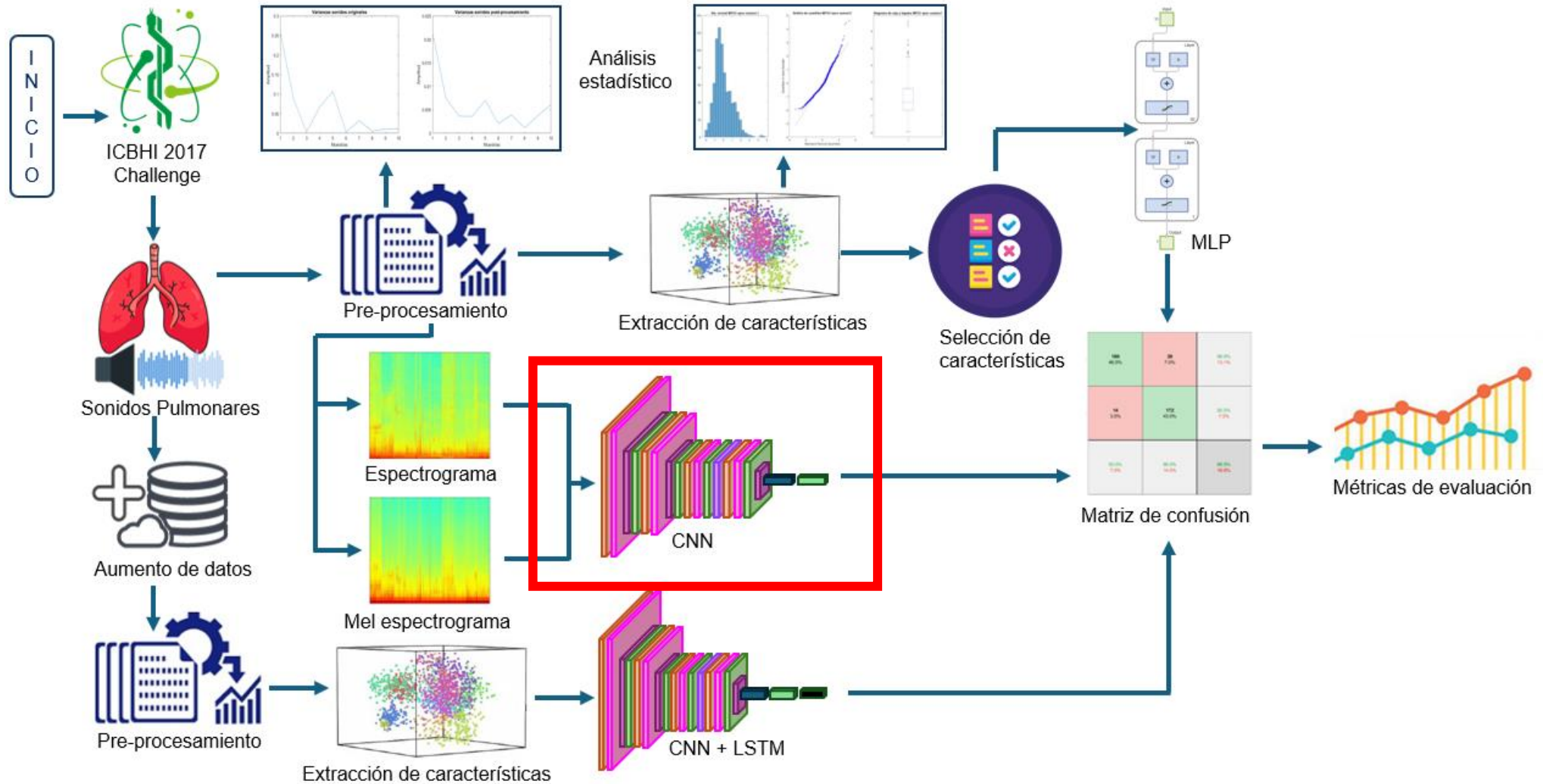


Figura 20. Metodología propuesta para la clasificación de señales de audio (**Modelo CNN**).

Arquitectura Red Convolucional (CNN)

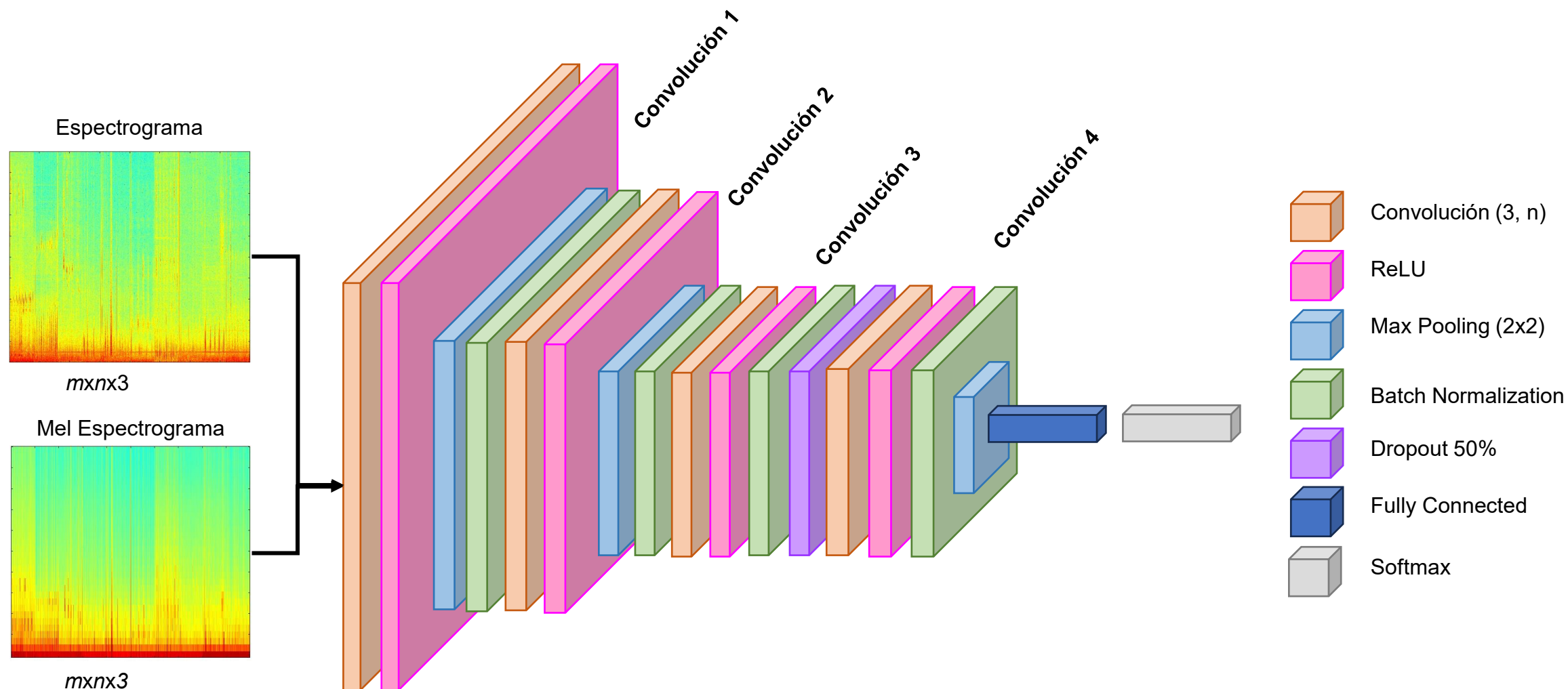


Figura 21. Arquitectura para el modelo CNN

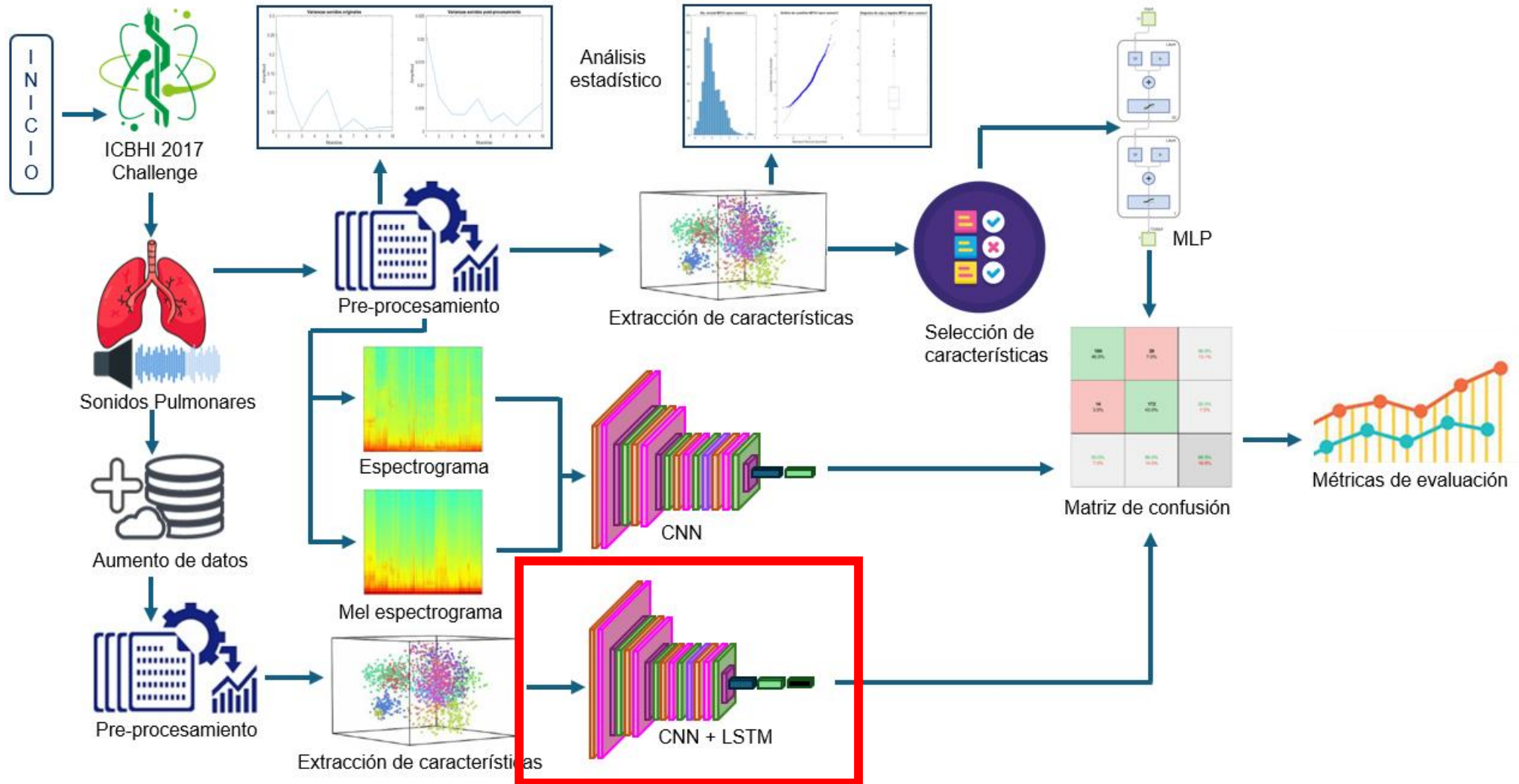


Figura 22. Metodología propuesta para la clasificación de señales de audio (Modelo CNN+LSTM).

Arquitectura Red Convolucional + LSTM (CNN+LSTM)

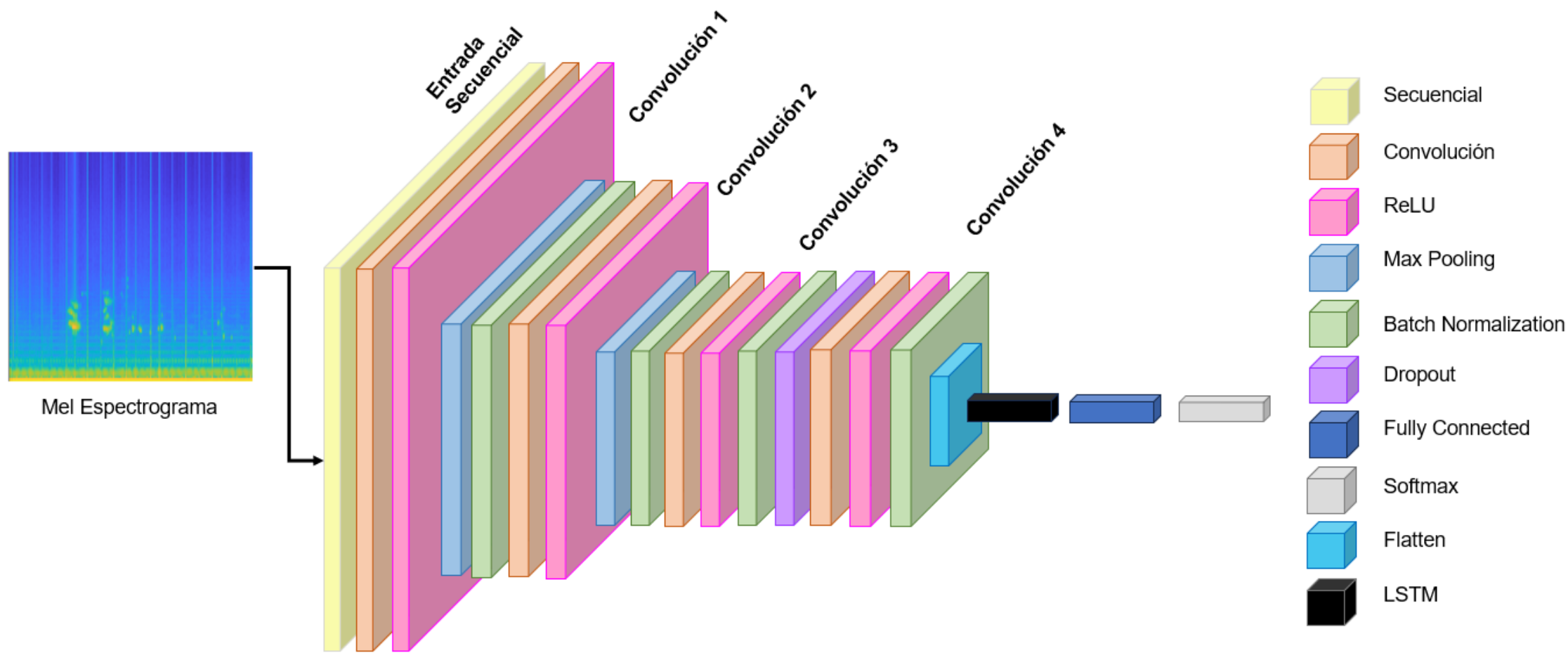


Figura 23. Arquitectura para el modelo CNN+LSTM

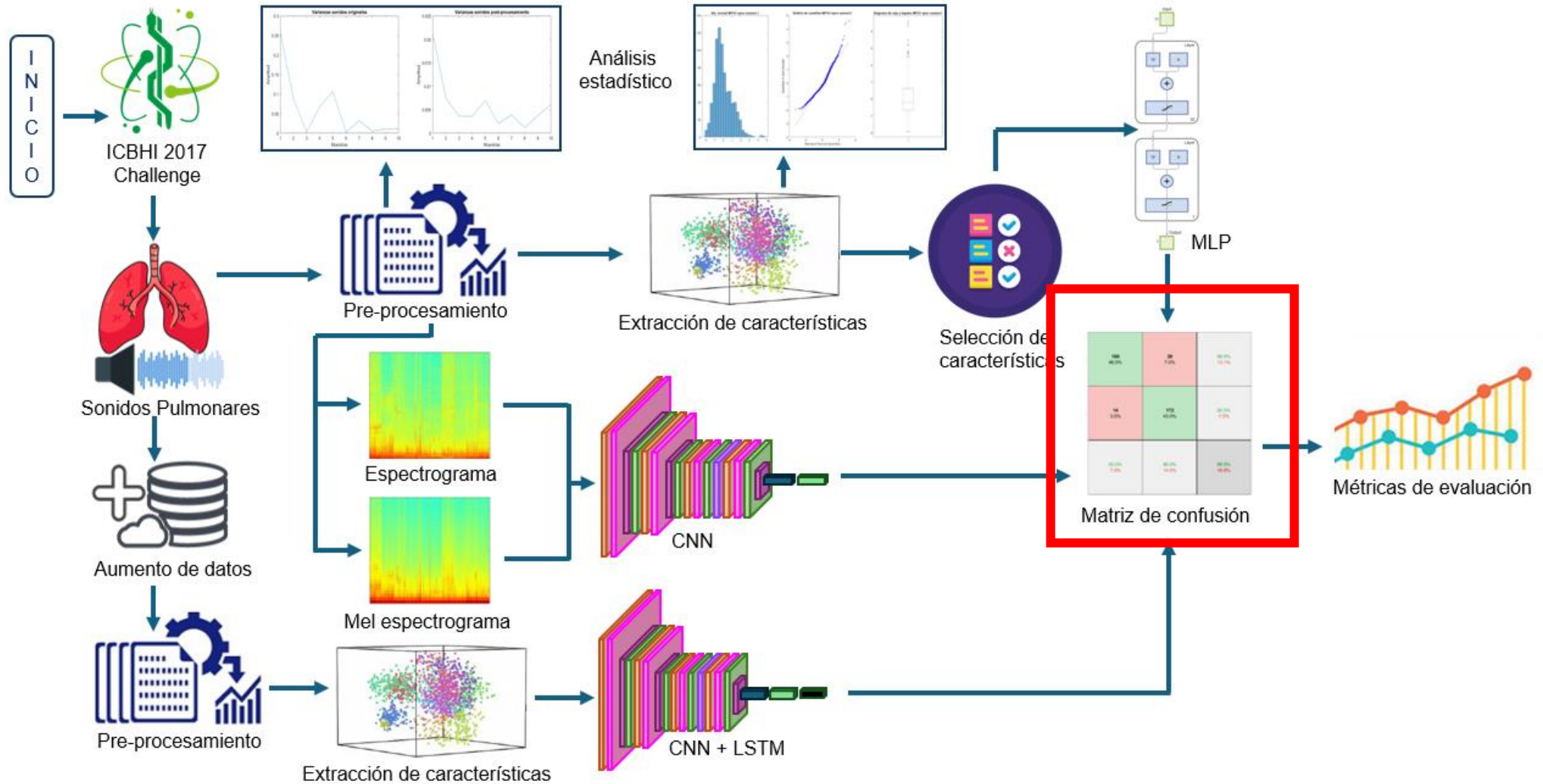


Figura 24. Metodología propuesta para la clasificación de señales de audio (Matriz de confusión).



Matriz de confusión modelo MLP

Matriz de Confusión

Etiqueta de Salida	SANOS	184 46.0% VN	24 6.0% FN	88.5% 11.5% Sensibilidad (SEN)
	EPOC	16 4.0% FP	176 44.0% VP	91.7% 8.3% Especificidad (ESP)
		92.0% 8.0% Precisión (PRC)	88.0% 12.0% VPN	90.0% 10.0% Exactitud (ACC)
	SANOS		EPOC	
		Etiqueta Verdadera		

Figura 25. Matriz de confusión modelo de red MLP



Matriz de confusión modelo CNN

True Class	Epoc	296 VP	24 FP
	Sanos	60 FN	260 VN
		Epoc	Sanos
		Predicted Class	

Figura 26. Matriz de confusión modelo de red CNN



Matriz de confusión modelo CNN+LSTM

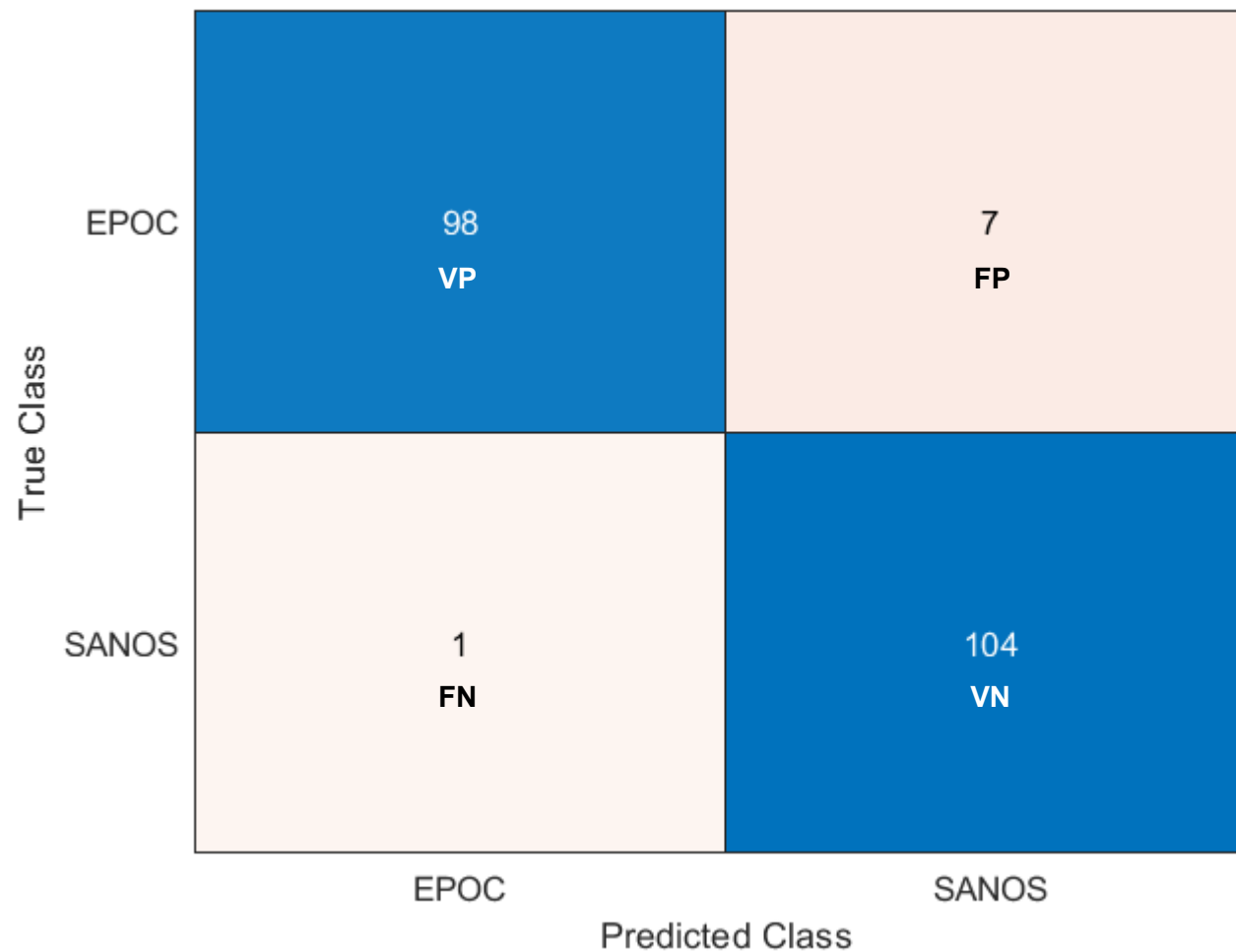


Figura 27. Matriz de confusión modelo de red CNN+LSTM

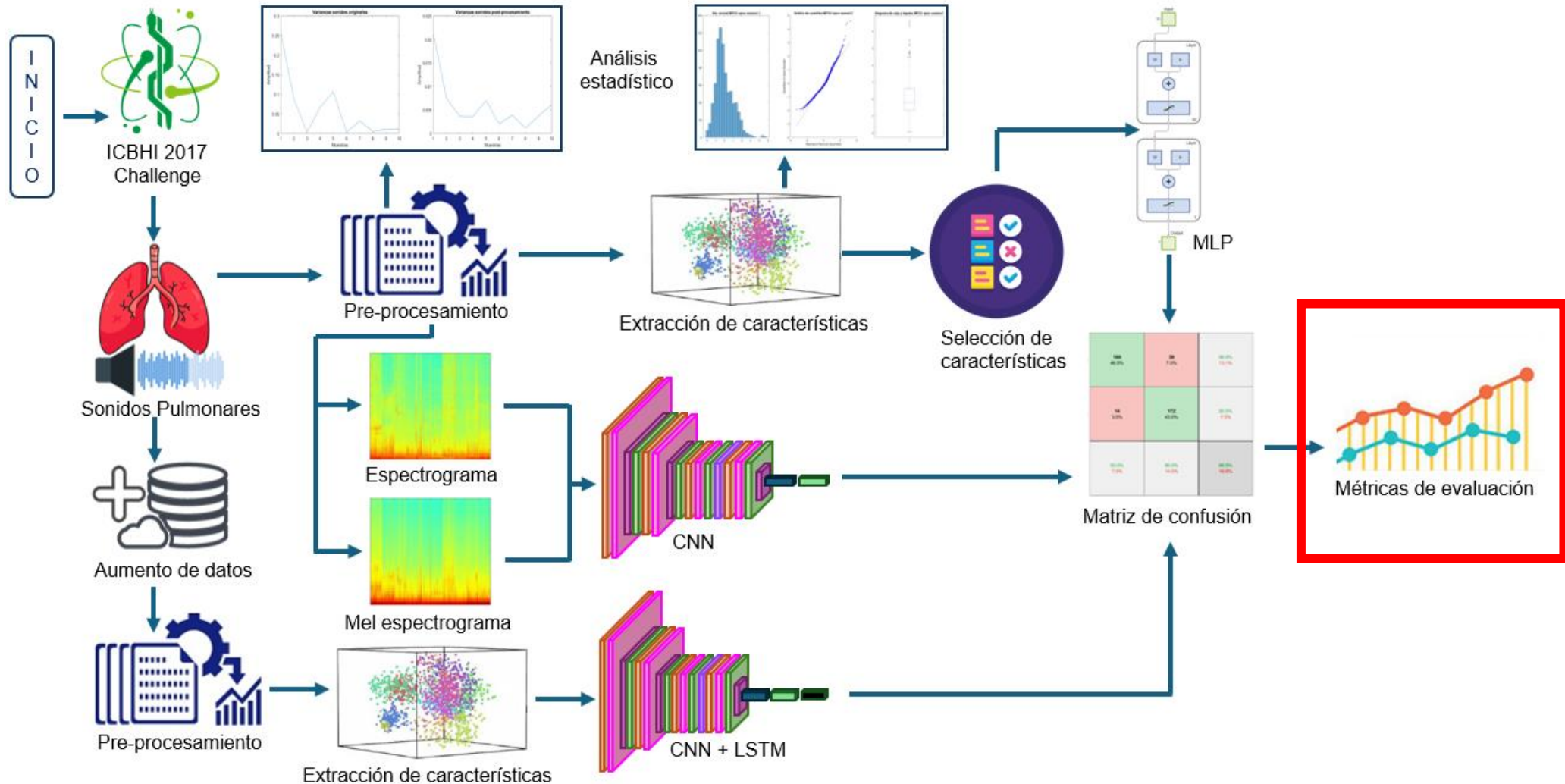


Figura 28. Metodología propuesta para la clasificación de señales de audio (Métricas de evaluación).



Métricas

Métrica	Ecuación
Exactitud (ACC)	$\frac{VP + VN}{VP + VN + FP + FN}$
Precisión (PRC)	$\frac{VP}{VP + FP}$
Sensibilidad (SEN)	$\frac{VP}{VP + FN}$
Especificidad (ESP)	$\frac{VN}{VN + FP}$
F1-Score (F1)	$2 * \frac{PRC * SEN}{PRC + SEN}$

Figura 29. Tabla con las métricas utilizadas para evaluar los modelos de RNA



Métricas

Métrica	Ecuación
Valor Predictivo Negativo (VPN)	$\frac{VN}{VN + FN}$
Media Geométrica (GM)	$\sqrt{SEN * ESP}$
Coefficiente de correlación de Matthews (MCC)	$\frac{(VP * VN) - (FP * FN)}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$
Coefficiente de correlación de Matthews Normalizado (MCCn)	$\frac{1 + MCC}{2}$
Coefficiente de Kappa de Cohen (MKn)	$VPP + VPN - 1$

Figura 30. Tabla con las métricas utilizadas para evaluar los modelos de RNA



Comparativa de los mejores resultados obtenidos

Arq. RNA	Métricas										
	VACC	ACC	PRC	SEN	ESP	F1	GM	MCC	MCCn	MKn	VPN
MLP (MFCC)	N/A	90.00	92.00	88.46	91.67	90.20	90.05	80.06	90.03	80.00	88.00
MLP (Delta)	N/A	84.58	85.83	83.74	85.47	84.77	84.60	69.19	84.59	69.17	83.33
MLP (Delta-Delta)	N/A	84.50	86.00	83.50	85.57	84.73	84.52	69.03	84.52	69.00	83.00
CNN	87.50	88.13	91.88	85.47	91.22	88.55	88.29	76.47	88.23	76.25	84.38
CNN+LSTM	97.12	96.19	93.33	98.99	93.69	96.08	96.31	82.23	91.12	92.38	99.07

Figura 31. Tabla con los mejores resultados obtenidos de cada modelo de red



MARVID®

© MARVID-Mexico

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162, 163 fraction I, 164 fraction I, 168, 169, 209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BMARVID is part of the media of MARVID-Mexico., E: 94-443.F: 008- (www.marvid.org/booklets)