Tecnológico de Monterrey

Proyecto Integrador (Gpo. 10)

Avance 5. Modelo Final.

Equipo #6

Julio César Pérez Zapata	A01793880
Christian Emilio Saldaña López	A00506509
Jorge Estivent Cruz Mahecha	A01793808

Modelo Binario

El modelo final definitivo se basará en la arquitectura Sincnet, no solo por los resultados obtenidos en cuanto a la métrica de precisión, sino también debido a que esta red neuronal ha sido diseñada específicamente para el procesamiento de señales de audio, como el habla.

Sincnet utiliza filtros de paso de banda diseñados específicamente para capturar características del habla, como formantes y transiciones del sonido, que resultan ser útiles para nuestro propósito. Al mismo tiempo, las capas convolucionales posteriores pueden capturar características globales y abstractas.

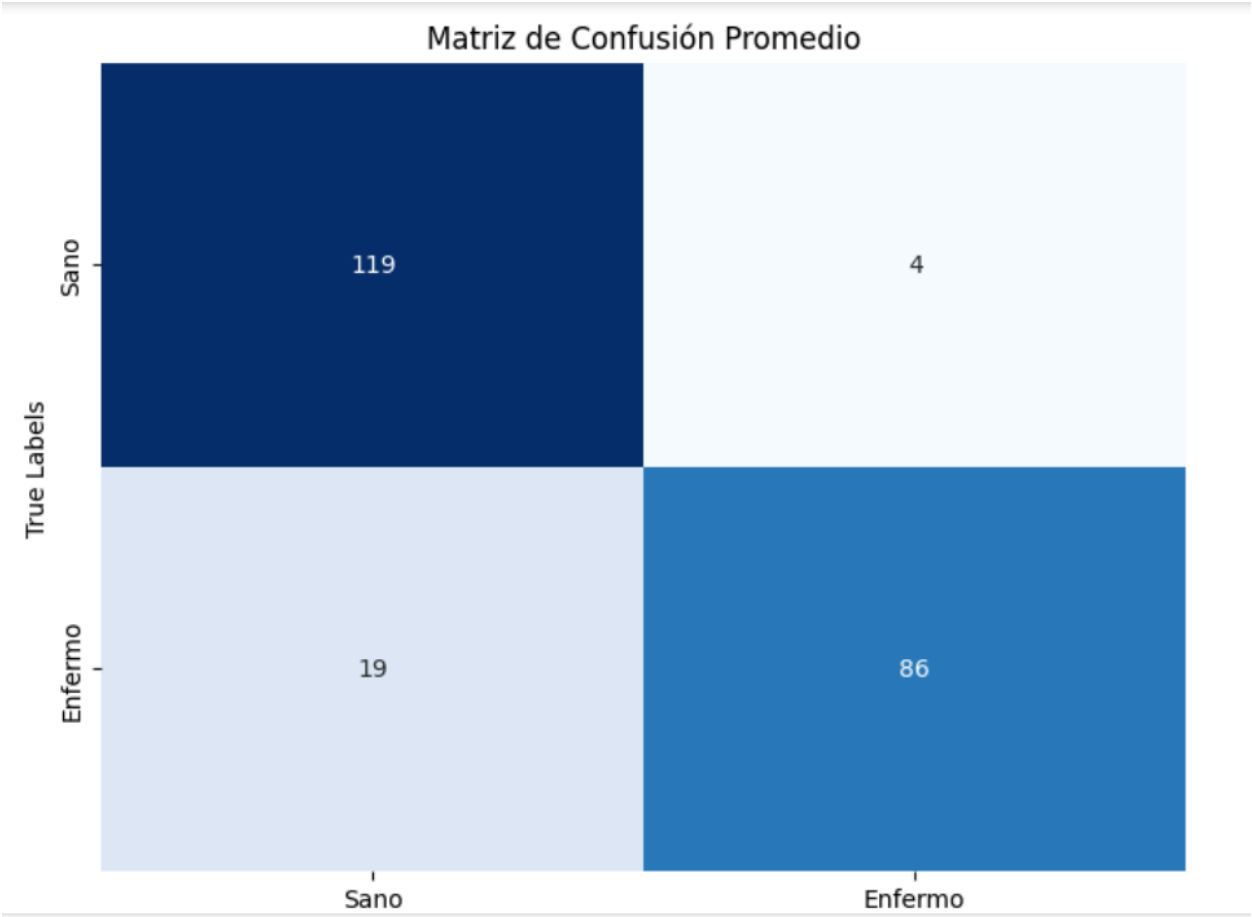
Una ventaja adicional radica en el uso de filtros sinc en lugar de los convencionales. Esto permite que SincNet aprenda representaciones más eficientes y específicas para las tareas de procesamiento de señales de audio. Este enfoque puede traducirse en una mejora notable en la generalización y el rendimiento en comparación con las arquitecturas convencionales.

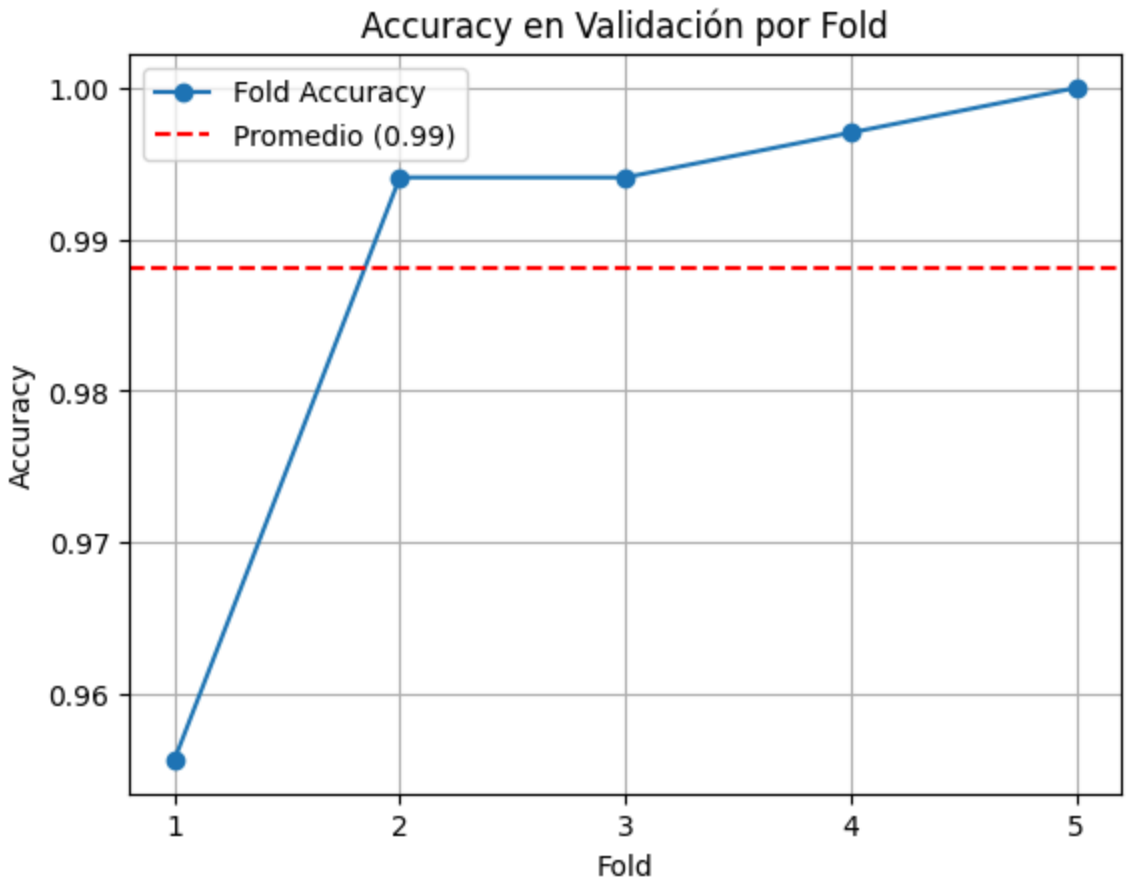
Comparativos de los modelos anteriores

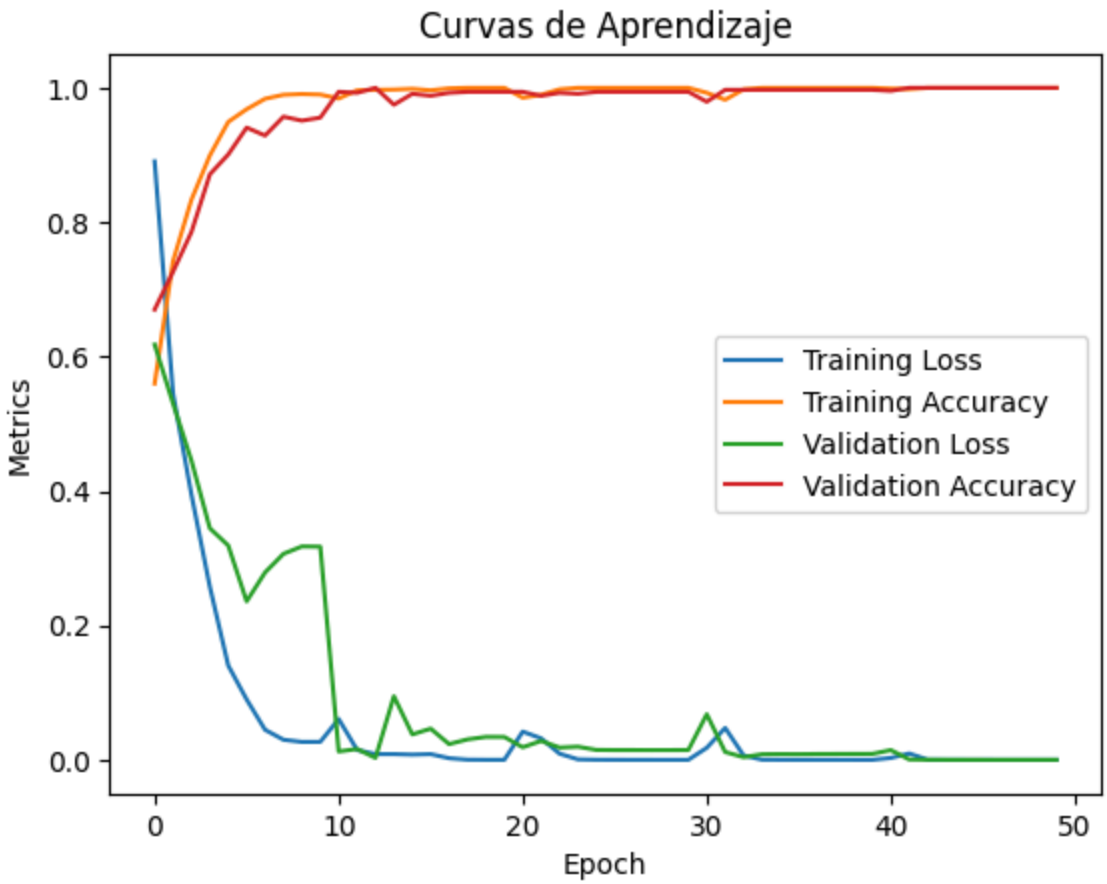
-

Modelo	F1-Score	Recall	Accuracy	Training (Min)	Data Augmentation
Regresion Logistica - Binaria	0.67	0.70	0.72	1	si
CNN MFCC- Binaria	0.67	0.62	0.62	20	si
SincNet sin balanceo - Binaria	0.64	0.62	0.65	40	no
SincNet con data augmentation y balanceo - Binaria	0.98	0.98	0.98	53	si

SincNet - Multi clase	0.67	0.58	0.66	10.3	no
CNN - Multi clase	0.65	0.58	0.59	9	no
Mfcc-Conv2D	0.60	0.64	0.61	5	si



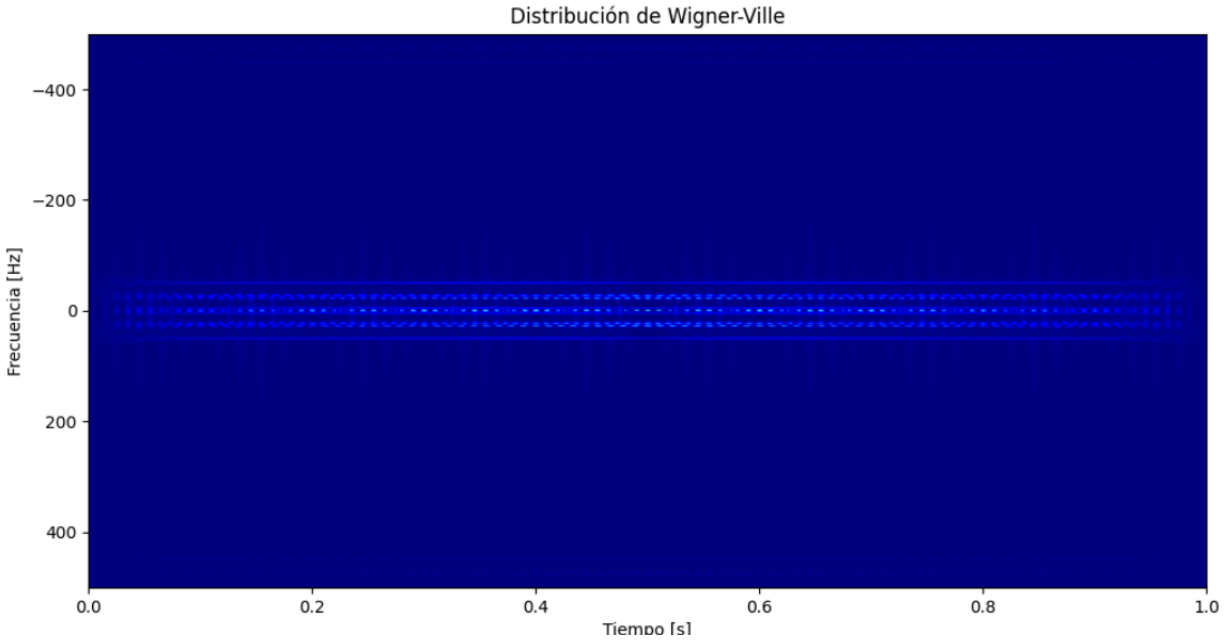

Accuracy en Validación por Fold
Fold Accuracy
Promedio (0.99)
1.00
0.99
0.98
0.97
0.96
Accuracy
1
2
3
4
5
Fold

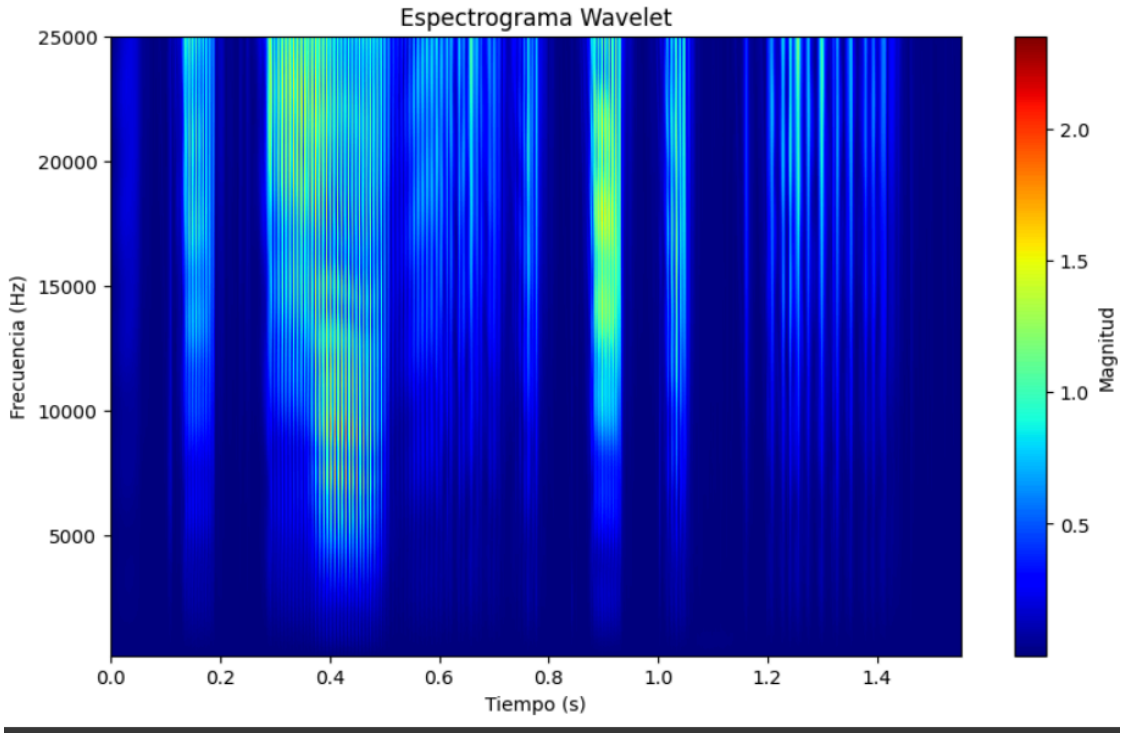

Curvas de Aprendizaje
Metrics
Epoch
Training Loss
Training Accuracy
Validation Loss
Validation Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50

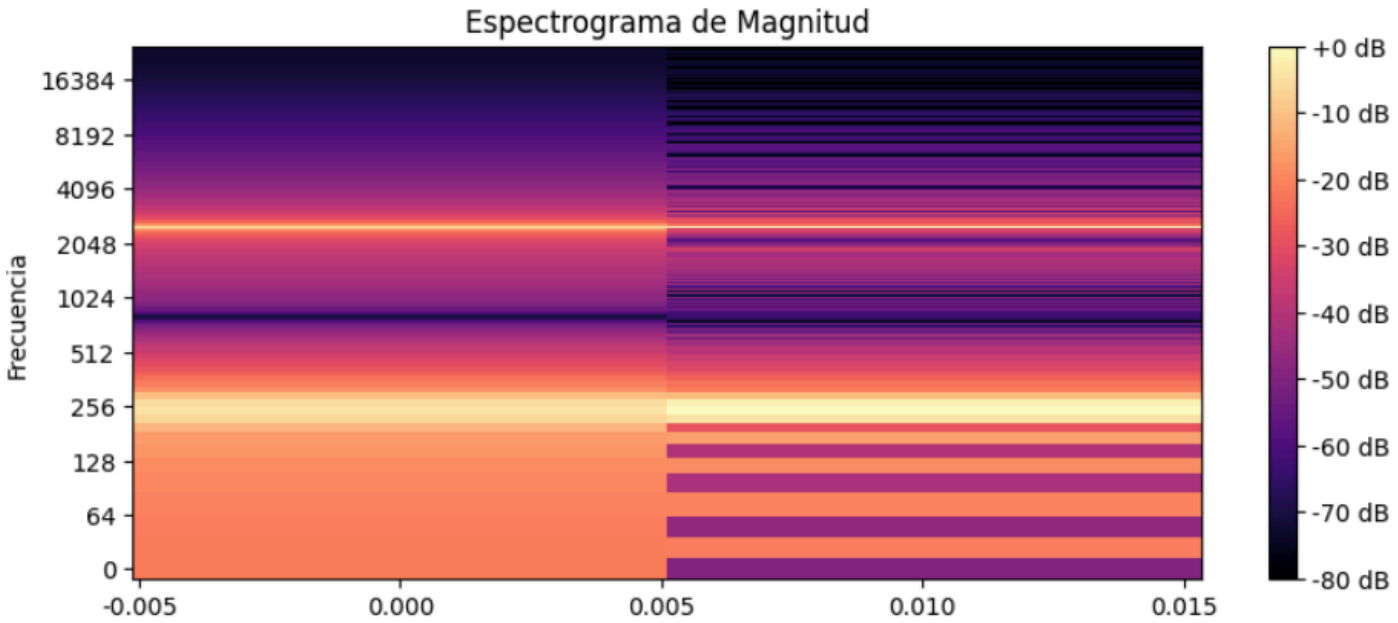
Modelo Multiclase

Durante el proceso de desarrollo del modelo multiclase, se llevaron a cabo diversas pruebas para optimizar el tratamiento de los datos y mejorar significativamente la evaluación del modelo. En la fase final de este proceso, se exploró la inclusión de información adicional extraída de los archivos de audio del conjunto de datos. Esto se realizó con el propósito de enriquecer los datos disponibles para el entrenamiento de la red neuronal.

Entre las técnicas utilizadas se encuentran la generación de diferentes espectrogramas y análisis de frecuencia de los datos originales. Estos incluyen la distribución de Wigner Ville, la Transformada de Wavelet y la Transformada de Fourier de Tiempo Corto, que puede ser representada en forma de espectrograma de magnitud. La integración de estos análisis a una red neuronal convolucional resulta especialmente efectiva debido a la capacidad inherente de estas redes para procesar información visual. Al incorporar directamente estos espectrogramas, se fortalece significativamente el desempeño del modelo.



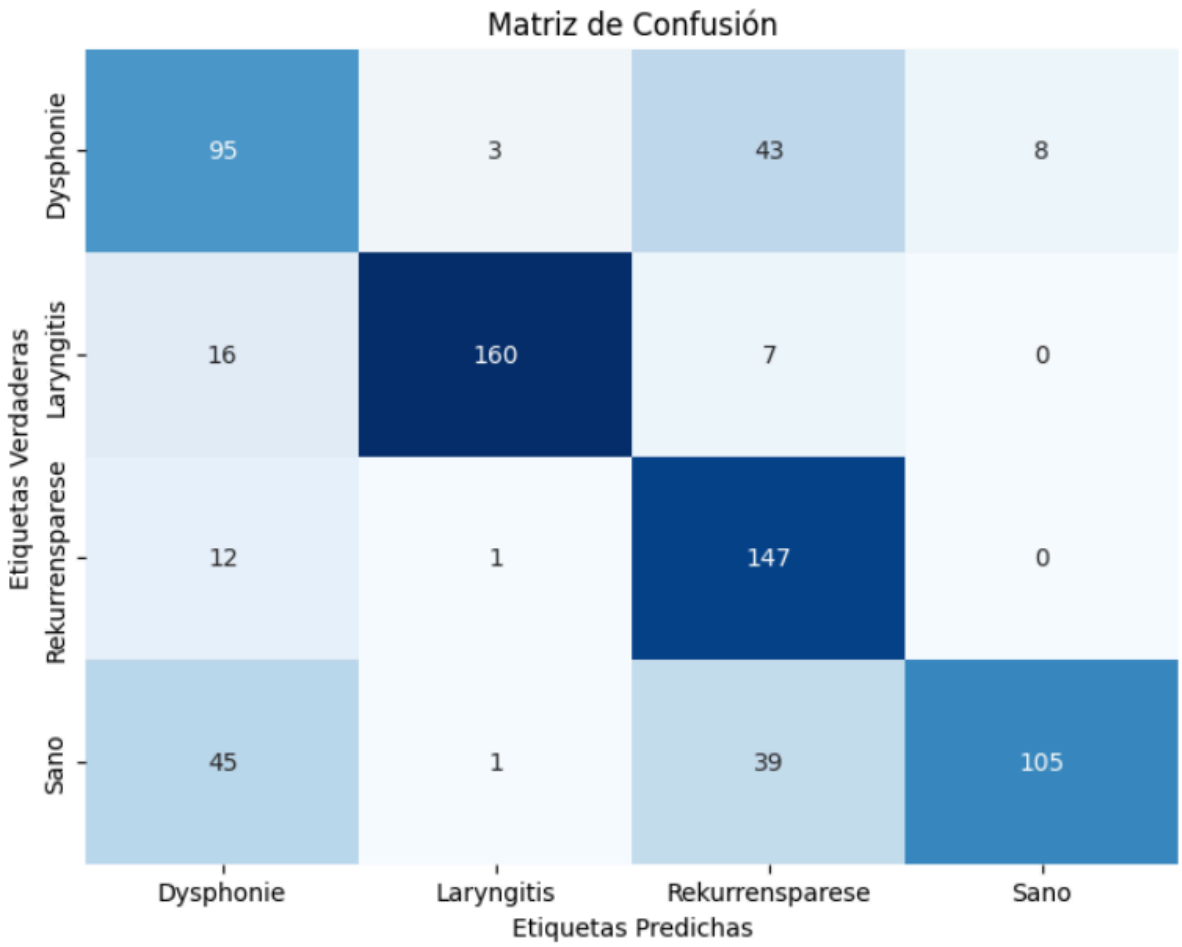




Se procede a realizar esta integración obteniendo los siguientes resultados:

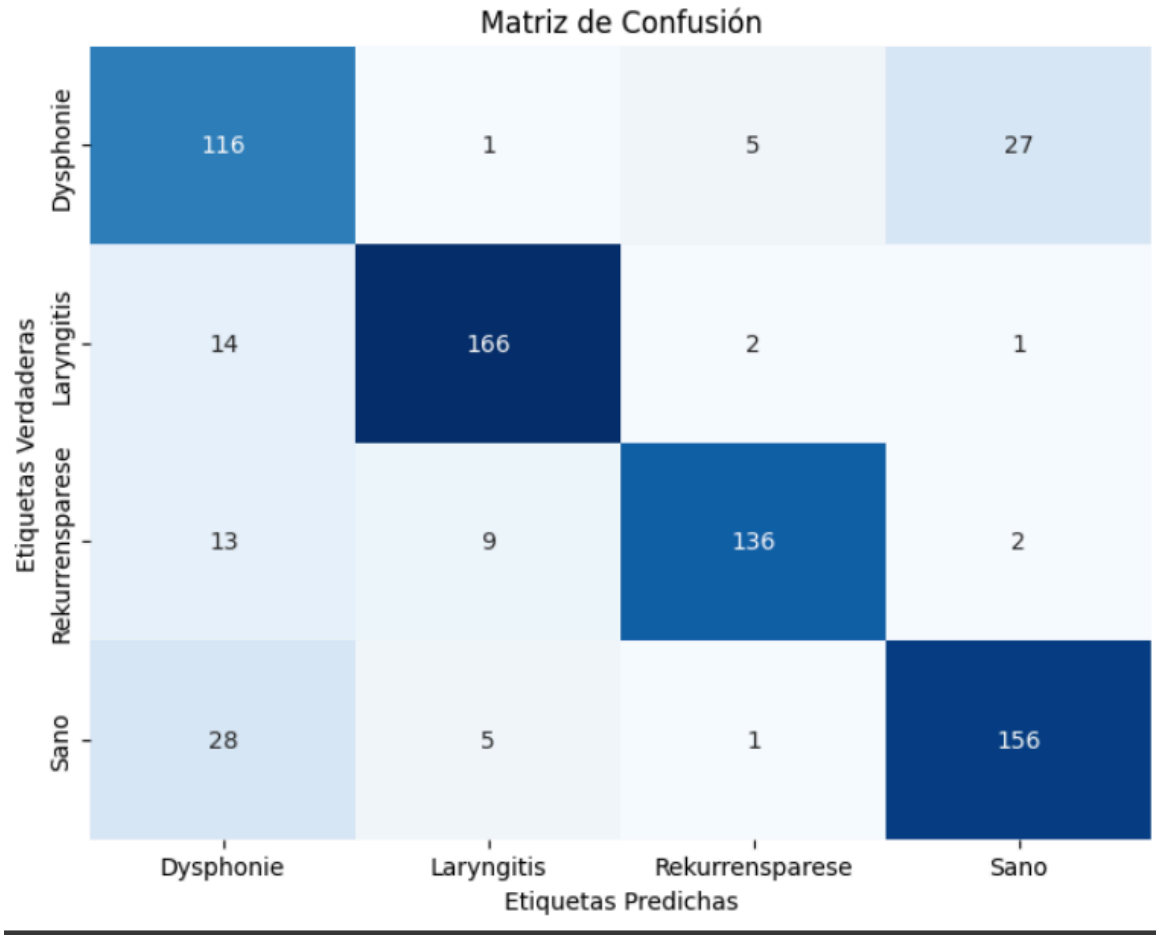
CNN Multiclase:

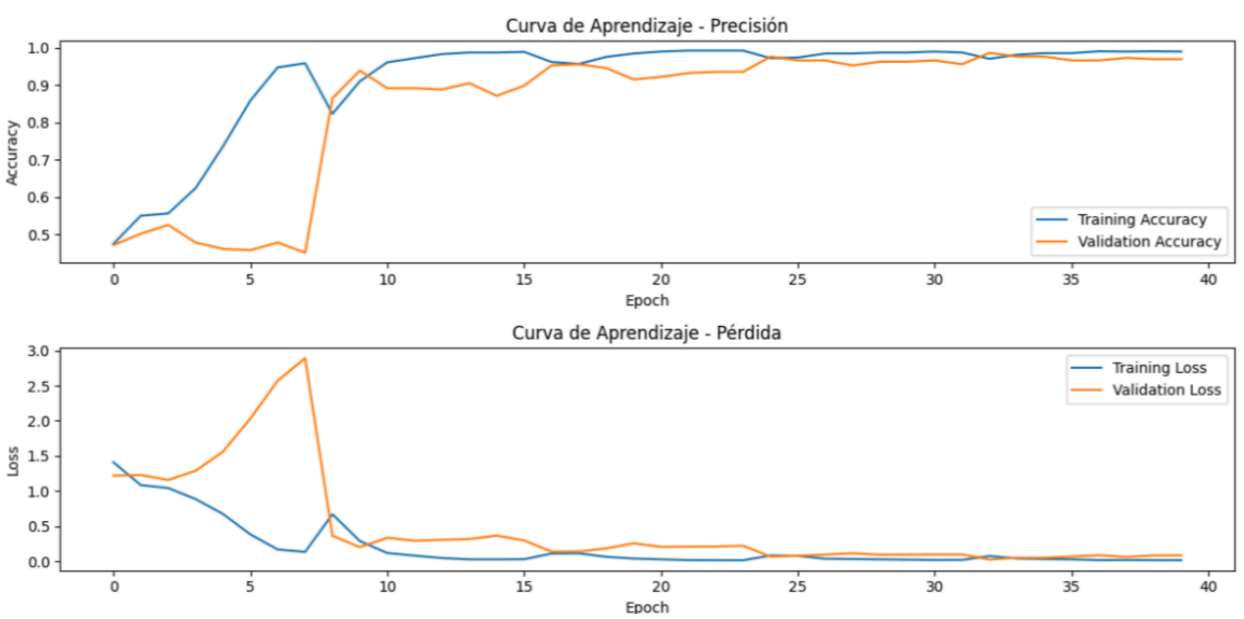
	Precision	Recall	F1-score	Support
Dysphonie	0.75	0.51	0.60	154
Laryngitis	0.95	0.97	0.96	184
Rekurrensparese	0.74	0.93	0.83	161
Sano	0.80	0.82	0.81	196
accuracy			0.82	695
macro avg	0.81	0.81	0.80	695
weighted avg	0.81	0.82	0.81	695



SincNet Multiclase:

	Precision	Recall	F1-score	Support
Dysphonie	0.68	0.82	0.63	154
Laryngitis	0.92	0.77	0.87	184
Rekurrensparese	0.94	0.88	0.86	161
Sano	0.84	0.58	0.68	196
accuracy			0.75	695
macro avg	0.80	0.76	0.76	695
weighted avg	0.81	0.75	0.76	695





Al analizar los modelos multiclase, se destaca que el rendimiento más sobresaliente lo exhibe la sincNet. Esta red, al ser una convolucional CNN con filtros específicamente diseñados para el procesamiento de audio, supera notablemente a una CNN convencional. Dada la complejidad inherente de los datos, alcanzar resultados óptimos siempre resulta desafiante. Por tanto, es recomendable considerar la implementación de ambos modelos en una aplicación eventual, comenzando con un enfoque binario y luego extendiéndose al multiclase.