



Tecnológico
de Monterrey

Proyecto Integrador (Gpo. 10)

Avance 4. Modelos Alternativos.

Equipo #6

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

A01793880
A00506509
A01793808

Las siguientes son acciones que deberás abordar en este avance:

- **Construir al menos 6 modelos diferentes (individuales, no ensambles), utilizando algoritmos variados.**

Se establecen los siguientes modelos para la evaluación de la solución de clasificación de patologías:

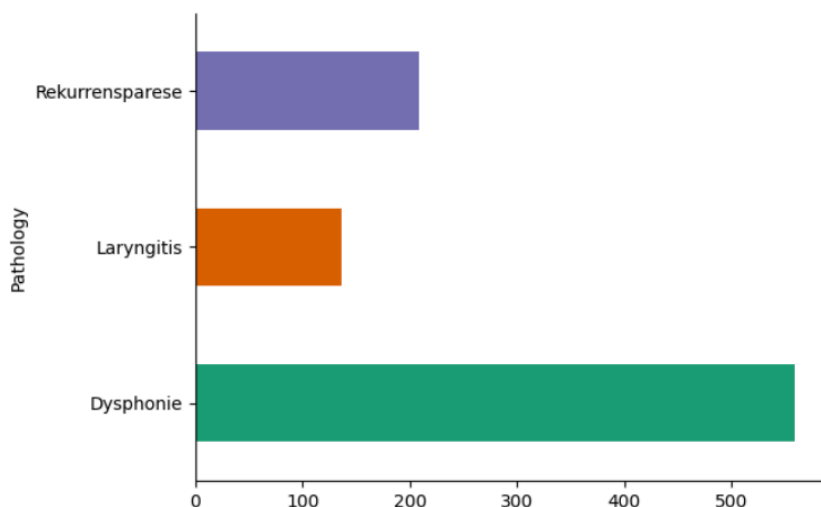
Binaria(Clasificación entre Sanos y Enfermos)

- Regresión Logística
- Red CNN
- Red Convolucional(CNN) Especifica Sinc Net
- Red Convolucional(CNN) Especifica Sinc Net + Data Augmentation - Balanceo de Clases

Multiclase - Clasificación entre cuatro clases(Dysphonie, Laryngitis, Rekurrensparese y Sano)

- Red Convolucional(CNN) Especifica Sinc Net
- Red CNN

Las patologías están distribuidas de la siguiente manera:



- **Comparar el rendimiento de los modelos obtenidos.**

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

- **Seleccionar los dos modelos que proporcionen el mejor rendimiento.**

De acuerdo con las métricas predefinidas, como la F1-score, Recall y accuracy, se seleccionaron los **modelos SincNet con aumento de datos** (por medio de white-noise) y **balanceo de clases** (sobremuestreo de la clase minoritaria) para la clasificación binaria. **Para la clasificación multiclase**, también se seleccionó el modelo SincNet, y se continuará ajustando sus **hiper parámetros** (como la tasa de aprendizaje y el número de épocas) y **seleccionando las características más relevantes** para garantizar los mejores resultados, también se aclara que para las redes entrenadas con la multiclase se utilizan la función de pérdida Sparse Categorical CrossEntropy que garantiza que el modelo pueda trabajar con las etiquetas de la clase.

- **Ajustar los dos mejores modelos.**

El proceso de ajuste de los modelos se realiza en dos etapas principales. En la primera etapa, se **aumentan los datos** (Data-Augmentation) disponibles para entrenar el modelo. Esto se puede lograr mediante técnicas utilizando el ruido blanco (white-noise) para la generación de datos sintéticos. Además, se aplica un **balanceo de clases** para evitar que la clase mayoritaria domine el proceso de aprendizaje. En la segunda etapa, se busca evitar el **sobreajuste** del modelo, que es un problema que ocurre cuando el modelo aprende demasiado de los datos de entrenamiento y no es capaz de generalizar bien a nuevos datos. Para ello, se nos hizo la recomendación de utilizar la **validación cruzada**.

- **Elegir el modelo individual final.**

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.

En resumen, SincNet ofrece una combinación única de eficiencia, interpretabilidad y capacidad para capturar características locales y globales. Estas cualidades lo hacen especialmente atractivo para el análisis de audio, particularmente en aplicaciones relacionadas con el procesamiento de señales de habla.