

#### PAPER NAME

TESISPacheco\_MT\_Final.pdf

AUTHOR Luis Pacheco

WORD COUNT

20096 Words

PAGE COUNT

100 Pages

SUBMISSION DATE

Jan 12, 2024 10:08 AM PST

CHARACTER COUNT

102011 Characters

FILE SIZE

1.5MB

REPORT DATE

Jan 12, 2024 10:10 AM PST

### • 11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Crossref database

Crossref Posted Content database

### Excluded from Similarity Report

Internet database

Publications database

• Submitted Works database







### TECNOLÓGICO NACIONAL DE MÉXICO INSTITUTO TECNOLÓGICO DE MEXICALI

### CLASIFICACIÓN DE SONIDOS RESPIRATORIOS ADVENTICIOS A TRAVÉS DEL HISTOGRAMA DE GRADIENTES ORIENTADOS

### TESIS

PARA OBTENER GRADO DE MAESTRO EN SISTEMAS COMPUTACIONALES

## PRESENTA

ING. LUIS ENRIQUE PACHECO TORRES

## DIRECTOR DE TESIS

DR. ARNOLDO RAMIREZ DIAZ (EN MEMORIA DE)

M.C. VERÓNICA QUINTERO ROSAS

## **CODIRECTOR DE TESIS**

DR. JESUS ELIAS MIRANDA VEGA

Av. Tecnológico S/N Col. Elías Calles C.P. 21376, Mexicali, B.C. Tel. 686 580 49 80 al 84 e-mail: direccion@itmexicali.edu.mx **tecnm.mx | itmexicali.edu.mx** 



Mexicali B.C. Enero de 2024





Instituto Tecnológico de Mexicali

Subdirección o Departamento que elabora el documento, Ejemplo: Departamento de Comunicación y Difusión

Mexicali, B.C., 12/Enero/2024 Memorándum DEPI-05/2024 Asunto: Autorización de impresión

#### LUIS ENRIQUE PACHECO TORRES PRESENTE

El que suscribe, Jefe de la División de Estudios de Posgrado e Investigación, comunica a usted que para la obtención del grado de Maestría en Sistemas Computacionales, se ha autorizado la reproducción de su trabajo de tesis, cuyo título es:

"CLASIFICACIÓN DE SONIDOS RESPIRATORIOS ADVENTICIOS A TRAVÉS DEL HISTOGRAMA DE GRADIENTES ORIENTADOS."

La autorización se emite con base a la revisión y dictamen emitido favorablemente y avalado con su rúbrica por los integrantes del Comité Tutorial, integrado por:

VERONICA QUINTERO ROSAS Presidente

**JESUS E. MIRANDA VEGA** 

Secretario

CAMPRILLO NAS210 Romanus

MARIO ALBERTO CAMARILLO RAMOS

Vocal

ATENTAMENTE Excelencia en Educación Tecnológica® La Tecnología para el Bien de la Humanidad

NORIO D'AMORILLO RAFINS



MARIO A. CAMARILLO RAMOS Jefe de la División de Estudios de Posgrado e Investigación

C.c.p. Departamento de Servicios Escolares C.c.p. Subdirección Académica

机除着更多不可能



Av. Tecnológico S/N Col. Elías Calles C.P. 21376, Mexicali, B.C. Tel. 686 580 49 80 al 84 e-mail: direccion@itmexicali.edu.mx tecnm.mx | itmexicali.edu.mx









### TECNOLÓGICO NACIONAL DE MÉXICO INSTITUTO TECNOLÓGICO DE MEXICALI

### CLASIFICACIÓN DE SONIDOS RESPIRATORIOS ADVENTICIOS A TRAVÉS DEL HISTOGRAMA DE GRADIENTES ORIENTADOS

### TESIS

PARA OBTENER GRADO DE MAESTRO EN SISTEMAS COMPUTACIONALES

## PRESENTA

ING. LUIS ENRIQUE PACHECO TORRES

## DIRECTOR DE TESIS

DR. ARNOLDO RAMIREZ DIAZ (EN MEMORIA DE)

M.C. VERÓNICA QUINTERO ROSAS

## **CODIRECTOR DE TESIS**

DR. JESUS ELIAS MIRANDA VEGA

Av. Tecnológico S/N Col. Elías Calles C.P. 21376, Mexicali, B.C. Tel. 686 580 49 80 al 84 e-mail: direccion@itmexicali.edu.mx **tecnm.mx | itmexicali.edu.mx** 



Mexicali B.C. Enero de 2024

## Resumen

Las enfermedades respiratorias<sup>44</sup>epresentan una de las principales causas de muerte en el mundo. En 2019, de acuerdo a informes<sup>81</sup> e la Organización Mundial de la Salud, aproximadamente 7.7 millones de personas murieron como consecuencia de estás enfermedades. Medidas como: la prevención, el diagnóstico oportuno y el tratamiento efectivo, son factores clave para reducir tanto la propagación de enfermedades como su impacto negativo en la sociedad. Entre las principales herramientas de los especialistas de la salud, para evaluar la condición del sistema respiratorio, se encuentra la auscultación. Este procedimiento le permite a los médicos escuchar los sonidos internos del cuerpo humano, con el fin de identificar diversas afecciones. Los sonidos respiratorios, en especial los sonidos respiratorios adventicios, contienen información vital acerca del estado de salud de los pulmones, vías respiratorias u otros órganos afines al sistema respiratorio. De manera que, son una pieza más para llegar al correcto diagnóstico. Sin embargo, la auscultación es un procedimiento subjetivo que depende en gran medida de la interpretación del médico. En este documento, se desarrolló un sistema computacional, utilizando el histograma de gradientes orientados a la par de algoritmos de aprendizaje de máquina, para clasificar sonidos respiratorios; en especial, sonidos respiratorios adventicios. Para ello, se empleo el marco de trabajo descrito por la base de datos ICBHI17. En donde, nuestro sistema alcanzo como resultado: 55.07 % AC, 34.37 % BA, 51.52 % AS, 75.87 % SP y 27.18 % SE; mostrando consistencia con el estado del arte.

## Abstract

<sup>62</sup> espiratory diseases represent one of the leading causes of death in the world. In 2019, 2019 Coording to World Health Organization reports, approximately 7.7 million people died as a result of these diseases. Measures such as prevention, timely diagnosis and effective treatment are key factors in reducing both the spread of diseases and their negative impact on society. Among the main tools used by health specialists evaluate the condition of the respiratory system is auscultation. This procedure allows physicians to listen<sup>51</sup>, the internal sounds of the human body in order to identify various conditions. Respiratory sounds, especially adventitious respiratory sounds, contain<sup>58</sup> tal information about the state of health of the lungs, airways or other organs related to the respiratory system. Thus, they are one more piece to reach the correct diagnosis. However, auscultations a subjective procedure that depends largely on the physician's interpretation. In this paper, a computational system was developed, using the histogram of gradient-oriented histograms coupled with machine learning algorithms, to classify respiratory sounds; in particular, adventitious respiratory sounds<sup>82</sup>or this purpose, we used the framework described by the ICBHI17 database. In which, our system achieved as a result: 55.07 % AC, 34.37 % BA, 51.52% AS, 75.87% SP and 27.18% SE; showing consistency with the state of the art.

## Dedicatoria

A mi madre, quién siempre ha estado a mi lado.

## Agradecimientos

Me gustaría agradecer a mis directores de tesis Dr. Arnoldo Díaz Ramírez y Dr. Jesús Elías Miranda Vega por su asesoramiento en este trabajo. También, a la Dr. Julia Díaz Escobar por brindarme su apoyo. Asimismo, quisiera <sup>79</sup> presar mi más profunda gratitud a mi familia, mi padre Luis por las enseñanzas que me inculco, mi madre Loreto por siempre estar a mi lado y mi hermana Gloria por siempre motivarme. Por último, pero no menos importante, agradezco al <sup>78</sup> onsejo Nacional de Humanidades, Ciencias y Tecnologías (Conahcyt) por la beca que me ha permitido realizar este trabajo.

# Índice general

1.	Intr	oducción	1
	1.1.	Pregunta de investigación	5
	1.2.	Objetivo general	5
	1.3.	Objetivos específicos	5
2.	Tral	bajos relacionados	7
3.	Ant	ecedentes teóricos	13
	3.1.	Auscultación	14
	3.2.	Sonidos respiratorios	16
		3.2.1. Sonidos respiratorios normales	17
		3.2.2. Sonidos respiratorios adventicios	19
	3.3.	Análisis computarizado de sonidos respiratorios	25
4.	$\mathbf{Sist}$	ema propuesto	27
	4.1.	Descripción de la base de datos (ICBHI17) $\ldots$	28
	4.2.	Preprocesamiento	32
	4.3.	Extracción de características	34
	4.4.	Clasificación (Algoritmos de aprendizaje de máquina)	37
	4.5.	Aumento de datos	39

5.	Resultados		
	5.1. Reducción de las representaciones TF	41	
	5.2. Combinación de métodos de preprocesamiento	43	
	5.3. Configuración del HOG	46	
	5.4. Ajuste de hiper-parámetros	51	
6.	6. Discusión		
7.	7. Conclusiones y trabajos futuros		
	Referencias	61	
A.	A. Detalles de implementación: Representaciones TF		
в.	Detalles de implementación: Configuración del HOG	73	
C.	Detalles de implementación: Espacio de parámetros	75	
D.	. Detalles de implementación: Hiper-parámetros obtenidos	77	

#### х

# Índice de figuras

4.1.	Etapas del sistema propuesto: (a) preprocesamiento, (b) extrac-	
	ción de características & (c) clasificación. $\ldots$	28
4.2.	Técnicas de rellenado: (a) Sin aplicar, (b) Zeros & (c) Réplica. $% \left( {\left( {{{\bf{x}}} \right)} \right)$ .	33
4.3.	Representaciones Tiempo-Frecuencia: (a) CWT, (b) STFT, (c)	
	MFCCs, (d) Gammatone, (e) Mel & (f) CQT	35
4.4.	Algoritmo del histograma de gradientes orientados: (a) imagen	
	original, (b) gradiente de la image (magnitud: izquierda y ángulo:	
	derecha), (c) división de la imagen en celdas, (d) histograma local,	
	(e) bloque de normalización & (f) concatenación del vector final.	36

# Índice de tablas

4.1.	Información técnica de la base de datos	29
4.2.	Información demográfica de la base de datos	31
4.3.	Distribución de los ciclos respiratorios en cada conjunto	31
5.1.	$\operatorname{Resultados}$ de los algoritmos ML en cada Representación Tiempo-	
	Frecuencia (RTF).	42
5.2.	Resultados individuales de los algoritmos ML en la combinación	
	Zeros-SF.	44
5.3.	Resultados individuales de los algoritmos ML en la combinación	
	Réplica-SF	44
5.4.	Resultados individuales de los algoritmos ML en la combinación	
	SR-Butterworth.	45
5.5.	Resultados individuales de los algoritmos ML en la combinación	
	Zeros-Butterworth	46
5.6.	Resultados individuales de los algoritmos ML en la combinación	
	Réplica-Butterworth	46
5.7.	Resultados individuales de los algoritmos ML en la serie de con-	
	figuraciones CON 1 a CON 4	48

5.8.	Resultados individuales de los algoritmos ML en la serie de con-	
	figuraciones CON 5 a CON 8	50
5.9.	Resultados individuales de los algoritmos ML en la serie de con-	
	figuraciones CON 9 a CON 12	52
5.10.	Resultados individuales de los algoritmos ML con el ajuste de	
	hiper-parámetros	53
5.11.	Resultados individuales de los algoritmos ML con el ajuste de	
	hiper-parámetros en la base de datos aumentada	53
6.1.	Comparación con el estado del arte.	58
A.1.	Detalles de implementación de la representación STFT	71
A.2.	Detalles de implementación de la representación CWT	71
A.3.	Detalles de implementación de la representación MFCCs	72
A.4.	Detalles de implementación de la representación Mel $\ .\ .\ .$ .	72
A.5.	Detalles de implementación de la representación CQT $\ . \ . \ .$ .	72
A.6.	Detalles de implementación de la representación Gammatone	72
B.1.	Detalles de implementación de cada configuración del HOG	73
C.1.	Espacio de parámetros del SVM	75
C.2.	Espacio de parámetros del KNN	75
C.3.	Espacio de parámetros del RF	76
D.1.	Hiper-parámetros obtenidos <sup>47</sup> h la base de datos sin aumento	78
D.2.	niper-parametros obtenidos en la base de datos con aumento	19

## Capítulo 1

## Introducción

Las enfermedades respiratorias<sup>44</sup>epresentan una de las principales causas de muerte en el mundo (World Health Organization, 2020). En 2019, de acuerdo a informes<sup>56</sup>e la Organización Mundial de la Salud (OMS), 3 de las 10 causas líderes de decesos en la población mundial eran enfermedades que atacaban al sistema respiratorio; ocupando el tercer puesto, las Enfermedades Pulmonares Obstructivas Crónicas (EPOCs), seguido de las infecciones respiratorias de vías bajas con el cuarto lugar y terminando con el sexto puesto, los cánceres de tráquea, pulmón y bronquios (World Health Organization, 2020). Se estima que, en ese año, murieron aproximadamente 7.7 millones de personas a causa de estás enfermedades. En otras palabras, en 2019, las enfermedades respiratorias fueron responsables de casi el 14% de las muertes en el mundo (World Health Organization, 2020). En 2021, según datos de Astrica y Geografía (INEGI), 3 de 10 principales causas de muerte en México eran enfermedades respiratorias; tomando el primer puesto, el coronavirus (CO-VID19), seguido de la neumonía e influenza con el quinto lugar y finalizando con el décimo puesto, los EPOCs (INEGI, 2022). Estás enfermedades representaron, en 2021, cerca del 28.5 % de los decesos en México; alcanzando, en conjunto, un aproximado de 320,000 muertes (INEGI, 2022). Esto sin lugar a dudas convierte a las enfermedades respiratorias en un asunto de gran relevancia tanto a nivel global como a nivel nacional.

Medidas como: la prevención, el diagnóstico oportuno y el tratamiento efectivo; son factores fundamentales para reducir tanto la propagación de enfermedades como su impacto negativo en la sociedad (Perna y Tagarelli, 2019). Entre las principales herramientas de los especialistas en la salud, para evaluar la condición del sistema respiratorio, se encuentra la auscultación (Bohadana, Izbicki, y Kraman, 2014; Sarkar, Madabhavi, Niranjan, y Dogra, 2015). Este es un procedimiento médico centrado en escuchar los sonidos internos del cuerpo humano, habitualmente a través de un instrumento como el estetoscopio acústico, con el fin de identificar diversas afecciones que pueden aparecer en distintos órganos, tales como: el corazón o los pulmones (Pramono, Bowyer, y Rodriguez-Villegas, 2017; Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013). Los sonidos generados por el sistema respiratorio, comúnmente conocidos como sonidos respiratorios, contiene información vital acerca del estado de salud de los pulmones, vías respiratorias u otros órganos afines al sistema respiratorio; en especial, los sonidos respiratorios de carácter anormal (o sonidos respiratorios adventicios) (Sarkar y cols., 2015; Pramono y cols., 2017; Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Reichert, Gass, Brandt, y Andrès, 2008). Los sonidos respiratorios adventicios son sonidos adicionales superpuestos a los sonidos respiratorios normales (Pramono y cols., 2017). Y su presencia indica algún tipo de padecimiento o enfermedad en el sistema respiratorio (Sarkar y cols., 2015). Entre los sonidos respiratorios adventicios más comunes, se encuentran: los sonidos crepitantes (o crackles, en inglés) y los sonidos sibilantes (o wheezes, en inglés). La presencia de sonidos crepitantes usualmente indica

un proceso patológico en el tejido pulmonar o en la vías respiratorias; mientras que, la presencia de sonidos sibilantes habitualmente está asociada a la obstrucción de las vías respiratorias (Sarkar y cols., 2015; Reichert y cols., 2008). Los sonidos respiratorios sirven como un apoyo para llegar al correcto diagnóstico (Sarkar y cols., 2015; Reichert y cols., 2008). No obstante, estos sonidos no son atribuibles a una única enfermedad o padecimiento; en cambio, pueden aparecer en múltiples condiciones, y en ocasiones, hasta a la vez (Sarkar y cols., 2015).

Existen otros procedimientos, a parte de la auscultación, tales como: la espirometría, los rayos X o el análisis de gases en sangre; que cuentan con un mayor grado de certeza en la evaluación del estado de salud del sistema respiratorio (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Kandaswamy, Kumar, Ramanathan, Jayaraman, y Malmurugan, 2004; Rocha y cols., 2019). Sin embargo, parte de estos métodos, como la espirometría; requieren tanto de la motivación como de la cooperación de los pacientes para obtener resultados óptimos (Hayes y Kraman, 2009). Además, suelen necesitar de un entorno clínico, como un hospital, para llevar a cabo las pruebas pertinentes; esto a causa de los equipos e instrumentos requeridos. En contraste, la auscultación resalta por su bajo costo, seguridad, fácil desempeño y su carácter no invasivo (Bohadana y cols., 2014; Sarkar y cols., 2015). Aún así, es necesario resaltar, que es un procedimiento subjetivo que recae<sup>53</sup>n gran medida en la experiencia, habilidad y capacidad auditiva del médico que lo desempeña (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Reichert y cols., 2008; Kandaswamy y cols., 2004). En otras palabras, depende de la interpretación particular que tenga el médico sobre lo que esta escuchando. Esto aumenta la probabilidad de confundir o clasificar erróneamente los distintos sonidos respiratorios, y por ende, retrasar o dificultar el diagnostico oportuno (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013). El análisis computarizado de sonidos respiratorios nace con la intención de sobrellevar estas limitaciones, al trasladar el problema a un contexto mas objetivo como lo es la computación (Sovijärvi, Vanderschoot, y Earis, 2000; Kandaswamy y cols., 2004). Para ello, se incorporaron dos componentes clave: (i) la grabación de los sonidos respiratorios, mediante dispositivos electrónicos como micrófonos o estetoscopios digitales; y (ii) la implementación de sistemas computacionales para el análisis de las señales de audio (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Gurung, Scrafford, Tielsch, Levine, y Checkley, 2011; Marques, Oliveira, y Jácome, 2014). Esto derivo en un foco de atención por parte de la comunidad científica, en el análisis de sonidos respiratorios, que culmino en múltiples propuestas relacionadas al preprocesamiento, <sup>16</sup>xtracción de características y clasificación de las señales de audio (Pramono y cols., 2017; Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Gurung y cols., 2011; J. E. Earis y Cheetham, 2000). Y dio pie a diversas aplicaciones, siendo una de las mas importantes: la clasificación automatizada de los sonidos respiratorios (Pramono y cols., 2017).

En este documento, se propone el desarrollo de un sistema computacional enfocado en la clasificación de sonidos respiratorios; en específico, sonidos respiratorios de carácter anormal, tales como: los sonidos crepitantes y los sonidos sibilantes. Con el fin, de apoyar en el diagnostico oportuno de enfermedades respiratorias. Para ello, se plantea el uso del descriptor de características: Histograma de Gradientes Orientados (HOG, por sus siglas en ingles) (Dalal y Triggs, 2005), originalmente concebido dentro de la visión computacional, como nuestro extractor de características. Puesto que, hasta donde conocemos, este descriptor no ha sido empleado dentro del análisis de sonidos respiratorios. Además, ha obtenido resultados favorables en la clasificación de otro tipo de señales de audio (escenas acústicas) (Rakotomamonjy y Gasso, 2014; Bisot, Essid, y Richard, 2015; Abidin, Togneri, y Sohel, 2018); consiguiendo, en casos particulares, competir con el estado del arte de ciertas bases de datos (Rakotomamonjy y Gasso, 2014). Asimismo, se propone utilizar algoritmos de aprendizaje de maquina tradicionales, tales como. Adaptina de Soporte Vectorial (SVM, por sus siglas en ingles), K-Vecinos mas Cercanos (KNN, por sus siglas en ingles) y Bosques Aleatorios (RF, por sus siglas en ingles); para clasificar los sonidos respiratorios. Finalizando, con una comparación entre los resultados obtenidos y los trabajos realizados en la base de datos ICBHI17 (Rocha y cols., 2019).

#### 1.1. Pregunta de investigación

En función al marco de trabajo provisto por la base de datos ICBHI17. ¿Qué tan efectivo es el histograma de gradientes orientados, como extractor de características, en la clasificación de sonidos respiratorios?

#### 1.2. Objetivo general

Desarrollar un sistema computacional, utilizando el histograma de gradientes orientados como extractor de características a la par de algoritmos de aprendizaje de máquina tradicionales, capaz de clasificar sonidos respiratorios; en específico, sonidos respiratorios de carácter anormal como: los sonidos crepitantes y los sonidos sibilantes; con el fin, de apoyar en el diagnóstico oportuno de enfermedades respiratorias.

#### 1.3. Objetivos específicos

- Examinar la literatura referente al análisis computarizado de sonidos respiratorios
- Elegir una base de datos relacionada al tema

- Elegir técnicas para el preprocesamiento de las señales de audio
- Determinar los métodos de análisis tiempo-frecuencia para convertir a imagen las señales de audio
- Seleccionar los parámetros más adecuados para HOG
- Definir los algoritmos de aprendizaje de máquina tradicionales
- Evaluar el desempeño del sistema propuesto

## Capítulo 2

## Trabajos relacionados

En los últimos 50 años, el análisis computarizado de sonidos respiratorios ha dado grandes avances en la reducción de las limitaciones de la auscultación tradicional, sin perder las cualidades que presenta este procedimiento, tales como: fácil desempeño, menor costo y carácter no invasivo; en comparación a otros procedimientos como: los rayos X o la tomografía computarizada (Pramono y cols., 2017; Sovijärvi v cols., 2000; Palaniappan, Sundaraj, Ahamed, Arjunan, v Sundaraj, 2013). Esto a atraído la atención de la comunidad científica al análisis de sonidos respiratorios, lo cual ha desembocado en múltiples propuestas relacionadas al preprocesamiento, extracción de características y clasificación de los sonidos respiratorios (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Gurung y cols., 2011; J. E. Earis y Cheetham, 2000). Particularmente, en el ámbito de la extracción de características, se han implementado una amplia gama de técnicas o métodos, que parten desde características habituales de las señales de audio, en los dominios: tiempo, frecuencia o tiempo-frecuencia; hasta la aplicación de aprendizaje de máquina profundo para la ejecución de está tarea (Pramono y cols., 2017; Palaniappan, Sundaraj, y Ahamed, 2013; Kim y cols., 2022). En este apartado, nos enfocaremos en los trabajos asociados a la base de datos ICBHI17 (Rocha y cols., 2019); en específico, aquellos que sigan las normas impuestas en la misma. Chambres y cols. (Chambres, Hanna, y Desainte-Catherine, 2018) proponen utilizar un conjunto de características, en las cuáles se incluyen: características con información espectral (p. ej. bandas de energía), características de ritmo, características tonales, características SFX, valores estadísticos de las características mencionadas (desviación, varianza, promedio, etc.) y Mel Frequency Cepstral Coefficients (MFCCs); aunque termina utilizando únicamente las características con información espectral y los MFCCs con un Arbol de Decisión (DT, por sus siglas en inglés). Asimismo, Jakovljević v Lončar-Turukalo (Jakovljević v Lončar-Turukalo, 2018) implementan los MFCCs a la par de su primera derivada junto con un Modelo Oculto de Márkov (HMM<sup>38</sup>) or sus siglas en inglés) combinado con un Modelo Gaussiano Mixto<sup>38</sup>MM, por sus siglas en inglés). Serbes y cols. (Serbes, Ulukaya, y Kahya, 2018) plantean descomponer las señales de audio con una transformada Wavelet y aplicar a cada parte de la señal descompuesta un Short Time Fourier Transform (STFT) así como también obtener características espectrales y valores estadísticos de los coeficientes de la transformada Wavelet; y clasificar los sonidos respiratorios mediante un SVM. De igual manera, Ntalampiras y Potamitis (Ntalampiras y Potamitis, 2019) dividen las señales de audio con una transformada Wavelet y obtienen el Área Envolvente de Auto-correlación (AEA) de las distintas partes para formar un vector de características final; este vector es utilizado con un Gráfico Acíclico Dirigido (DAG, por sus siglas en inglés) en combinación con un HMM. Pham y cols. (Pham, Phan, Palaniappan, Mertins, y McLoughlin, 2021) proponen el uso de una Red Neuronal Convolucional (CNN, por sus siglas en inglés) que incorpora la técnica Mixture of Experts (MoE) en distintas representaciones TF: espectrograma Log-Mel, MFCCs, Gammatone y Constant-Q Transform (CQT). Al igual que Ngo y cols. (Ngo y cols., 2021) que utilizan una CNN a la par de un Autoencoder en un espectrograma Gammatone. Petmezas y cols. (Petmezas y cols., 2022) proponen combinar una CNN con una Red Neuronal de Memoria a Corto Plazo Largo (LSTM, por sus siglas en inglés) y aplicar la función de perdida Focal Loss; sobre una representación STFT. Por su parte, Chang y cols. (Chang, Ren, Nguyen, Nejdl, y Schuller, 2022) hacen uso de una CNN así como un modelo ResNet (Residual Network) e incluye en las estructuras una capa Attention y un bloque Dilation; entrenando los modelos sobre un espectrograma Mel. Moummad y Farrugia (Moummad y Farrugia, 2023) emplean una CNN a la par de la técnica de aprendizaje constrativo en un espectrograma Mel. Mientras que, Neto y cols. (Neto, Arrais, Vinuto, y Lucena, 2022) proponen el uso de un transformador de visión convolucional en las representaciones: espectrograma Mel, MFCCs v CQT. Ma v cols. (Ma v cols., 2019) implementan un modelo ResNet bilineal, tomando en una entrada la representación STFT y en la otra una matriz generada por un análisis Wavelet. Asimismo, Yang y cols. (Yang, Liu, Song, Parada-Cabaleiro, y Schuller, 2020) utilizan un modelo ResNet de 18 capas (ResNet-18) e incorpora a la estructura diferentes bloques de tipo Attention; entrenando el modelo en una representación STFT. Ma y cols. (Ma, Xu, y Li, 2020) proponen un modelo ResNet añadiendo un bloque Non-Local, en una representación STFT. Asimismo, Gairola y cols. (Gairola, Tom, Kwatra, y Jain, 2021) hace uso de un modelo ResNet de 34 capas (ResNet-34) en un espectrograma Mel. Nguyen y Pernkopf (Nguyen y Pernkopf, 2022) utilizan un modelo ResNet de 50 capas (ResNet-50) así como un modelo ResNet de 101 capas (ResNet-101) en un espectrograma Log-Mel. Por su parte, Tong y cols. (Tong, Liu, Xie, Hong, y Li, 2022) implementan un modelo ResNet incorporando a la estructura un bloque Squeeze-Excitation en un espectrograma Mel. Pham y cols. (Pham, Ngo, Hoang, Schindler, y McLoughlin, 2022) realizan una comparación de distintos modelos: ResNet, VGG, MobileNet, Inception, Xception y DenseNet; en un espectrograma Wavelet y un espectro grama Log-Mel. Asimismo, Pham y cols. (Pham, Phan, Schindler, y cols., 2021) emplean de base un modelo Inception en un espectrograma Wavelet. Mientras que, Vinh y cols. (Vinh, Chau, Nam, y Long, 2022) proponen un modelo VGG de 12 capas (VGG-12) incorporando un modulo de atención convolucional en una representación Gammatone. Como se puede observar, hasta los limites de nuestra revisión, el descriptor de características HOG no fue explorado dentro de la clasificación de sonidos respiratorios. No obstante, existen otros trabajos asociados a la clasificación de escenas acústicas, que muestran resultados favorables por parte del HOG en la clasificación de señales de audio. Rakotomamonjy y Gasso (Rakotomamonjy y Gasso, 2014) proponen el uso del histograma de gradientes orientados en representaciones TF, con el fin de demostrar que este extractor de características provee información relacionada a las representaciones, como la dirección de la potencia espectral de la variación, que otros métodos (p. ej. MFCCs) no poseen. Para ello, implementa el HOG en cuatro bases de datos distintas con una representación CQT a la par de una SVM; en adición, agregan una etapa extra al algoritmo original, que incluye un proceso de pooling para reducir la dimensionalidad del vector de características. Obteniendo con esto, resultados comparables con el estado del arte de las distintas bases de datos. Por su parte, Bisot y cols. (Bisot y cols., 2015) emplean el HOG en conjunto con la característica SPD (Subband Power Distribution), en una representación TF de la percepción del volumen así como también un CQT, para entrenar una SVM; alcanzando en solitario, una exactitud 92 % y, en conjunto con SPD, una exactitud 93.4%. Finalmente, Abidin y Sohel (Abidin y cols., 2018) aplican el HOG en una representación VQT (Variable-Q Transform) con una SVM; obteniendo, una exactitud de 81%. Aunque cabe recalcar, que en este ultimo no

emplean el HOG como el tema central de sus experimentos.

## Capítulo 3

## Antecedentes teóricos

El sistema respiratorio es un grupo de órganos, que actúan de manera coordinada, para llevar a cabo el proceso de la respiración (Ionescu, 2013). Este mecanismo se encarga de intercambiar el dióxido de carbono  $(CO_2)$ , generado como desperdicio por las células, por el oxigeno  $(O_2)$  que se encuentra en el ambiente; a través de dos etapas: (i) la inspiración, al inhalar el O2 a los pulmones y (ii) espiración, al exhalar el  $CO_2$  al ambiente (Ionescu, 2013; AL-Khalidi, Saatchi, Burke, Elphick, y Tan, 2011). La secuencia de estas dos fases (inspiración/espiración) es conocida como ciclo respiratorio. En la inspiración, el oxigeno ingresa al cuerpo por la nariz o la boca, pasa por la garganta (laringe y traquea) hasta llegar a dos grandes conductos (bronquios) conectados a los pulmones; los cuáles, ramifican en pequeños caminos, denominados: bronquiolos, que desembocan en diminutos sacos de aire, conocidos como: alvéolos; donde, el  $O_2$  se difunde en los capilares pulmonares a cambio de dióxido de carbono; produciendo aire con  $CO_2$ , que se expulsa, en la espiración, recorriendo el camino inverso hasta salir por la boca o la nariz (Ionescu, 2013; Robinson y Furlow, 2007; Tu, Inthavong, y Ahmadi, 2013). El sistema respiratorio se puede dividir en regiones de acuerdo a su función o anatomía (Tu y cols., 2013). Anatómicamente, existen dos grandes agrupaciones: (i) las vías respiratorios superiores, formadas por órganos que se encuentran fuera del tórax: nariz, faringe y laringe; y (ii) las vías respiratorias inferiores, constituidas por órganos que se localizan dentro del tórax: traquea, pulmones, bronquios, bronquiolos y alvéolos (Ionescu, 2013; Patwa y Shah, 2015; Tu y cols., 2013). Mientras que, funcionalmente, existen dos zonas: (i) conductiva (nariz a bronquiolos), con órganos que forman el camino para llevar el oxigeno a los pulmones; y (ii) respiratoria (alvéolos), en donde se lleva a cabo el intercambio de gases (Tu y cols., 2013; Patwa y Shah, 2015). El sistema respiratorio cuenta, también, con otros propósitos, tales como: filtrar, humidificar y calentar el aire inhalado (nariz); producir sonido (laringe) o regular el pH del cuerpo (pulmones) (Tu y cols., 2013; Strohl, Butler, y Malhotra, 2012).

#### 3.1. Auscultación

La auscultación es un procedimiento médico centrado en escuchar los sonidos internos del cuerpo humano, usualmente mediante un instrumento como el estetoscopio acústico, con el fin de identificar distintas afecciones que se pueden presentar en órganos como: el corazón o los pulmones (Pramono y cols., 2017; Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013). Documentos tan antiguos como el Papiro Ebers (1500 A.C) o las Vedas Hindúes (1500 A.C – 1200 A.C) hacen referencia ha esté procedimiento (Bishop, 1980; Markel, 2006; Andrès, Gass, Charloux, Brandt, y Hentzler, 2018). Aunque, es Hipócrates (460 A.C – 360 A.C) quién consolida el concepto, al posar su oído en el pecho de los pacientes para escuchar los sonidos respiratorios; acuñando el término auscultación inmediata (Sarkar y cols., 2015; Markel, 2006). Sin embargo, no es hasta la invención del estetoscopio acústico, por parte de René Théophile Hyacinthe
Laennec en 1816, que este procedimiento se populariza alrededor del mundo; gracias a la sustitución del contacto directo por un instrumento (Sarkar y cols., 2015; Andrès y cols., 2018). Volviendo a la auscultación en un procedimiento más cómodo para los pacientes, particularmente para aquellos en el que la auscultación inmediata podría considerarse inadecuada o embarazosa (Sarkar y cols., 2015; Bishop, 1980; Weinberg, 1993). En los años venideros, Laennec publicaría una de las obras clásicas dentro de la literatura médica *De l' Auscultation Médiate, etc'*, donde plasmaría sus hallazgos con el estetoscopio; describiendo una gran variedad de enfermedades del pecho y, en menor medida, padecimientos del corazón; a la par de los sonidos asociados a cada una de ellas (Bishop, 1980).

En la actualidad, a más de 200 años de su invención, el estetoscopio se ha convertido en un símbolo de la profesión médica así como la auscultación se ha consagrado como parte de los conocimientos básicos del especialista en la salud (Markel, 2006; Jr, 2008). No obstante, con el avance de la tecnología, el valor clínico de este procedimiento se ha puesto en tela de juicio; puesto que, herramientas como: los rayos X o la tomografía computarizada, ofrecen una mayor certeza en la detección de padecimientos (p. ej. enfermedades respiratorias) (Jr, 2008). Este cuestionamiento se avala de la subjetividad del procedimiento, que recaen<sup>53</sup>h gran medida en la experiencia, habilidad y capacidad auditiva del médico que desempeña el proceso (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Reichert v cols., 2008; Kandaswamy v cols., 2004). En otras palabras, depende de cómo el médico interpreta lo escuchado; lo cuál, aumenta la probabilidad de confundir o clasificar erróneamente los sonidos respiratorios, y por ende, retrasar o dificultar el diagnóstico (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013). Adicionalmente, el estetoscopio acústico plantea otros problemas como: la susceptibilidad al ruido, los efectos de resonancia tubular o la atenuación de los sonidos en frecuencias altas; inherentes al diseño (Sovijärvi y cols., 2000; Gurung y cols., 2011). A pesar de lo dicho, la auscultación no da indicios de desaparecer, puesto que ofrece ciertas cualidades como: un bajo costo, fácil desempeño, carácter no invasivo, seguridad y poca necesidad de cooperación; que los procedimientos mencionados no poseen (Bohadana y cols., 2014; Sarkar y cols., 2015). Por ello, al contrario de lo sugerido, el siguiente paso para la auscultación es adaptarse a los tiempo, con la inclusión de tecnología al procedimiento (Jr, 2008; Kim y cols., 2022).

### 3.2. Sonidos respiratorios

Los sonidos respiratorios son sonidos producidos en el sistema respiratorio, al momento de respirar (Sarkar y cols., 2015; Sovijarvi y cols., 2000). Estos sonidos se caracterizan de tres atributos: frecuencia, amplitud y calidad o timbre; que nos permiten discernir entre dos sonidos similares (Sarkar y cols., 2015). La frecuencia, medida en hertz (Hz), es el número de ondas de sonido por segundo; a la par, el tono (o pitch, en inglés) es la interpretación subjetiva de la frecuencia. Por su parte, la amplitud define la distancia entre la posición media de la onda y su pico, relacionándose con la energía que contiene; asimismo, la intensidad sonora (o loudness, en inglés) es la percepción subjetiva de la amplitud, y se mide en decibelios (dB). Mientras tanto, la calidad o timbre es la propiedad del sonido que nos permite distinguir entre dos sonidos con tono e intensidad sonora idénticos (Sarkar y cols., 2015). Los sonidos respiratorios contienen información relevante acerca del estado de salud de las vías respiratorias, pulmones u otros órganos afines al sistema respiratorio; particularmente, los sonidos de carácter anormal (o sonidos respiratorios adventicios) (Sarkar y cols., 2015; Pramono y cols., 2017; Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Reichert y cols., 2008). En este apartado, se discutirán los distintos sonidos respiratorios.

### 3.2.1. Sonidos respiratorios normales

Los sonidos respiratorios normales son sonidos generados por un sistema respiratorio sano, con el paso de una corriente de aire a través del árbol traqueobronquial (Sarkar y cols., 2015; Pasterkamp, Kraman, y Wodicka, 1997). Estos sonidos son producidos por corrientes turbulentas o corrientes de vórtice (Sarkar y cols., 2015; Pasterkamp y cols., 1997; J. Earis, 1992). Las corrientes de aire turbulento generan sonido conforme las moléculas de aire colisionan unas con otras, y con las paredes de las vías respiratorias, como consecuencia de su naturaleza desordenada; mientras que, las corrientes de vórtice se forman cuando el flujo de aire que sale de un orificio circular se dirige a un canal más ancho (Sarkar y cols., 2015). Otras corrientes de aire, como las corrientes laminares, son incapaces de crear sonido, ya que transitan en paralelo a las vías respiratorias (Sarkar y cols., 2015; Forgacs, 1978). Los sonidos respiratorios normales se pueden clasificar como: vesicular (pulmonar), traqueal o bronquial; de acuerdo a lugar de adquisición (Sarkar y cols., 2015; Bohadana y cols., 2014).

### Sonidos vesiculares normales

Los sonidos vesiculares (pulmonares) normales son sonidos suaves, susurrantes, no musicales, de tonalidad baja, escuchados sobre la superficie del pecho; que se observan solo en la inspiración y en la espiración temprana; esto último, por la naturaleza pasiva de la etapa que genera menos corrientes de aire turbulentas (Bohadana y cols., 2014; Sarkar y cols., 2015). Esté tipo de sonidos respiratorios normales se caracterizan por contar con<sup>16</sup>n rango de frecuencias de 100 Hz a 1000 Hz, con una perdida de energía entre 100 Hz y 200 Hz; a consecuencia del efecto, tipo filtro pasa-baja, creado por las paredes torácicas (Bohadana y cols., 2014; Pramono y cols., 2017). Los sonidos vesiculares normales se originan en diferentes puntos, en función a la etapa del ciclo respiratorio; durante la inspiración estos sonidos provienen de las vías respiratorias lobulares y segmentarias, mientras que, en la espiración proceden de fuentes más centrales (Bohadana y cols., 2014; Sarkar y cols., 2015).

#### Sonidos traqueales normales

Los sonidos traqueales normales son sonidos ásperos, fuertes, de tonalidad alta, escuchados en la escotadura supraesternal o en el lateral del cuello; que se presentan en ambas fases del ciclo respiratorio, con duraciones similares entre sí (Bohadana y cols., 2014; Sarkar y cols., 2015). Estos sonidos respiratorios normales se caracterizan por tener un amplio rango de frecuencias, que varían de 100 Hz a 5000 Hz; con una perdida de energía en los 800 Hz y poca energía después de los 1500 Hz (Bohadana y cols., 2014). Los sonidos traqueales normales son generados por el paso de una corriente de aire turbulenta en la faringe, glotis y región subglótica (Bohadana y cols., 2014). La auscultación de la traquea no es un procedimiento común, sin embargo, estos sonidos poseen información valiosa puesto que reciben los sonidos de los pulmones sin el efecto del filtro (Sarkar y cols., 2015; Pramono y cols., 2017).

### Sonidos bronquiales normales

Los sonidos bronquiales normales son sonidos huecos, fuertes, de alta tonalidad, similares a los sonidos traqueales, escuchados sobre las vías respiratorias largas del pecho, en específico, en el segundo y tercer espacio intercostal; que se observan en ambas etapas del ciclo respiratorio, donde la espiración es más intensa así como prolongada, en el aspecto audible, en comparación con la inspiración (Bohadana y cols., 2014; Sarkar y cols., 2015; Pramono y cols., 2017). Estos sonidos respiratorios normales se caracterizan por contener más energía en bandas de frecuencias altas, que los sonidos vesiculares normales (Pramono y cols., 2017). Los sonidos bronquiales normales son un sonido intermedio entre los sonidos traqueales normales y los sonidos vesiculares normales (Bohadana y cols., 2014).

### 3.2.2. Sonidos respiratorios adventicios

Los sonidos respiratorios adventicios son sonidos superpuesto a los sonidos respiratorios normales (Pramono y cols., 2017). Y su presencia, usualmente, indica la existencia o desarrollo de algún padecimiento en el sistema respiratorio (Sarkar y cols., 2015). Esta clase de sonidos se pueden clasificar en: continuos (sibilancia, roncus & estridor) o discontinuos (crepitante, de grazna & roce pleural); de acuerdo a la duración de los sonidos (Sarkar y cols., 2015; Pramono y cols., 2017). Dentro de cada categoría, la sibilancia así como el crepitante son los sonidos respiratorios adventicios más comunes.

#### Sonidos sibilantes (wheezes)

Los sonidos sibilantes son sonidos respiratorios adventicios continuos, de carácter musical, que se presentan en una o ambas etapas del ciclo respiratorio, ya sea en la inspiración o en la espiración (Bohadana y cols., 2014; Forgacs, 1978). Estos sonidos continuos se caracterizan por durar típicamente más de 100 ms así como por contar con<sup>16</sup>n rango de frecuencias de 100 a 1000 Hz; a la par de armónicos que ocasionalmente superan está marca (Bohadana y cols., 2014; Sarkar y cols., 2015). El carácter musical de está clase de sonidos, es determinado por su espectro; almacenando casi toda su energía en las frecuencias armónicas relacionadas; mientras que, el tono de la nota es definido por la frecuencia fundamental (Forgacs, 1978). Los armónicos son frecuencias superiores, formadas por múltiplos completos, de la frecuencia fundamental (la frecuencia

menor); dígase 100 Hz o 150 Hz de la frecuencia base: 50 Hz (Sarkar y cols., 2015). Los sonidos sibilantes se pueden clasificar en dos categorías: monofónicos o polifónicos; de acuerdo a las notas musicales percibidas (Forgacs, 1978). La primera categoría consiste de una única nota que comienza y termina en momentos diferentes (Sarkar y cols., 2015). La presencia de los sonidos monofónicos se asocia con enfermedades como: asma, EPOCs, obstrucción bronquial por tumor, broncostenosis por inflamación, acumulación de moco, un cuerpo extraño en las vías respiratorias y, en algunos casos, con enfermedades vinculadas a la fibrosis intersticial difusa; donde se pueden observar en conjunto con los sonidos crepitantes (Sarkar y cols., 2015; Forgacs, 1978). Mientras tanto, los sonidos polifónicos están formados por múltiples notas musicales que comienzan y terminan al mismo tiempo; de manera similar a un acorde disonante (Sarkar y cols., 2015: Forgacs, 1978). La presencia de está clase de sonidos también se relaciona con la obstrucción de las vías respiratorias; sin embargo, su valor clínico se encuentra en un punto ambiguo, ya que los sonidos polifónicos pueden producirse, a voluntad, por pacientes sanos con una espiración forzada (Sarkar y cols., 2015; Forgacs, 1978).

Múltiples teorías se han presentado para esclarecer cómo se producen los sonidos sibilantes. En un inicio, se creía que estos sonidos eran el resultado de un mecanismo similar al de un tubo de órgano; donde el gas contenido en el interior de un bronquio estrechado generaba una nota musical al comenzar a oscilar (Forgacs, 1978). No obstante, las aseveraciones de está explicación, como: que el tono de la nota dependía de la longitud de las vías respiratorias contiguas o que una simple reducción en el calibre del tracto respiratorio bastaba para generar los sonidos musicales; se contraponía a distintas observaciones, como: que ciertos sonidos sibilantes de baja frecuencia requerían de una longitud de 4 a 8 pies, mientras que, la vía axial más larga del árbol bronquial medía menos de un pie de longitud; que ningún sonido musical podía producirse sin el contacto de las paredes opuestas del bronquio comprimido; o que el tono de estos sonidos continuos no cambiaba, en contraste con los instrumentos de viento, al sustituir el oxigeno por otro gas (helio) (Forgacs, 1978; Meslier, Charbonneau, y Racineux, 1995). Más adelante, Paul Forgacs (Forgacs, 1978) propondría que los sonidos sibilantes eran creados por un bronquio estrechado hasta el punto de cierre, cuyas paredes estarían en constante cambio entre los estados: cerrado o apenas abierto; donde el tono de la nota estaría en función a la masa y las propiedades elásticas de las paredes (Sarkar y cols., 2015; Meslier y cols., 1995). Aunque, serían Grotberg y Gavriely (Grotberg y Gavriely, 1989), quienes establecerían que los sonidos sibilantes iban acompañados de una corriente de aire restringida, pero que la limitación del flujo no iba necesariamente acompañada de estos sonidos: planteando que su producción se debía a la oscilación conjunta entre el fluido y las paredes de las vías respiratorias, que daba comienzo cuando la velocidad del aire alcanzaba un valor crítico (velocidad de aleteo); en donde el tono de la nota era definido por el grosor de las paredes así como por su rigidez a la flexión y la tensión longitudinal (Sarkar y cols., 2015; Meslier y cols., 1995).

### Sonidos estridores (stridors)

Los sonidos estridores (stridors, en inglés) son sonidos respiratorios adventicios continuos, agudos fuertes, de carácter musical, parecidos a los sonidos sibilantes; que se observan, frecuentemente, durante la inspiración y, en menor medida, en la espiración, en cuyo caso son bifásicos (Bohadana y cols., 2014; Sovijärvi y cols., 2000; Pramono y cols., 2017). Estos sonidos continuos se caracterizan por una duración mayor a 250 ms en conjunto con una frecuencia fundamental de 500 Hz; además, de múltiples armónicos (Bohadana y cols., 2014; Pramono y cols., 2017). Los sonidos estridores son causados por una corriente de aire turbulenta que pasa por una sección estrecha de las vías respiratorias superiores, por ello, son más evidentes sobre la traquea (Bohadana y cols., 2014; Pramono y cols., 2017). Estos sonidos adventicios están relacionados a padecimientos como: epiglotitis aguda, anafilaxia, disfunción de las cuerdas vocales, inhalación de cuerpos extraños, tumores laríngeos, tiroiditis, edema de las vías respiratorias tras la extracción de un dispositivo y carcinoma traqueal (Bohadana y cols., 2014).

### Sonidos de grazna (squawks)

Los sonidos de grazna (squawks, en inglés), también conocidos como sibilantes cortos, son sonidos respiratorios adventicios continuos, de carácter tanto musical como no musical; que se presentan durante la inspiración tardía, habitualmente, precedidos por sonidos crepitantes (Bohadana y cols., 2014; Sarkar y cols., 2015). Estos sonidos adventicios se caracterizan por tener una duración menor a 200 ms a la par de una frecuencia fundamental que varía entre 200 Hz a 300 Hz (Bohadana y cols., 2014). Se teoriza que los sonidos de grazna son producidos por la oscilación de las vías respiratorias periféricas, en zonas pulmonares desinfladas (Bohadana y cols., 2014; Sarkar y cols., 2015). Estos sonidos continuos están asociados a enfermedades como: neumonitis por hipersensibilidad, neumonía y bronquiectasias (Bohadana y cols., 2014).

### Sonidos roncosos (rhonchus)

Los sonidos roncosos (rhonchus, en inglés) son sonidos sibilantes de baja tonalidad similares a los ronquidos; que se presentan tanto en la inspiración como en la espiración, o en ambas etapas del ciclo respiratorio (Bohadana y cols., 2014). Estos sonidos adventicios se caracterizan por una duración mínima de 80 ms a la par de una frecuencia máxima de 200 Hz (Bohadana y cols., 2014; Pramono y cols., 2017). Los sonidos roncosos son causados por el espesamiento de la mucosidad en las vías respiratorias (Pramono y cols., 2017). Estos sonidos continuos están asociados a enfermedades como: EPOCs y bronquitis (Bohadana y cols., 2014).

#### Sonidos crepitantes (crackles)

Los sonidos crepitantes (o crackles) son sonidos respiratorios adventicios discontinuos, de carácter explosivo, que se observan, en gran medida, durante la inspiración y, ocasionalmente, en la espiración (Bohadana y cols., 2014; Sarkar y cols., 2015). Estos sonidos discontinuos se caracterizan por una duración menor a 20 ms a la par de un amplio espectro de frecuencias, que oscilan entre 100 Hz a 2000 Hz (J. Earis, 1992). Los crackles se dividen en dos grupos: crepitantes finos (o fine crackles, en inglés) y crepitantes gruesos (o coarse crackles, en inglés); según la duración y frecuencia del sonido (Piirila y Sovijarvi, 1995). El primer grupo cuenta con una duración cercana a los 5 ms junto con una frecuencia aproximada de 650 Hz (Bohadana y cols., 2014; Pramono y cols., 2017). Los crepitantes finos tienden a presentarse en la inspiración media o tardía, no son audibles en la boca e inician, habitualmente, en la parte basal de los pulmones; asimismo, no se ven afectados por la tos, solo la gravedad puede modificar su estado; desapareciendo con los cambios de posición del cuerpo (Bohadana y cols., 2014; Sarkar y cols., 2015; Pasterkamp y cols., 1997). Este grupo de sonidos adventicios se relaciona con enfermedades como: fibrosis pulmonar idopática, asbestosis, neumonitis intersticial inespecífica y la fibrosis intersticial asociada a trastornos del tejido conjuntivo; asemejándose al sonido que se crea al separar lentamente un velcro (Bohadana y cols., 2014; Sarkar y cols., 2015; Pramono y cols., 2017). Mientras tanto, los crepitantes gruesos poseen una duración aproximada de 15 ms a la par de una frecuencia cercana a los 350 Hz (Bohadana y cols., 2014; Sarkar y cols., 2015; Pramono y cols., 2017). Este grupo de sonidos discontinuos pueden aparecer en la etapa temprana de la inspiración, o a lo largo de la espiración; no son procedentes de ninguna área particular de los pulmones, suelen transmitirse a la boca y, en contraste con el otro grupo, se ven alterados, con la posibilidad de desaparecer, por la tos (Bohadana y cols., 2014; Sarkar y cols., 2015). Los crepitantes gruesos se asocian con enfermedades como: EPOCs, bronquiectasias, asma, neumonía e insuficiencia cardíaca congestiva; descrito como un "pop" (Bohadana y cols., 2014; Sarkar y cols., 2015; Pramono y cols., 2017).

A lo largo de los años, numerosas propuestas han surgido para explicar qué mecanismo crea los sonidos crepitantes. Inicialmente, se tenía la noción de que el paso del aire a través de secreciones acumuladas en las vías respiratorias generaba estos sonidos (Sarkar y cols., 2015; J. Earis, 1992; Forgacs, 1978; Piirila y Sovijarvi, 1995). Sin embargo, aunque en ciertos casos es verdad, la persistencia de este sonido discontinuo después de toser así como su repetida aparición en la inspiración, contradice está idea como una solución general (Sarkar y cols., 2015; J. Earis, 1992; Forgacs, 1978; Piirila y Sovijarvi, 1995). Tiempo después, Forgacs (Forgacs, 1978) plantea que los sonidos crepitantes se producen por una igualación de presión, de carácter explosivo, que se manifiesta al eliminar una barrera que separa dos compartimientos que reúnen gas a distintas presiones; causando este fenómeno, la apertura repentina de vías respiratorias pequeñas durante la inspiración, que colapsaron en la espiración (Sarkar y cols., 2015; J. Earis, 1992; Piirila y Sovijarvi, 1995). Posteriormente, Fredberg y Holford (Fredberg y Holford, 1983), concuerdan con Forgacs en la causa pero no en el fenómeno; señalando que este tipo de sonido se debe al cambio de tensión elástica (vibración) en las paredes de las vías respiratorias; que precede al cambio repentino de estado, cerrado a abierto (crepitante de inspiración) o abierto a cerrado (crepitante de espiración), de pequeñas vías respiratorias (Sarkar y cols., 2015; Piirila y Sovijarvi, 1995). Teorías más recientes, proponen que los sonidos crepitantes de inspiración son producidos por la ruptura de puentes líquidos,

que se generan por la inestabilidad mecánica de las vías respiratorias pequeñas; mientras que, los sonidos crepitantes de espiración surgen por la formación de los mismos (Almeida, Buldyrev, y Alencar, 2013).

#### Sonidos de roce pleural (pleural rub)

Los sonidos de roce pleural (pleural rub, en inglés) son sonidos respiratorios adventicios discontinuos, chirriantes, de carácter explosivo no musical; que se observan en ambas fases del ciclo respiratorio, donde la secuencia de espiración imita a la secuencia de inspiración (Bohadana y cols., 2014). Estos sonidos discontinuos se caracterizan por contar con una duración aproximada de 15 ms junto con una frecuencia menor a 350 Hz (Bohadana y cols., 2014). Los sonidos de roce pleural son causados por la fricción de las membranas pleurales al respirar (Pramono y cols., 2017). Estos sonidos adventicios están relacionados con enfermedades como: pleuritis y mesotelioma (Bohadana y cols., 2014).

### 3.3. Análisis computarizado de sonidos respiratorios

El análisis computarizado de sonidos respiratorios surge con la intención de sobrellevar las limitaciones del método tradicional (auscultación acústica) al trasladar el problema a un contexto más objetivo, tal como lo es la computación (Sovijärvi y cols., 2000; Kandaswamy y cols., 2004). Para ello, establece una aproximación sistemática, al incorporar dos elementos clave al proceso: (i) la grabación de los sonidos respiratorios, a través de dispositivos electrónicos (micrófonos o estetoscopios digitales); y (ii) la implementación de sistemas computacionales dedicados al análisis de las señales de audio (Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013; Gurung y cols., 2011; Marques y cols., 2014). Esté nuevo acercamiento, captó la atención de la comunidad científica, al inhibir las problemáticas de la auscultación acústica y mantener sus cualidades, como: el bajo costo y su carácter no invasivo; desembocando, en múltiples propuestas relacionadas al preprocesamiento, extracción de características y clasificación de las señales de audio (Pramono y cols., 2017; Gurung y cols., 2011; ARA, 2000; Palaniappan, Sundaraj, Ahamed, Arjunan, y Sundaraj, 2013). Así como también, dando pie a una nueva gama de aplicaciones, dentro del análisis de sonidos respiratorios, como: almacenar, replicar y compartir sonidos respiratorios; graficar y comparar señales de audio en distintos períodos de tiempo, detectar o clasificar automáticamente sonidos o enfermedades respiratorias; e incluso, posibilitar el monitoreo en tiempo real de pacientes a distancia (Kim y cols., 2022; Pramono y cols., 2017; Sovijärvi y cols., 2000). Recientemente, la inteligencia artificial (aprendizaje de máquina) ha tomado la batuta en el análisis computarizado de sonidos respiratorios; particularmente, el aprendizaje profundo (p. ej. redes neuronales convolucionales), permitiendo la aplicación de nuevos enfoques, como el análisis de los sonidos en formato de imagen (representaciones tiempo-frecuencia); o un análisis más preciso de los sonidos respiratorios (Kim y cols., 2022).

### Capítulo 4

### Sistema propuesto

El sistema, véase Figura 4.1, consta de tres etapas: (i) preprocesamiento, (ii) extracción de características y (iii) clasificación; de acuerdo a lo observado en el estado del arte. La etapa de preprocesamiento adecua las señales de audio según las necesidades de las fases posteriores. En este caso, igualando tanto la duración (o longitud) comolo frecuencia de muestreo de las señales de audio, reduciendo los ruidos inherentes (p. ej. latidos del corazón) a las grabaciones y transformando los audios a un formato bidimensional (imagen). La etapa de extracción de características se enfoca en generar nuevas características (o datos) a partir de los datos originales; de manera que, la información resultante sea más significativa. Dentro del sistema, esto se cumple por medio del histograma de gradientes orientados. Finalmente, la etapa de clasificación, implica el entrenamiento de algoritmos de aprendizaje de máquina; con el fin de discernir entre las diferentes clases de las señales de audio (o sonidos respiratorios). Adicionalmente, se crea un conjunto de entrenamiento aumentado, para contrarrestar el desbalance de los datos. En este capítulo, se describe a detalle cada etapa del sistema propuesto así como también la base de datos utilizada.



Figura 4.1: Etapas del sistema propuesto: (a) preprocesamiento, (b) extracción de características & (c) clasificación.

### 4.1. Descripción de la base de datos (ICBHI17)

Parte esencial para el entrenamiento de algoritmos de aprendizaje de máquina son los datos. En este contexto, audios que contengan sonidos respiratorios. A creación de una base de datos no estaba dentro de nuestros planes; por lo cual, optamos por buscar, en la literatura, una opción adecuada. Así bien, solo tres bases de datos resaltaron: R.A.L.E (Owens, 2002), ICBHI17 (Rocha y cols., 2019) y SPRSound (Zhang y cols., 2022); gracias a su contenido o uso. R.A.L.E era una base de datos dirigida a la enseñanza, por ello incluía pocas grabaciones. Además, estaba disponible únicamente bajo una licencia comercial. Esto, junto con la poca información tanto demográfica como técnica, nos hizo descartar directamente está opción. Por su parte, ICBHI17, contenía 920 grabaciones de uso libre, a la par de un grupo de participantes demográficamente diverso. Mientras que, SPRSound poseía un poco más de 2000 grabaciones, también sin ninguna restricción de uso. Aunque, se enfocaba únicamente en participantes pediátricos, exclusivamente, de origen Chino. Esta particularidad, nos inclinó a elegir a ICBHI17 como nuestra base de datos.

Dato	Detalles
No. de grabaciones	920
Herramienta (No. de grabaciones)	Meditron $(127)$ ; Littmann $3200 (60)$ ;
	Classic II (87); AKG (646)
Frecuencia de muestreo (No. de gra-	<sup>5</sup> kHz (90); 10 kHz (6); 44.1 kHz
baciones)	(824)
Bits por muestra	16
Duración promedio de las grabacio-	21.5 s
nes	
Duración promedio de los ciclos res-	2.7 s
piratorios	
Punto de adquisición (No. de graba-	A-D (168); A-I (162); P-D (132); P-I
ciones)	(139); L-D (112); L-I (77); Tráquea
	(130)

Tabla 4.1: Información técnica 47 e la base de datos

La base de datos ICBHI17 fue desarrollada para apoyar al desafío científico, sobre el análisis de sonidos respiratorios, llevado a cabo en la conferencia ICBHI International Conference on Biomedical and Health Informatics) del año 2017. Está base de datos se recopiló con ayuda de dos equipos de investigación independientes, en diferentes países (Portugal y Grecia) durante varios años. Las grabaciones se adquirieron de participantes de todas las edades; tomando como puntos de adquisición, la tráquea y seis ubicaciones torácicas: anterior (A), posterior (P) y lateral (L); izquierdo (I) y derecho (D). Asimismo, los audios se grabaron por medio de cuatro herramientas: dos estetoscopios digitales <sup>64</sup>Velch Allyn Meditron Master Elite Plus Stethoscope Model 5079 & 3M Littmann 3200), un dispositivo formado por siete estetoscopios acústicos (3M Littmann Classic II SE) con micrófonos en el tubo principal y un dispositivo que consistía de siete micrófonos electret con acoplamiento de aire (C 417 PP, AKG Acoustics); en entornos tanto clínicos como no clínico (hogares). La Tabla 4.1 muestra en detalle información técnica de la base de datos. Los participantes eran pacientes diagnosticados con: neumonía, EPOCs, asma, bronquiolitis, bronquiectasias, fibrosis quísticas, infecciones de las vías bajas (LIRT, por sus siglas en inglés) e infecciones de las vías superiores (UIRT, por sus siglas en inglés). La Tabla 4.2 muestra la información demográfica de la base de datos. En total, ICBHI17 consta de 920 grabaciones de 126 participantes, que reúnen 6898 ciclos respiratorios; de los cuáles, 3640 no contienen sonidos respiratorios adventicios, 1864 tienen sonidos crepitantes (crackles), 886 sonidos sibilantes (wheezes) y 506 ambos tipos de sonidos adventicios (wheeze+crackle: both). Cada grabación, posee anotaciones que describen los siguientes campos: inicio del ciclo respiratorio (en segundos), fin del ciclo respiratorios, presencia/ausencia (0/1) de sonidos crepitantes y presencia/ausencia de sonidos sibilantes.

También, ICBHI17 propone un marco de trabajo para el análisis comparativo (o benchmarking, en inglés) de propuestas relacionadas al análisis de sonidos respiratorios. Para ello, establece una tarea en común: clasificar los ciclos respiratorios de acuerdo a las etiquetas provistas por las anotaciones. En otras palabras, clasificar los sonidos respiratorios en: Normal, Crackle, Wheeze y Both. Asimismo, proporciona una división de la base de datos en los conjuntos de entrenamiento (60%) y prueba (40%). La Tabla 4.3 presenta la distribución de los ciclos respiratorios en cada conjunto. Además, define una serie de métricas para evaluar el desempeño de las propuestas, siendo estás métricas: Average Score (AS) (4.1), Harmonic Score (HS) (4.2), Specificity (SP) (4.3) y Sensitivity (SE) (4.4). En nuestro caso, decidimos omitir la métrica HS por su uso casi nulo en el estado del arte. Y agregar dos métricas más: (i) Accuracy (AC) (4.5) que

Dato	Detalles		
No. de participantes	126: 77 adultos, 49 menores de edad		
Sexo	79 hombres, 46 mujeres (No disponi-		
	ble: 1)		
Sexo: Menores de edades	26 hombres, 23 mujeres		
Sexo: Adultos	53 hombres, 23 mujeres (No disponi-		
	ble: 1)		
Edad (promedio $\pm$ desviación	$43.0 \pm 32.2$ años (No disponible: 1)		
estándar)			
Edad: Menores de edad	$4.8 \pm 4.6$ años		
Edad: Adultos	$67.6 \pm 11.6$ años (No disponible: 1)		
Índice de masa corporal en los adul-	27.2 ± 5.4 $kg/m^2$ (No disponible: 2)		
tos			
Peso de los menores de edad	$21.4 \pm 17.2$ kg (No disponible: 5)		
Altura de los menores de edad	$1044 \pm 30.8$ cm (No disponible: 7)		
Diagnóstico (No. de participantes)	EPOCs $(64)$ ; Saludable $(26)$ ; URTI		
	(14); Bronquiectasias $(7)$ ; Neumonía		
	(6); Bronquiolitis $(6)$ ; LRTI $(2)$ ; As-		
	$\max(1)$		

Tabla 4.2: Información demográfica de la base de datos

	Entrenamiento	Prueba	Total
Normal	2063	1579	3640
Sibilantes (Wheeze)	501	385	886
Crepitantes (Crackle)	1215	649	1864
Ambos (Both)	363	143	506

Tabla 4.3: Distribución de los ciclos respiratorios en cada conjunto.

nos permite observar que también se están clasificando los datos individualmente y (ii) Balanced Accuracy (BA) (4.6) que toma en cuenta el desbalance de las clases al evaluar la correcta clasificación de los datos.

$$AS = (SP + SE)/2 \tag{4.1}$$

$$HS = (2 * SP * SE)/(SP + SE)$$

$$(4.2)$$

$$SP = \frac{Normal_{Correcto}}{Normal_{Total}} \tag{4.3}$$

$$SE = \frac{(Crackle_{Correcto} + Wheeze_{Correcto} + Both_{Correcto})}{(Crackle_{Total} + Wheeze_{Total} + Both_{Total})}$$
(4.4)

$$AC = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

$$\tag{4.5}$$

$$BA = \frac{\sum_{i=1}^{n} \left( \frac{TP_i}{(TP_i + FN_i)} \right)}{n}$$
(4.6)

donde, TP representa los verdaderos positivos, TN los verdaderos negativos, <sup>55</sup>P los falsos positivos, FN los falsos negativos y n el número de clases.

### 4.2. Preprocesamiento

En esta etapa, las señales de audio fueron adecuadas para las fases posteriores. Primero, se estableció una sola frecuencia de muestreo para todas las grabaciones. Esto debido a que los audios fueron grabados con distintos dispositivos, y por ende, en diferentes frecuencias de muestreo; véase Tabla 4.1. Por ello, re-muestreamos las grabaciones a 4 kHz: la frecuencia mínima común; reduciendo la complejidad computacional sin afectar a las señales de interés (menores a 2000 Hz) (Sarkar y cols., 2015). A continuación, segmentamos las grabaciones, de acuerdo a las anotaciones proveídas, en los respectivos ciclos respiratorios. Los segmentos diferían en duración: oscilando entre los 0.2 y los 16 segundos, con un promedio de 2.7 segundos. Para contrarrestar este hecho, optamos por implementar técnicas de rellenado (o padding, en inglés); cuya finalidad, es igualar la longitud de un vector a un tamaño deseado, reduciendo (si es mayor) o agregando (si es menor) valores al vector original. Para ello, seleccionamos dos candidatos, denominados: Zeros y Réplica; que tenían la tarea de ajustar los ciclos respiratorios a 3 segundos. Zeros, como su nombre lo indica, aplicaba el número cero (0) como valor de rellenado; mientras que, Réplica, duplicaba parte del segmento. También, se consideró mantener la discrepancia en la duración de los segmentos; puesto que, el tamaño de la imagen final era independiente a la duración de los ciclos respiratorios (Rakotomamonjy y Gasso, 2014). La Figura 4.2 muestra el efecto de las técnicas de rellenado en las imágenes generadas.



Figura 4.2: Técnicas de rellenado: (a) Sin aplicar, (b) Zeros & (c) Réplica.

Posteriormente, implementamos un filtro Butterworth de 5to orden, sobre las señales de audio, en el rango de frecuencias [100 Hz - 1800 Hz]; para reducir los ruidos inherentes a las grabaciones. De esta manera, obteníamos una versión relativamente más fiel de las señales de interés (sonidos respiratorios); al suprimir frecuencias que podían incluir otro tipo de sonidos, como: latidos del corazón, la fricción de los dispositivos al grabar o el propio ambiente. Asimismo, se contempló conservar el estado innato de las señales de audio; debido a que, no sabíamos que impacto podría tener el filtro en los resultados finales. Por último, convertimos los segmentos en imágenes, por medio de una serie de métodos de análisis tiempo-frecuencia: STFT, MFCCs, CWT, CQT, Mel y Gammatone; que sobresalían en el estado del arte. Las imágenes obtenidas, conocidas como representaciones Tiempo-Frecuencia (TF), eran almacenadas en un formato PNG con un tamaño de 128x128 píxeles. La Figura 4.3 ilustra las imágenes resultantes de cada método.

### 4.3. Extracción de características

En esta etapa, las imágenes eran utilizadas como punto de partida para generar nuevas características (o datos); que contuvieron información relevante de los datos originales. De esta manera, se lograba reducir tanto la información redundante como el nivel de dificultad que poseían los datos iniciales (Khalid, Shehryar, y Nasreen, 2014). Para ello, aplicamos al descriptor de características: histograma de gradientes orientados, sobre las imágenes; re-escaladas a un tamaño de 64x128 píxeles. Originalmente, el HOG se propuso dentro del campo de la visión computacional, Dalal y Triggs (Dalal y Triggs, 2005), para la detección de seres humanos. Sin embargo, no tardo en expandirse a otras áreas de aplicación. Obteniendo, resultados favorables en la clasificación de escenas acústicas (Rakotomamonjy y Gasso, 2014; Bisot y cols., 2015; Abidin y cols., 2018); llegando a competir, en casos particulares, con el estado del arte de ciertas bases de datos (Rakotomamonjy y Gasso, 2014). Esto, más su inexistente uso, hasta donde alcanza nuestro conocimiento, en el ámbito del análisis de sonidos respiratorios; convirtió al histograma de gradientes orientados, en una opción razonable para la resolución del problema.

El HOG se basa en la evaluación de histogramas locales, normalizados, de la orientación del gradiente de la imagen; en una cuadricula densa (Dalal, Triggs, y Schmid, 2006). Y se rige por los siguientes pasos: (i) se calcula el gradiente de la imagen, obteniendo así, la magnitud y el ángulo asociado a cada píxel; (ii) se divide la imagen en pequeñas regiones espaciales sin traslape, denominados *celdas*, n x m píxeles; (iii) se crea un histograma de *n* puntos, espaciados uniformemente entre 0°-180° (gradiente sin signo) o 0°-360° (gradiente con signo, por



Figura 4.3: Representaciones Tiempo-Frecuencia: (a) CWT, (b) STFT, (c) MFCCs, (d) Gammatone, (e) Mel & (f) CQT.

cada celda); (iv) cada magnitud es acumulada en un punto del histograma local según su ángulo asociado, si el ángulo se encuentra entre dos puntos la magnitud es dividida proporcionalmente entre ambos; (v) las celdas son agrupadas en *bloques* de n x m; (vi) en cada bloque se emplea un proceso de normalización, con el fin de proporcionar una fuerte invarianza a la iluminación; y (vii) los histogramas de cada bloque son concatenados en un vector de características final. La Figura 4.4 ilustra el algoritmo de características de gradientes orientados.



Figura 4.4: Algoritmo del histograma de gradientes orientados: (a) imagen original, (b) gradiente de la image (magnitud: izquierda y ángulo: derecha), (c) división de la imagen en celdas, (d) histograma local, (e) bloque de normalización & (f) concatenación del vector final.

# 4.4. Clasificación (Algoritmos de aprendizaje de máquina)

En esta etapa, algoritmos de Aprendizaje de Máquina (ML, por sus siglas en inglés) eran entrenados, sobre las características del histograma de gradientes orientados, para discernir entre las distintas clases de sonidos respiratorios; que contenían la base de datos. Para ello, recurrimos a algoritmos ML tradicionales; los cuáles, en comparación con los algoritmos profundos, tienden a desempeñarse mejor en bases de datos relativamente pequeñas (Zhang y cols., 2022). Así bien, elegimos a tres candidatos: máquina de soporte vectorial, k-vecinos más cercanos y bosques aleatorios; según su uso o potencial dentro del análisis de sonidos respiratorios (Zhang y cols., 2022; Pramono y cols., 2017). A continuación, se describirá brevemente cada uno de ellos:

- Bosque aleatorio: es uno de los algoritmos de aprendizaje por conjuntos (o ensemble learning, en inglés) más populares (Chen y Ishwaran, 2012). Este busca etiquetar los datos de acuerdo a las predicciones de un grupo de árboles de decisión. Para ello, los datos de entrenamiento son remuestreados en n sub-conjuntos tipo bootstrap (Chen y Ishwaran, 2012); los cuáles se crean al tomar N datos con repetición del conjunto inicial (ordenado aleatoriamente) (Hesterberg, 2011). Cada sub-conjunto, genera un único árbol de decisión; donde cada división de nodo, se forma por las mejores características reunidas en un sub-conjunto aleatorio de las características originales. De manera que, cada árbol de decisión produce un resultado independiente a los demás. Así bien, los datos son encasillados en la clase predicha con más frecuencia.
- K-vecinos más cercanos: posee uno de los algoritmos más simples dentro del aprendizaje de máquina (Mahesh, 2019). Este intenta clasificar

nuevos datos en función a su proximidad con los datos de entrenamiento (Jiang, Cai, Wang, y Jiang, 2007). En otras palabras, asume que entidades similares se encuentran cercanas unas de otras. Para esto, considera la proximidad de k datos, determinados de acuerdo a una métrica de distancia, al punto de interés; adjudicando a este, la clase con más datos dentro de los k vecinos más cercanos.

• Máquina de soporte vectorial: es uno de los algoritmos más implementados dentro del ML (Mahesh, 2019). Este tiene como fin generar un hiperplano que divida, de forma lineal, los datos en sus respectivas clases. El hiperplano puede ser un punto (1D), una línea (2D) o un plano según la dimensión de los datos (Noble, 2006); y representa un límite de decisión, donde cada lado está asociado a una clase en particular. Pueden existir múltiples hiperplanos, sin embargo, la opción predilecta es aquel que cuenta con la mayor distancia (margen) entre los puntos más cercanos de cada categoría (vectores de soporte). En el mundo real, los datos no suelen ser linealmente separables; en cambio, son casi o no linealmente separables. Para estos casos, el algoritmo permite "suavizar" las restricciones del margen, al consentir la clasificación errónea de ciertos datos; aunque, esto implica un intercambio entre el tamaño del margen y la cantidad de clasificaciones erróneas aceptables (Noble, 2006; Mammone, Turchi, y Cristianini, 2009). Asimismo, en casos más complicados, con datos no linealmente separables; es posible implementar una técnica, conocida como: truco de kernel; la cuál permite proyectar los datos, de acuerdo a una función (o kernel), a una dimensión superior, en donde se pueda cumplir con el objetivo del algoritmo (Noble, 2006).

### 4.5. Aumento de datos

Adicionalmente, se creó otra base de datos con el conjunto de entrenamiento aumentado; para contrarrestar el desbalance, en cantidad, que existía entre las distintas clases. De esta manera, disminuíamos el efecto que ejercía la clase superior (Normal), cinco veces más grande que la clase menor (Both), en las predicciones. Para ello, implementamos la técnica de estiramiento del tiempo (time stretch, en inglés) que modificaba la velocidad, y por ende, la longitud de las señales de audio; con una proporción de cambio, uniforme aleatoria, de 0.1 a 0.3 de la longitud original.

### Capítulo 5

### Resultados

### 5.1. Reducción de las representaciones TF

Dada la variedad de representaciones tiempo-frecuencia, que aumentaban en gran medida la cantidad de experimentos, fue necesario reducir el número de opciones a la mitad; puesto que, no contábamos con el tiempo ni los recursos suficientes para llevar a cabo todos ellos. Así, creamos una base de datos por cada posibilidad; que reunía las imágenes generadas a partir de las señales de audio sin modificar (no rellenado/no filtro). Para esto, empleamos la biblioteca de software Librosa (McFee y cols., 2015), que nos permitió crear las representaciones: STFT, MFCCs, CQT y Mel; mientras que, las imágenes de tipo: CWT y Gammatone; se generaron con ayuda de la librería de software Ssqueezepy (Muradeli, 2020) y el código provisto por el repositorio de GitHub "Gammatone Filterbank Toolkit" (Heeris, 2013), respectivamente. El Apéndice A muestra los detalles de implementación de cada una de las representaciones tiempofrecuencia utilizadas. Cada base de datos fue evaluada empleando los valores de defecto, tanto del histograma de gradientes orientados como de los algoritmos ML. Para ello, utilizamos las bibliotecas de software Scikit-Image (Van Der Walt y cols., 2014) y Scikit-Learn (Kramer, 2016); con las cuáles, implementamos el histograma de gradientes orientados y los algoritmos ML, respectivamente. La Tabla 5.1 muestra los resultados individuales de los algoritmos ML en cada representación tiempo-frecuencia.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
-	SVM	52.98	32.14	49.43	73.72	25.15
CWT	KNN	50.18	28.48	46.34	72.7	19.97
	RF	54.21	28.58	49.49	81.82	17.16
	SVM	53.56	29.5	49.37	78.09	20.65
STFT	KNN	53.56	29.41	49.27	78.66	19.88
	RF	55.22	27.57	49.85	86.7	13
	SVM	53.77	28.8	49.46	79.04	19.88
MFCCs	KNN	49.31	29.47	46.21	67.45	24.98
	RF	55.99	29.06	51.07	84.8	17.33
	SVM	54.57	28.22	49.73	83.2	16.27
Gammatone	KNN	49.76	28.66	45.86	72.82	18.91
	RF	54.94	26.67	49.26	88.48	10.05
	SVM	54.68	27.42	49.29	86.26	12.32
Mel	KNN	47.39	27.25	43.96	67.45	20.48
	RF	53.27	25.49	47.36	87.84	6.88
CQT	SVM	56.97	25.08	49.82	98.8	0.85
	KNN	47.06	25.13	42.89	71.5	14.27
	RF	53.23	24.11	46.88	90.44	3.31

Tabla 5.1: Resultados de los algoritmos ML en cada Representación Tiempo-Frecuencia (RTF).

Para determinar las tres mejores opciones consideramos como principal indicador a la métrica BA; puesto que, a nuestro criterio, expresaba de mejor manera el intercambio que existía entre la correcta clasificación de los sonidos respiratorios normales y la correcta clasificación de los sonidos respiratorios adventicios (crackle, wheeze & both); al tomar en cuenta el desbalance de los datos. Así bien, observamos que las representaciones: CWT (SVM), STFT (SVM) y MFCCs (KNN); obtuvieron los mejores resultados con 32.14 %, 29.5 % y 29.47 % en la métrica BA, respectivamente. En donde, el algoritmo SVM con el método CWT, consiguió el mejor desempeño general con: 52.98 % AC, 32.14 % BA, 49.43 % AS, 73.72 % SP y 25.15 % SE. Mientras tanto, el resto de representaciones: Gammatone (KNN), Mel (SVM) y CQT (KNN); alcanzaron los siguientes resultados: 28.66 %, 27.42 % y 25.13 % en la métrica BA, respectivamente.

## 5.2. Combinación de métodos de preprocesamien-

 $\mathbf{to}$ 

Decididas las tres representaciones tiempo-frecuencia, continuamos con la evaluación de la etapa de preprocesamiento. Para esto, definimos cinco posibles combinaciones: (i) Zeros-SF (Sin Filtro), (ii) Réplica-SF, (ii) SR-Butterworth (Sin Rellenado), (iv) Zeros-Butterworth y (v) Réplica-Butterworth; que incluían tanto a las técnicas de rellenado como al filtro digital. Las primeras dos combinaciones se enfocaban en igualar únicamente las longitudes de las señales de audio sin alterar su contenido. Así bien, dentro de la combinación Zeros-SF, los modelos: SVM (CWT), KNN (MFCCs) y KNN (STFT); consiguieron los mejores resultados: 27.94%, 27.69% y 26.83% en la métrica BA, respectivamente. En donde, la representación CWT con el algoritmo SVM, logró el desempeño general más alto con: 52.97 % AC, 27.94 %, 48.13 % AS, 81.31 % SP y 14.95 % SE. La Tabla 5.2 presenta los resultados individuales de los algoritmos ML en la combinación Zeros-SF. Mientras tanto, en la combinación Réplica-SF, los modelos: KNN (MFCCs), SVM (CWT) y KNN (STFT); obtuvieron los resultados más altos con: 29.86 %, 28.7 % y 28 % en la métrica BA, respectivamente. Donde, el algoritmo KNN con el método CWT, alcanzó el mejor desempeño general: 46.8 % AC, 29.86 % BA, 43.7 % AS, 64.97 % SP v 22.42 % SE. La Tabla 5.2 muestra los resultados individuales de los algoritmos ML en la combinación Réplica-SF.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
	SVM	52.97	27.94	48.13	81.31	14.95
CWT	KNN	44.88	26.86	41.27	65.99	16.56
	RF	53.12	26.39	47.82	84.1	11.52
	SVM	52.46	25.25	46.91	84.99	8.83
MFCCs	KNN	44.95	27.69	41.86	63.07	20.64
	$\mathbf{RF}$	53.84	25.42	47.9	88.66	7.13
STFT	SVM	51.66	24.71	46.08	84.35	7.81
	KNN	48.4	26.83	44.17	73.14	15.2
	RF	52.06	24.35	46.2	86.3	6.03

Tabla 5.2: Resultados individuales de los algoritmos ML en la combinación Zeros-SF.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
	SVM	51.56	25.67	46.44	81.5	11.38
STFT	KNN	49.56	28	45.49	73.4	17.58
	RF	54.75	25.68	48.62	90.62	6.62
	SVM	53.01	28.7	48.5	79.41	17.58
CWT	KNN	46.98	26.08	42.81	71.43	14.18
	RF	52.64	26.2	47.38	83.47	11.29
MFCCs	SVM	54.57	28.87	50.03	81.12	18.94
	KNN	46.8	29.86	43.7	64.97	22.42
	RF	55.69	28.11	50.42	86.57	14.27

Tabla 5.3: Resultados individuales de los algoritmos ML en la combinación Réplica-SF.

La combinación SR-Butterworth representaba el caso opuesto a las combinaciones anteriores. Es decir, conservaba la discrepancia de longitudes y modificaba el contenido de las señales de audio, al suprimir ciertas frecuencias. Para ello, utilizamos la biblioteca de software SciPy (Virtanen y cols., 2020), que nos permitió implementar el filtro pasa-banda Butterworth. Dentro de está combinación, los modelos: SVM (CWT), SVM (STFT) y SVM (MFCCs); consiguieron los mejores resultados: 32.75%, 30.98% y 29.64% en la métrica BA, respectivamente. En donde, la representación CWT con el algoritmo SVM, alcanzó el desempeño general más alto con: 54.13% AC, 32.75% BA, 50.42% AS, 75.87% SP y 24.97% SE. La Tabla 5.4 presenta los resultados individuales de

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
	SVM	54.13	32.75	50.42	75.87	24.97
CWT	KNN	50.03	29.37	46.43	71.12	21.75
	$\mathbf{RF}$	53.77	28.77	49	81.69	16.31
	SVM	50.43	30.98	47.14	69.72	24.55
STFT	KNN	50.72	29.58	46.99	72.57	21.41
	RF	55.15	27.23	49.25	89.67	8.83
MFCCs	SVM	45.46	29.64	42.65	60.16	25.74
	KNN	38.53	27.66	36.57	50.03	23.1
	RF	46.29	28.57	43.08	65.1	21.07

los algoritmos ML en la combinación SR-Butterworth.

Tabla 5.4: Resultados individuales de los algoritmos ML en la combinación SR-Butterworth.

Finalmente, las dos últimas combinaciones se centraban en la unión de ambos aspectos. Dentro de la combinación Zeros-Butterworth, los modelos: KNN (MFCCs), KNN (STFT) y KNN (CWT); lograron los resultados más altos con: 29%, 28.38% y 27.2% en la métrica BA, respectivamente. Donde, el algoritmo KNN con el método MFCCs; consiguió el mejor desempeño general: 40.13% AC, 29% BA, 37.75% AS, 54.02% SP y 21.49% SE. La Tabla 5.5 muestra los resultados individuales de los algoritmos ML en la combinación Zeros-Butterworth. Mientras tanto, en la combinación Réplica-Butterworth, los modelos: KNN (MFCCs), SVM (CWT) y SVM (STFT); alcanzaron los mejores resultados: 29.94%, 28.59% y 27.92% en la métrica BA, respectivamente. En donde, la representación MFCCs con el algoritmo KNN; obtuvo el desempeño general más alto con: 43.21% AC, 29.94% BA, 40.46% AS, 59.34% SP y 21.58% SE. La Tabla 5.6 presenta los resultados individuales de los algoritmos ML en la combinación Réplica-Butterworth. Así bien, en base a los resultados, elegimos a la combinación SR-Butterworth para concretar la etapa de preprocesamiento.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
MFCCs	SVM	43.61	25.93	39.84	65.67	14.01
	KNN	40.13	29	37.75	54.02	21.49
	RF	45.68	25.8	41.46	70.36	12.57
	SVM	51.92	26.11	46.61	83.02	10.19
STFT	KNN	49.96	28.38	45.87	73.9	17.84
	RF	53.15	24.35	46.92	89.67	4.16
CWT	SVM	52.43	26.59	47.05	83.91	10.19
	KNN	48.04	27.2	43.64	73.78	13.5
	RF	52.14	25.65	46.59	84.61	8.58

Tabla 5.5: Resultados individuales de los algoritmos ML en la combinación Zeros-Butterworth.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
	SVM	47.46	28.82	43.99	67.76	20.22
MFCCs	KNN	43.21	29.94	40.46	59.34	21.58
	RF	47.16	28.65	43.52	68.52	18.52
	SVM	54.35	28.59	49.29	84.16	14.35
CWT	KNN	50.36	26.94	45.79	77.13	14.44
	RF	52.83	25.98	47.17	85.94	8.41
STFT	SVM	49.34	27.92	45.29	73.08	17.5
	KNN	49.05	27.72	45.25	71.31	19.2
	RF	55.58	25.23	48.95	94.42	3.48

Tabla 5.6: Resultados individuales de los algoritmos ML en la combinación Réplica-Butterworth.

### 5.3. Configuración del HOG

Concluidos los experimentos dedicados al preprocesamiento, proseguimos con la búsqueda de valores para los parámetros del histograma de gradientes orientados. Para esto, establecimos empíricamente una serie de configuraciones (CON), que implicaban el incremento o decremento de: los n puntos de cada histograma local, los n x m píxeles en cada celda o las n x m celdas en cada bloque. En total, se crearon doce configuraciones; cuatro para cada parámetro. El Apéndice B, detalla los valores de cada configuración. Las configuraciones fueron implementadas por medio de la biblioteca de software Scikit-Image (Van Der Walt y cols., 2014). Así bien, el primer grupo de configuraciones (CON 1 - CON 4), se centraba en modificar el tamaño de los bloques. Dentro de la configuración CON 1, los modelos: KNN (MFCCs), SVM (STFT) v SVM (CWT); obtuvieron los resultados más altos con: 31.75 %, 29.01 % y 28.63 % en la métrica BA, respectivamente. Donde, la representación MFCCs con el algoritmo KNN; consiguió el mejor desempeño general: 42.59 % AC, 31.75 % BA, 41.02 % AS, 51.5 % SP y 30.24 % SE. Mientras tanto, en CON 2, los modelos: SVM (CWT), KNN (MFCCs) v SVM (STFT); lograron los mejores resultados: 31.17%, 30.95% v 30.62 % en la métrica BA, respectivamente. En donde, el algoritmo SVM con el método CWT; tuvo el desempeño general más alto con: 54.1 % AC, 31.17 % BA, 49.98 % AS, 78.21 % SP y 21.75 % SE. En la configuración CON 3, los modelos: SVM (CWT), KNN(MFCCs) y SVM (STFT); consiguieron los resultados más altos con: 33.79 %, 31.02 % y 30.59 % en la métrica BA, respectivamente. Donde, el algoritmo SVM con la representación CWT: alcanzó el mejor desempeño general: 55.4% AC, 33.79% BA, 51.77% AS, 76.69% SP y 26.84% SE. Mientras que, en CON 4, los modelos: SVM (CWT), SVM (STFT) y KNN (MFCCs); lograron los mejores resultados: 34.37 %, 30.99 % y 30.75 % en la métrica BA, respectivamente. Donde, la representación CWT con el algoritmo SVM; logró el desempeño general más alto con: 55.07 % AC, 34.37 % BA, 51.52 % AS, 75.87 % SP y 27.18 % SE. La Tabla 5.7 muestra los resultados de los algoritmos ML en la serie de configuraciones CON 1 a CON 4.

La siguiente serie de configuraciones (CON 5 a CON 8) se enfocaba en cambiar el tamaño de las celdas. Así, dentro de la configuración CON 5, los modelos: SVM (CWT), SVM (STFT) y SVM (MFCCs); consiguieron los resultados más altos con: 32.07 %, 30.99 % y 28.79 % en la métrica BA, respectivamente. En donde, el algoritmo SVM con el método CWT; alcanzó el mejor desempeño general: 55.18 % AC, 32.07 % BA, 50.99 % AS, 79.73 % SP y 22.25 % SE. Por su parte, en la configuración CON 6, los modelos: SVM (CWT), KNN (STFT) y KNN

CON	RTF	Modelo	AC	BA	AS	SP	SE
			(%)	(%)	(%)	(%)	(%)
		SVM	47.2	28.12	43.42	69.34	17.5
	MFCCs	KNN	42.59	31.75	41.02	51.8	30.24
		RF	47.31	27.78	43.29	70.86	15.71
		SVM	57.47	29.01	51.99	89.55	14.44
1	STFT	KNN	49.34	26.4	45.07	74.35	15.8
		RF	57.47	26.69	50.92	95.82	6.03
		SVM	51.17	28.63	47.34	80.49	14.18
	CWT	KNN	43.75	25.84	40.22	64.47	15.97
		RF	54.28	27.92	48.73	86.76	10.7
		SVM	54.1	31.17	49.98	78.21	21.75
	CWT	KNN	49.6	29.15	46.34	68.65	24.04
		RF	53.48	27.94	48.41	83.15	13.67
		SVM	48.07	29	44.43	69.41	19.45
2	MFCCs	KNN	39.84	30.95	38.55	47.37	29.73
		RF	47.82	28.62	44.07	69.79	18.35
	STFT	SVM	50.61	30.62	47.32	69.91	24.27
		KNN	45.02	27.57	42.03	62.57	21.49
		RF	56.53	27.8	50.58	91.32	9.85
		SVM	55.4	33.79	51.77	76.69	26.84
	CWT	KNN	50.5	31.08	47.13	70.23	24.04
		RF	54.64	29.48	49.95	82.07	17.84
		SVM	49.34	30.11	45.82	69.98	21.66
3	MFCCs	KNN	41.18	31.02	39.73	49.65	29.82
		RF	47.42	28.9	43.87	68.2	19.54
		SVM	51.05	30.59	47.5	71.81	23.19
	STFT	KNN	51.7	29.95	47.74	74.92	20.56
		RF	54.46	26.41	48.58	88.85	8.32
		SVM	55.07	34.37	51.52	75.87	27.18
	CWT	KNN	50.25	31.28	46.8	70.42	23.19
		RF	53.55	29.3	48.92	80.68	17.16
		SVM	51.52	30.99	47.96	72.38	23.53
4	STFT	KNN	52.28	29.72	48.1	76.75	19.45
		RF	54.68	26.87	48.8	89.1	8.49
		SVM	49.27	30.32	45.78	69.72	21.83
	MFCCs	KNN	41.32	30.75	39.54	51.74	27.35
		RF	48.04	30.02	44.73	67.38	22.09

Tabla 5.7: Resultados individuales de los algoritmos ML en la serie de configuraciones CON 1 a CON 4

(MFCCs); lograron los mejores resultados: 32.74%, 30.25% y 29.9% en la métrica BA, respectivamente. Donde, la representación CWT con el algoritmo SVM;

obtuvo el desempeño general más alto con: 54.42 % AC, 32.74 % BA, 50.64 % AS, 76.56 % SP y 24.72 % SE. En CON 7, los modelos: SVM (CWT), KNN (MFCCs) y SVM (STFT); alcanzaron los resultaron más altos con: 32.37 %, 31.08 % y 30.46 % en la métrica BA, respectivamente. En donde, el algoritmo SVM con la representación CWT; tuvo el mejor desempeño general: 54.13 % AC, 32.37 % BA, 50.45 % AS, 75.68 % SP y 25.23 % SE. Mientras que, en la configuración CON 8, los modelos: SVM (CWT), SVM (STFT) y KNN (MFCCs); obtuvieron los mejores resultados: 31.01 %, 30.74 % y 29.57 % en la métrica BA, respectivamente. Donde, el método CWT con el algoritmo SVM; consiguió el desempeño general más alto con: 53.26 % AC, 31.01 % BA, 49.31 % AS, 76.37 % SP y 22.25 % SE. La Tabla 5.8 presenta los resultados individuales de los algoritmos ML en la serie de configuraciones CON 5 a CON 8.

Finalmente, la última serie de configuraciones (CON 9 a CON 12) se centraba en modificar la cantidad de puntos en cada histograma local. En la configuración CON 9, los modelos: SVM (CWT), KNN (STFT) y KNN (MFCCs); lograron los mejores resultados: 32.58%, 30.76% y 30.33% en la métrica BA, respectivamente. Donde, el algoritmo SVM con la representación CWT; consiguió el desempeño más alto con: 54.89% AC, 32.58% BA, 51.03% AS, 77.51% SP y 24.55% SE. Mientras que, en la configuración CON 10, los modelos: SVM (CWT), KNN (MFCCs) y KNN (STFT); alcanzaron los resultados más altos con: 32.94%, 30.31% y 30.15% en la métrica BA, respectivamente. En donde, la representación CWT con el algoritmo SVM; tuvo el mejor desempeño general: 54.71% AC, 32.94% BA, 51.01% AS, 76.37% SP y 25.65% SE. Por su parte, en CON 11, los modelos: SVM (CWT), KNN (MFCCs) y SVM (STFT); obtuvieron los mejores resultados: 32.39%, 31.25% y 30.04% en la métrica BA, respectivamente. Donde, el método CWT con el algoritmo SVM; logró el desempeño más alto con: 54.6% AC, 32.39% BA, 50.69% AS, 77.51% SP y 23.87%

CON	RTF	Modelo	AC	BA	AS	SP	SE
			(%)	(%)	(%)	(%)	(%)
		SVM	55.18	32.07	50.99	79.73	22.25
	CWT	KNN	50.65	29.79	46.75	73.46	20.05
	0.111	RF	53.8	29.11	49.06	81.57	16.56
		SVM	51.19	30.99	47.8	71.05	24.55
5	STFT	KNN	50.25	28.95	46.48	72.32	20.64
		RF	56.78	27.68	50.78	91.89	9.68
		SVM	48.87	28.79	44.9	72.13	17.67
	MFCCs	KNN	39.29	28.41	37.45	50.09	24.8
		RF	47.49	28.4	43.71	69.66	17.75
		SVM	54.42	32.74	50.64	76.56	24.72
	CWT	KNN	50.21	31.47	46.89	69.66	24.12
		RF	53.41	29.88	49.01	79.16	18.86
		SVM	49.7	30.21	46.57	68.08	25.06
6	STFT	KNN	50.47	30.25	46.91	71.31	22.51
		RF	54.28	26.76	48.63	87.33	9.94
	MFCCs	SVM	49.16	27.88	44.62	75.74	13.5
		KNN	39.62	29.9	37.94	49.46	26.42
		RF	47.6	29.56	44.17	67.7	20.64
		SVM	54.13	32.37	50.45	75.68	25.23
	CWT	KNN	50.32	30.94	46.89	70.42	23.36
		RF	53.41	29.58	49.01	79.16	18.86
		SVM	48.69	26.54	43.67	78.08	9.26
7	MFCCs	KNN	41	31.08	39.37	50.53	28.2
		RF	47.64	29.98	44.46	66.24	22.68
		SVM	50.21	30.46	46.92	69.47	24.38
	STFT	KNN	50.58	30.03	47.08	71.05	23.1
		RF	54.68	27.82	49.11	87.27	10.96
		SVM	53.26	31.01	49.31	76.37	22.25
	CWT	KNN	48.54	28.21	44.93	69.72	20.13
		RF	53.41	30.12	49.14	78.4	19.88
		SVM	50.61	30.74	47.41	69.34	25.48
8	STFT	KNN	50.68	29.42	47.04	72	22.09
		RF	54.1	27.04	48.57	86.4	10.7
		SVM	48.8	26.63	43.79	78.15	9.43
	MFCCs	KNN	38.96	29.57	37.24	49.08	25.4
		RF	46.95	29.53	43.72	65.86	21.58

Tabla 5.8: Resultados individuales de los algoritmos ML en la serie de configuraciones CON 5 a CON 8

SE. Por último, en la configuración CON 12, los modelos: SVM (CWT), SVM (STFT) y KNN (MFCCs); consiguieron los resultados más altos con: 32.23 %,
30.18 % y 29.29 % en la métrica BA, respectivamente. En donde, el algoritmos SVM con el método CWT; alcanzó el mejor desempeño general: 54.86 % AC, 32.23 % BA, 50.92 % AS, 77.89 % SP y 23.95 % SE. La Tabla 5.9 muestra los resultados individuales de los algoritmos ML en la serie de configuraciones CON 9 a CON 12. Así bien, considerando los resultados, seleccionamos a CON 4 como la configuración más apta para el histograma de gradientes orientados.

#### 5.4. Ajuste de hiper-parámetros

Por último, de manera similar al apartado anterior, buscamos los parámetros más adecuados para los algoritmos ML. Para ello, utilizamos la biblioteca de software Scikit-Learn (Kramer, 2016), que nos permitió implementar la técnica de ajuste de hiper-parámetros: Halving GridSearchCV. Está técnica se enfoca en encontrar, dentro de un espacio finito de parámetros, el mejor candidato; por medio de una sucesión de rondas, que aumentan el número de recursos en cada iteración. Cada algoritmo ML fue entrenado bajo un espacio de parámetros, relativamente pequeño, que incluía las variables más relevantes. El Apéndice C detalla los parámetros explorados por cada algoritmo. Así bien, dentro de este escenario, los modelos: RF (MFCCs), RF (CWT) y SVM (STFT); obtuvieron los mejores resultados: 38.32 %, 36.26 % y 35.28 % en la métrica BA, respectivamente. Donde, el algoritmo RF con la representación MFCCs, alcanzó el desempeño general más alto con: 42.16% AC, 38.32% BA, 41.94% AS, 43.45%SP y 40.44 % SE. La Tabla 5.10 muestra los resultados individuales de los algoritmos ML con el ajuste de hiper-parámetros. Adicionalmente, aplicamos esté mismo procedimiento sobre el conjunto de entrenamiento aumentado. Para ello, utilizamos la biblioteca de software Librosa (McFee v cols., 2015), que nos permitió aplicar la técnica: estiramiento del tiempo; en las señales de audio originales. En este caso, los modelos: SVM (CWT), SVM (MFCCs) y SVM (STFT); logra-

CON	RTF	Modelo	AC	BA	AS	SP	SE
			(%)	(%)	(%)	(%)	(%)
		SVM	54.89	32.58	51.03	77.51	24.55
	CWT	KNN	49.78	29.07	46.08	71.43	20.73
		RF	53.55	28.72	48.76	81.63	15.88
		SVM	50.14	30.4	46.74	70.04	23.44
9	STFT	KNN	52.57	30.76	48.82	74.54	23.1
		RF	55.66	27.01	49.64	90.88	8.41
		SVM	49.49	28.5	45.08	75.3	14.86
	MFCCs	KNN	40.23	30.33	38.38	51.1	25.65
		RF	46.51	28.74	43.32	65.23	21.41
		SVM	54.71	32.94	51.01	76.37	25.65
	CWT	KNN	51.34	30.7	47.83	71.88	23.78
		RF	54.93	29.81	50.1	83.21	16.99
		SVM	49.45	29.12	45.4	73.14	17.67
10	MFCCs	KNN	41.1	30.31	39.5	50.53	28.46
		RF	48.47	29.86	45.16	67.89	22.42
	STFT	SVM	50.39	29.94	46.81	71.37	22.25
		KNN	52.17	30.15	48.35	74.54	22.17
		RF	55.37	25.94	49.11	92.02	6.2
	CWT	SVM	54.6	32.39	50.69	77.51	23.87
		KNN	49.27	29	45.73	69.98	21.49
		RF	54.2	28.03	48.94	85.05	12.82
		SVM	48.8	29.35	45.09	70.48	19.71
11	MFCCs	KNN	40.23	31.25	38.74	49.01	28.46
		RF	48.51	29.06	44.72	70.67	18.77
		SVM	50.58	30.04	46.92	72	21.85
	STFT	KNN	52.57	29.82	48.4	77.01	19.79
		RF	55.47	25.58	49.03	93.22	4.84
		SVM	54.86	32.23	50.92	77.89	23.95
	CWT	KNN	48.87	29.69	45.73	67.25	24.21
		RF	54.02	27.72	48.76	84.86	12.65
	~~~~	SVM	51.81	30.18	47.87	74.85	20.9
12	STFT	KNN	52.14	29.52	48.02	76.25	19.79
		RF	56.24	25.84	49.67	94.68	4.67
	10000	SVM	48.73	29.03	44.87	71.31	18.43
	MFCCs	KNN	40.6	29.29	38.66	51.93	25.4
		RF	48.04	28.42	44.13	70.93	17.33

Tabla 5.9: Resultados individuales de los algoritmos ML en la serie de configuraciones CON 9 a CON 12

ron los resultados más altos con: 36.53%, 32.76% y 31.99% en la métrica BA, respectivamente. En donde, el algoritmo SVM con el método CWT, consiguió

el desempeño más alto: 51.78% AC, 36.53% BA, 49.45% AS, 65.42% SP y 33.47% SE. La Tabla 5.11 presenta los resultados individuales de los algoritmos ML con el ajuste de hiper-parámetros en la base de datos aumentada. Y el Apéndice D muestra los valores obtenidos en el ajuste de hiper-parámetros.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
	SVM	39.26	33.65	38.33	44.71	31.95
MFCCs	KNN	37.45	30.98	36.67	41.99	31.35
	$\mathbf{RF}$	42.16	38.32	41.94	43.45	40.44
	SVM	50.25	35.29	48.17	62.44	33.9
CWT	KNN	42.71	34.19	41.54	49.53	33.56
	$\mathbf{RF}$	45.46	36.26	44.75	49.65	39.85
	SVM	49.46	35.28	47.39	61.56	33.22
STFT	KNN	45.79	27.03	42.28	66.37	18.18
	RF	45.75	30.75	43.76	57.19	30.33

Tabla 5.10: Resultados individuales de los algoritmos ML con el ajuste de hiperparámetros.

RTF	Modelo	AC (%)	BA (%)	AS (%)	SP (%)	SE (%)
	SVM	51.78	36.53	49.45	65.42	33.47
CWT	KNN	51.05	30.07	46.85	75.68	18.01
	RF	54.06	27.35	48.48	86.76	10.2
	SVM	44.19	32.76	42.38	54.84	29.91
MFCCs	KNN	43.51	30.34	41.43	55.67	27.19
	RF	49.35	30.14	45.7	70.68	20.73
	SVM	41.87	31.99	40.46	50.16	30.76
STFT	KNN	50.73	28.31	46.79	73.78	19.8
	$\mathbf{RF}$	55.15	26.14	48.79	92.4	5.18

Tabla 5.11: Resultados individuales de los algoritmos ML con el ajuste de hiperparámetros en la base de datos aumentada.

### Capítulo 6

### Discusión

A partir de los experimentos, observamos que la representación CWT se adecuó mejor al descriptor de características HOG; puesto que, frecuentemente, acompañaba a los modelos con los desempeños más altos. Seguido del método MFCCs; de uso común en el campo del análisis de sonidos respiratorios. Mientras tanto, STFT así como Mel quedaron rezagados en comparación; a pesar de contar con una gran frecuencia de aparición en el estado del arte, véase Tabla 6.1. En donde, Mel a la par de Gammatone y CQT fueron descartados en la reducción de representaciones de tiempo-frecuencia. Esté último, implementado junto con HOG en los trabajos relacionados a la clasificación de escenas acústicas (Rakotomamonjy y Gasso, 2014; Bisot y cols., 2015). Puede que, CQT al lograr el peor desempeño, véase Tabla 5.1; indique en gran medida la diferencia de complejidad al evaluar sonidos respiratorios en comparación a las escenas acústicas. Por otra parte, en la implementación de distintas combinaciones de métodos de preprocesamiento, percibimos resultados más altos por parte de la técnica de rellenado Réplica en contraste con el método Zeros, tanto al omitir (véase Tabla 5.2 & Tabla 5.3) como al no omitir el filtro digital (véase Tabla 5.5 & Tabla 5.6). Esto probablemente sea consecuencia de la naturaleza de cada método; en donde, Zeros no agrega información relacionada a la señal de audio, por el contrario, utiliza un valor constante que puede entenderse como un espacio "vacío" en el vector, véase Figura 4.2; mientras que, Réplica si proporciona información al repetir segmentos de la señal. No obstante, ninguna de las dos técnicas supero al hecho de no alterar la duración (o longitud) de las señales de audio; ya que, la combinación SR-Butterworth concretó la etapa de preprocesamiento con los resultados más altos, véase Tabla 5.4. Adicionalmente, no notamos un gran impacto en los resultados al usar (combinación SR-Butterworth) o no usar el filtro digital (sin preprocesamiento); puesto que, estás dos alternativas difieren poco entre sí, véase Tabla 5.1 & Tabla 5.4. Ahora bien, en la configuración del histograma de gradientes orientados, observamos un comportamiento lineal al modificar el tamaño de los bloques de normalización; debido a que, conforme este valor incrementaba también lo hacía el desempeño de los mejores modelos, según la métrica BA; véase Tabla 5.7. Sin embargo, esto no sucedía con los otros dos parámetros, en los cuáles no parecía existir un cambio sustancial. De hecho, ambos, mínimamente, tendían a decrementar conforme se aumentaba el valor, véase Tabla 5.8 & Tabla 5.9. Aún así, es necesario resaltar que la configuración HOG fue realizada de manera empírica con un pequeño grupo de alternativas para cada parámetro; por esto, no podemos confirmar que lo observado se cumpla en todos los casos. Finalmente, en el ajuste de los hiper-parámetros, se obtuvieron los resultados más altos de todos los experimentos; tanto con<sup>26</sup> base de datos original (véase Tabla 5.10) como con<sup>26</sup>, base de datos aumentada (Tabla 5.11). Resaltando el modelo RF con la representación MFCCs, que alcanzo unas métricas de 42.16 % AC, 38.32 % BA, 41.94 % AS, 43.45 % SP y 40.44 % SE. Sin embargo, este modelo consiguió el mejor desempeño (según BA) a costa de clasificar correctamente los sonidos respiratorios normales (reducción de SP). Por ello, bajo nuestro criterio, determinamos que el modelo SVM con el método CWT, de la configuración de parámetros CON 4, contaba con la mejor relación entre la correcta clasificación de sonidos normales (SP) y la correcta clasificación de sonidos adventicios (SE). Teniendo los siguientes resultados: 55.07% AC, 34.37% BA, 51.52% AS, 75.87% SP y 27.18% SE. Además de poseer la métrica BA más alta antes de aplicar el ajuste de hiper-parámetros así como también el segundo resultado más alto de la métrica AS. Así bien, en comparación con el estado del arte, véase Tabla 6.1; nuestra propuesta muestra resultados consistentes en comparación a los demás trabajo; en gran medida, por la complejidad así como por el marco de trabajo de la base de datos. No obstante, logra superar a la mitad de las propuesta en la métrica AS. Aunque, queda detrás de las propuestas que utilizan algoritmos de aprendizaje profundo como extractores de características. En estos casos, si bien no sencillo competir en cuestión de métricas, por las capacidades de estás técnicas, podemos destacar beneficios del HOG como: un menor consumo energético al implementarse en hardware, menor tiempo de entrenamiento y un buen balance entre exactitud y complejidad (Ayalew, Salau, Abeje, y Enyew, 2022); que no poseen los métodos de aprendizajes profundos como CNN. Así, el HOG se posiciona como una opción viable para el análisis de sonidos respiratorios.

Método	AS (%)	SP (%)	SE (%)
MFCCs+DT (Chambres y cols., 2018)	49.43	78.05	20.81
MFCCs+HMM-GMM (Jakovljević y Lončar-	39.56	-	-
Turukalo, 2018)			
STFT-Wavelet+SVM (Serbes y cols., 2018)	49.86	-	-
Wavelet-AEA+DAG-HMM (Ntalampiras y	50.1	-	-
Potamitis, 2019)			
Gammatone+CNN-MoE (Pham, Phan, Pa-	47	68	26
laniappan, y cols., 2021)			
Gammatone+CNN-Autoencoder (Ngo y	49	69	30
cols., 2021)			
STFT+CNN-LSTM-Focal Loss (Petmezas y	64.92	82.46	47.37
cols., 2022)			
Mel+CNN-Attention-Dilatation (Chang y	52.89	69.92	35.85
cols., 2022)			
Mel+CNN-Contrastive Learning (Moummad	57.55	72.96	39.15
y Farrugia, 2023)			
Mel-MFCCs-CQT+Convolutional Vision	57.36	78.31	36.41
Transformer (Neto y cols., 2022)			
STFT-Wavelet+Bi-ResNet (Ma y cols.,	50.16	69.2	31.12
2019)			
STFT+ResNet-18-Attention (Yang y cols.,	49.55	81.25	17.8
2020)			
STFT+ResNet-Non local layer (Ma y cols.,	52.26	63.2	41.32
2020)			
Mel+ResNet-34 (Gairola y cols., 2021)	56.2	72.3	40.1
Mel+ResNet-50 (Nguyen y Pernkopf, 2022)	58.29	79.34	37.24
Mel+ResNet-Squeeze-Excitation (Tong y	50	70	30
cols., 2022)			
CWT-Mel+Inception-VGG (Pham y cols.,	57.3	85.6	30
2022)			
CWT+Inception (Pham, Phan, Schindler, y	52	73	30
cols., 2021)			
Gammatone+VGG-12-Attention (Vinh y	52.18	-	-
cols., 2022)			
CWT-HOG+SVM (Nuestro sistema)	51.52	75.87	27.18

Tabla 6.1: Comparación con el estado del arte.

### Capítulo 7

# Conclusiones y trabajos futuros

En este documento, se propuso un sistema computacional para clasificar sonidos respiratorios; en especial, sonidos respiratorios de carácter anormal (o adventicios). Para ello, implementamos al descriptor de características: histograma de gradientes orientados, originalmente concebido para la detección de seres humanos, como nuestro extractor de características; puesto que, hasta donde alcanza nuestro conocimiento, no se había aplicado en el análisis de sonidos respiratorios, pero si en la clasificación de escenas acústicas; en donde, obtuvo resultados favorables. Así, a través de una serie de experimentos, que se vieron limitados tanto por el tiempo como por los recursos disponibles, evaluamos distintas representaciones tiempo-frecuencia, métodos de rellenado y configuraciones que resaltaran el potencial del HOG. Comparando nuestra propuesta con otros trabajos guiados por las convenciones descritas por la base de datos de nuestra elección (ICBHI17). Logrando, la combinación CWT-HOG (CON 4) con la máquina de soporte vectorial, el mejor desempeño y balance entre la correcta clasificación de los sonidos respiratorios normales y la correcta clasificación de los sonidos respiratorios adventicios, con: 55.07% AC, 34.37% BA, 51.52% AS, 75.87% SP y 27.18% SE. Mostrando al histograma de gradientes orientados como un extractor viable para la resolución del problema; con resultados consistentes con el estado del arte, que superaban a la mitad de los trabajos comparados. Así bien, como futuras aproximaciones al problema, se planea aumentar el número de experimentos, centrándonos en la representación CWT y profundizando en las configuraciones de los parámetros del HOG. Además, de contemplar una estrategia más elaborada contra el desbalance de los datos, como aumentar las imágenes con técnicas de aprendizaje profundo.

#### Referencias

#### Referencias

- <sup>20</sup>bidin, S., Togneri, R., y Sohel, F. (2018, noviembre). Spectrotemporal Analysis Using Local Binary Pattern Variants for Acoustic Scene Classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(11), 2112–2121. doi: 10.1109/TASLP.2018.2854861
- <sup>50</sup>L-Khalidi, F., Saatchi, R., Burke, D., Elphick, H., y Tan, S. (2011, junio).
   Respiration rate monitoring methods: A review. *Pediatric Pulmonology*, 46(6), 523–529. doi: 10.1002/ppul.21416
- Almeida, A. B., Buldyrev, S. V., y Alencar, A. M. (2013, agosto). <sup>2</sup>rackling sound generation during the formation of liquid bridges: A lattice gas model. *Physica A: Statistical Mechanics and its Applications*, 392(16), 3409–3416. doi: 10.1016/j.physa.2013.03.038
- 27 ndrès, E., Gass, R., Charloux, A., Brandt, C., y Hentzler, A. (2018). Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *Journal of Medicine and Life*, 11(2), 89–106.
- ARA, S. (2000). Characteristics of breath sounds and adventitious respiratory sounds. *Eur Respir Rev*, 10, 591–596.
- <sup>13</sup>yalew, A. M., Salau, A. O., Abeje, B. T., y Enyew, B. (2022). Detection and classification of covid-19 disease from x-ray images using convolutional neural networks and histogram of oriented gradients. *Biomedical Signal Processing and Control*, 74, 103530.
- <sup>48</sup>ishop, P. J. (1980, junio). Evolution of the Stethoscope. Journal of the Royal Society of Medicine, 73(6), 448–456. doi: 10.1177/014107688007300611
- <sup>21</sup>isot, V., Essid, S., y Richard, G. (2015, agosto). HOG and subband power distribution image features for acoustic scene classification. En 2015 23rd European Signal Processing Conference (EUSIPCO) (pp. 719–723). Nice: IEEE. doi: 10.1109/EUSIPCO.2015.7362477

- <sup>6</sup>ohadana, A., Izbicki, G., y Kraman, S. S. (2014, febrero). Fundamentals of Lung Auscultation. New England Journal of Medicine, 370(8), 744–751. doi: 10.1056/NEJMra1302901
- <sup>24</sup>hambres, G., Hanna, P., y Desainte-Catherine, M. (2018, septiembre). Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis. En 2018 International Conference on Content-Based Multimedia Indexing (CBMI) (pp. 1–6). La Rochelle: IEEE. doi: 10.1109/ CBMI.2018.8516489
- <sup>42</sup>hang, Y., Ren, Z., Nguyen, T. T., Nejdl, W., y Schuller, B. W. (2022, marzo). Example-based Explanations with Adversarial Attacks for Respiratory Sound Analysis (n.º arXiv:2203.16141). arXiv.
- <sup>35</sup>hen, X., y Ishwaran, H. (2012, junio). Random forests for genomic data analysis. *Genomics*, 99(6), 323–329. doi: 10.1016/j.ygeno.2012.04.003
- <sup>23</sup>alal, N., y Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. En 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp. 886–893). San Diego, CA, USA: IEEE. doi: 10.1109/CVPR.2005.177
- <sup>52</sup>alal, N., Triggs, B., y Schmid, C. (2006). Human Detection Using Oriented Histograms of Flow and Appearance. En<sup>26</sup>. Leonardis, H. Bischof, y A. Pinz (Eds.), *Computer Vision – ECCV 2006* (Vol. 3952, pp. 428–441). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/11744047\_33
  - Earis, J. (1992, septiembre). Lung sounds. Thorax, 47(9), 671-672.
  - Earis, J. E., <sup>57</sup>heetham, BMG. (2000). Current methods used for computerized respiratory sound analysis. *European Respiratory Review*, 10.
- <sup>59</sup>orgacs, P. (1978, marzo). The Functional Basis of Pulmonary Sounds. *Chest*, 73(3), 399–405. doi: 10.1378<sup>84</sup>nest.73.3.399
- Fredberg, J. J., y Holford, S. K. (1983, marzo).<sup>10</sup>iscrete lung sounds: Crackles

(rales) as stress-relaxation quadrupoles. The Journal of the Acoustical Society of America, 73(3), 1036–1046. doi: 10.1121/1.389151

- <sup>12</sup>airola, S., Tom, F., Kwatra, N., y Jain, M. (2021, noviembre). RespireNet: A Deep Neural Network for Accurately Detecting Abnormal Lung Sounds in Limited Data Setting. En 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 527–530). Mexico: IEEE. doi: 10.1109/EMBC46164.2021.9630091
- Totberg, J. B., y Gavriely, N. (1989, mayo). Flutter in collapsible tubes: A theoretical model of wheezes. *Journal of Applied Physiology*, 66(5), 2262–2273. doi: 10.1152/jappl.1989.66.5.2262
- <sup>4</sup>Jurung, A., Scrafford, C. G., Tielsch, J. M., Levine, O. S., y Checkley, W. (2011, septiembre). Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis. *Respiratory Medicine*, 105(9), 1396–1403. doi: 10.1016/j.rmed .2011.05.007
- Dayes, D., y Kraman, S. S. (2009, diciembre). The Physiologic Basis of Spirometry. *Respiratory Care*, 54(12), 1717–1726.
- 62eeris,
   J.
   (2013).
   Gammatone
   Filterbank
   Toolkit.

   https://github.com/detly/gammatone.
- <sup>6</sup>Lesterberg, T. (2011). Bootstrap. Wiley Interdisciplinary Reviews: Computational Statistics, 3, 497–526.
- INEGI. (2022, octubre). Estadísticas de Defunciones Registradas 2021.
- Ionescu, C. M. (2013). <sup>80</sup>he Human Respiratory System. En *The Human Respiratory System* (pp. 13–22). London.<sup>73</sup>pringer London. doi: 10.1007/ 978-1-4471-5388-7\_2

- <sup>3</sup>akovljević, N., y Lončar-Turukalo, T. (2018). Hidden Markov Model Based Respiratory Sound Classification. En N. Maglaveras, I. Chouvarda, y P. De Carvalho (Eds.), *Precision Medicine Powered by pHealth and Connected Health* (Vol. 66, pp. 39–43). Singapore: Springer Singapore. doi: 10.1007/978-981-10-7419-6\_7
- <sup>18</sup> ang, L., Cai, Z., Wang, D., y Jiang, S. (2007). Survey of Improving K-Nearest-Neighbor for Classification. En Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) (pp. 679–683). Haikou, China: IEEE. doi: 10.1109/FSKD.2007.552
- Jr, R. L. M. 2008). In Defense of the Stethoscope. *RESPIRATORY CARE*, 53(3).
- <sup>1</sup> Andaswamy, A., Kumar, C. S., Ramanathan, Rm. Pl., Jayaraman, S., y Malmurugan, N. (2004, septiembre). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34(6), 523–537. doi: 10.1016/S0010-4825(03)00092-1
- <sup>33</sup> halid, S., Shehryar, T., y Nasreen, S. (2014, octubre). A survey of feature selection and feature extraction techniques in machine learning. Proceedings of 2014 Science and Information Conference, SAI 2014, 372–378. doi: 10.1109/SAI.2014.6918213
- <sup>8</sup> Xim, Y., Hyon, Y., Lee, S., Woo, S.-D., Ha, T., y Chung, C. (2022, marzo). The coming era of a new auscultation system for analyzing respiratory sounds. BMC Pulmonary Medicine, 22(1), 119. doi: 10.1186/s12890-022-01896-1
- Kramer, O. (2016). Scikit-Learn. En O. Kramer (Ed.),<sup>49</sup> Jachine Learning for Evolution Strategies (pp. 45–53). Cham: Springer International Publishing. doi: 10.1007/978-3-319-33383-0\_5

<sup>23</sup>Ia, Y., Xu, X., y Li, Y. (2020, octubre). LungRN+NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation. En *Interspeech 2020* (pp. 2902–2906). ISCA. doi: 10.21437/Interspeech.2020-2487

- Zia, Y., Xu, X., Yu, Q., Zhang, Y., Li, Y., Zhao, J., y Wang, G. (2019, octubre). LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using bi-ResNet Deep Learning Algorithm. En 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1–4). Nara, Japan: IEEE. doi: 10.1109/BIOCAS.2019.8919021
- Mahesh, B. (2019). Machine Learning Algorithms -A Review. doi: 10.21275/ ART20203995
- <sup>45</sup>Jammone, A., Turchi, M., y Cristianini, N. (2009). Support vector machines.
   WIREs Computational Statistics, 1(3), 283–289. doi: 10.1002/wics.49
- Aarkel, H. (2006, febrero). The Stethoscope and the Art of Listening. New England Journal of Medicine, 354(6), 551–553. doi: 10.1056/NEJMp048251
- <sup>32</sup>Jarques, A., Oliveira, A., y Jácome, C. (2014, mayo). Computerized Adventitious Respiratory Sounds as Outcome Measures for Respiratory Therapy: A Systematic Review. *Respiratory Care*, 59(5), 765–776. doi: 10.4187/respcare.02765
- CFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., y Nieto,
   O. (2015). Librosa: Audio and Music Signal Analysis in Python. En
   Python in Science Conference (pp. 18–24). Austin, Texas. doi: 10.25080/
   Majora-7b98e3ed-003
- <sup>2</sup> Jeslier, N., Charbonneau, G., y Racineux, J. (1995, noviembre). Wheezes. European Respiratory Journal, 8(11), 1942–1948. doi: 10.1183/09031936 .95.08111942
- Moummad, I., y Farrugia, N. (2023, agosto). <sup>70</sup> retraining Respiratory Sound Representations using Metadata and Contrastive Learning (n.º ar-Xiv:2210.16192). arXiv.

Muradeli, J. (2020). Squeezepy. GitHub. Note: https://github.com/OverLordGoldDragon/ssqueezepy/. doi: 10.5281/ zenodo.5080508

- <sup>36</sup>eto, J., Arrais, N., Vinuto, T., y Lucena, J. (2022, octubre). Convolution-Vision Transformer for Automatic Lung Sound Classification. En 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRA-PI) (pp. 97–102). Natal, Brazih<sup>51</sup>EEE. doi: 10.1109/SIBGRAPI55357.2022 .9991756
- <sup>15</sup>go, D., Pham, L., Nguyen, A., Phan, B., Tran, K., y Nguyen, T. (2021, abril). Deep Learning Framework Applied For Predicting Anomaly of Respiratory Sounds. En 2021 International Symposium on Electrical and Electronics Engineering (ISEE) (pp. 42–47). Ho Chi Minh, Vietnam.<sup>83</sup>EEE. doi: 10.1109/ISEE51682.2021.9418742
- <sup>1</sup>Jguyen, T., y Pernkopf, F. (2022, septiembre). Lung Sound Classification Using Co-Tuning and Stochastic Normalization. *IEEE Transactions on Biomedical Engineering*, 69(9), 2872–2882. doi: 10.1109/TBME.2022.3156293
- <sup>46</sup>oble, W. S. (2006, diciembre). What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. doi: 10.1038/nbt1206-1565
- Atalampiras, S., y Potamitis, I. (2019). Classification of Sounds Indicative of Respiratory Diseases. En J. Macintyre, L. Iliadis, I. Maglogiannis, y C. Jayne (Eds.), *forgineering Applications of Neural Networks* (Vol. 1000, pp. 93–103). Cham: Springer International Publishing. doi: 10.1007/978 -3-030-20257-6\_8
- Owens, D. (2002). Ral ang sounds 3.0. CIN: Computers, Informatics, Nursing, 5, 9–10.
- <sup>34</sup>alaniappan, R., Sundaraj, K., Ahamed, N., Arjunan, A., y Sundaraj, S. (2013). Computer-based Respiratory Sound Analysis: A Systematic Review. *IETE*

Technical Review, 30(3), 248. doi: 10.4103/0256-4602.113524

- <sup>4</sup>alaniappan, R., Sundaraj, K., y Ahamed, N. U. (2013, enero). Machine learning in lung sound analysis: A systematic review. *Biocybernetics and Biomedical Engineering*, 33(3), 129–135. doi: 10.1016/j.bbe.2013.07.001
- <sup>68</sup>asterkamp, H., Kraman, S. S., y Wodicka, G. R. (1997, septiembre). Respiratory Sounds, *Dimerican Journal of Respiratory and Critical Care Medicine*, 156(3), 974–987. doi: 10.1164/ajrccm.156.3.9701115
- <sup>41</sup>atwa, A., y Shah, A. (2015). Anatomy and physiology of respiratory system relevant to anaesthesia. *Indian Journal of Anaesthesia*, 59(9), 533. doi: 10.4103/0019-5049.165849
- Uerna, D., y Tagarelli, A. (2019, junio). Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks. En 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 50–55). doi: 10.1109/CBMS.2019.00020
- <sup>1</sup>etmezas, G., Cheimariotis, G.-A., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K., y Maglaveras, N. (2022, febrero). Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. Sensors, 22(3), 1232. doi: 10.3390/s22031232
- <sup>3</sup>ham, L., Ngo, D., Hoang, T., Schindler, A., y McLoughlin, I. (2022, enero). An Ensemble of Deep Learning Frameworks Applied For Predicting Respiratory Anomalies (n.º arXiv:2201.03054). arXiv.
- <sup>9</sup>ham, L., Phan, H., Palaniappan, R., Mertins, A., y McLoughlin, I. (2021, agosto). CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection. *IEEL*<sup>69</sup>ournal of Biomedical and Health Informatics, 25(8), 2938–2947. doi: 10.1109/JBHI.2021.3064237
- <sup>9</sup>ham, L., Phan, H., Schindler, A., King, R., Mertins, A., y McLoughlin, I. (2021, noviembre). Inception-Based Network and Multi-Spectrogram En-

semble<sup>25</sup>, pplied To Predict Respiratory Anomalies and Lung Diseases. En 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 253–256). Mexico: IEEE. doi: 10.1109/EMBC46164.2021.9629857

- <sup>2</sup>iirila, P., y Sovijarvi, A. (1995, diciembre). Crackles: Recording, analysis and clinical significance. *European Respiratory Journal*, 8(12), 2139–2148. doi: 10.1183/09031936.95.08122139
- Pramono, R. X. A., Bowyer, S., y Rodriguez-Villegas, E. (2017, mayo). Automatic adventitious respiratory sound analysis: A systematic review. *PLOS ONE*, 12(5), e0177926. doi: 10.1371/journal.pone.0177926
- Rakotomamonjy, A., y Gasso, G. (2014). <sup>31</sup> istogram of gradients of Time-Frequency Representations for Audio Scene Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 1–1. doi: 10.1109/TASLP.2014.2375575
- <sup>30</sup> eichert, S., Gass, R., Brandt, C., y Andrès, E. (2008, enero). Analysis of Respiratory Sounds: State of the Art. *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, 2, CCRPM.S530. doi: 10.4137/CCRPM.S530
- <sup>66</sup>Obinson, N. E., y Furlow, P. W. (2007). Anatomy of the respiratory system. Equine respiratory medicine and surgery, 3–17.
- Zocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P., ... de Carvalho, P. (2019, marzo).<sup>8</sup> n open access database for the evaluation of respiratory sound classification algorithms. *Physiological Measurement*, 40(3), 035001. doi: 10.1088/1361-6579/ab03ea
- <sup>2</sup>Jarkar, M., Madabhavi, I., Niranjan, N., y Dogra, M. (2015). Auscultation of the respiratory system. Annals of Thoracic Medicine, 10(3), 158–168. doi: 10.4103/1817-1737.160831
- Serbes, G., Ulukaya, S., y Kahya, Y. P. (2018). An Automated Lung Sound Pre-

processing and Classification System Based OnSpectral Analysis Methods. En N. Maglaveras, I. Chouvarda, y P. De Carvalho (Eds.), *Precision Medicine Powered by pHealth and Connected Health* (Vol. 66, pp. 45–49). Singapore: Springer Singapore. doi. 0.1007/978-981-10-7419-6\_8

- Oovijarvi, A., Dalmasso, F., Vanderschoot, J., Malmberg, L., Righini, G., y Stoneman, S. (2000). Definition of terms for applications of respiratory sounds. *European Respiratory Review*, 10(77), 597–610.
- Sovijärvi, A. Q. A., Vanderschoot, J., y Earis, J. E. (2000). Standardization of computerized respiratory sound analysis. *European Respiratory Review*.
- Strohl, K. P., Butler, J. P., y Malhotra, A. (2012, julio). Mechanical Properties of the Upper Airway. En Y. S. Prakash (Ed.), *Comprehensive Physiology* (1.<sup>a</sup> ed., pp. 1853–1872). Wiley. doi: 10.1002/cphy.c110053
- <sup>40</sup>ong, F., Liu, L., Xie, X., Hong, Q., y Li, L. (2022). Respiratory Sound Classification: From Fluid-Solid Coupling Analysis to Feature-Band Attention. *IEEE Access*, 10, 22018–22031. doi: 10.1109/ACCESS.2022.3151789
- Tu, J., Inthavong, K., y Ahmadi, G. (2013). The Human Respiratory System. En Computational Fluid and Particle Dynamics in the Human Respiratory System (pp. 19–44). Dordrecht: Springer Netherlands. doi: 10.1007/978 -94-007-4488-2\_2
- <sup>19</sup>an Der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... Yu, T. (2014, junio). Scikit-image: Image processing in Python. *PeerJ*, 2, e453. doi: 10.7717/peerj.453
- Vinh, T. Q., Chau, N. M., Nam, T. N. N., y Long, N. T. (2022, julio). Implementation<sup>43</sup>: an Irregular Lung Sound Diagnostic System. En 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE) (pp. 170–175). Nha Trang, Vietnam.<sup>43</sup>EEE. doi: 10.1109/ICCE55644.2022.9852035

- Irtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... van Mulbregt, P. (2020, marzo). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. doi: 10.1038/s41592-019-0686-2
- Weinberg, F. (1993, octubre). The history of the stethoscope. Canadian Family Physician, 39, 2223–2224.

<sup>5</sup> World Health Organization. (2020, diciembre). The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

- <sup>39</sup>ang, Z., Liu, S., Song, M., Parada-Cabaleiro, E., y Schuller, B. W. (2020, octubre). Adventitious Respiratory Classification Using Attentive Residual Neural Networks. En *Interspeech 2020* (pp. 2912–2916). ISCA. doi: 10.21437/Interspeech.2020-2790
- Zhang, Q.,<sup>22</sup>hang, J., Yuan, J., Huang, H., Zhang, Y., Zhang, B., ... Lian,
  Y. (2022).<sup>22</sup>PRSound: Open-Source SJTU Paediatric Respiratory Sound
  Database. *IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND* SYSTEMS, 99(99).

### Apéndice A

# Detalles de implementación: Representaciones TF

Parámetro	Valor asignado
n fft	256
window	hann
hop length	256
window length	256

Tabla A.1: Detalles de implementación de la representación STFT

Parámetro	Valor asignado
Wavelet	Morlet

Tabla A.2: Detalles de implementación de la representación CWT

#### 72APÉNDICE A. DETALLES DE IMPLEMENTACIÓN: REPRESENTACIONES TF

Parámetro	Valor asignado
n fft	512
n mfcc	$13 \ge 10$ frames (130)
window	hann
hop length	512
window length	512

Tabla A.3: Detalles de implementación de la representación MFCCs

Parámetro	Valor asignado
n fft	512
n mels	50
window	hann
hop length	512
window length	512

Tabla A.4: Detalles de implementación de la representación Mel

Parámetro	Valor asignado
n bins	70
window	hann
hop length	512

Tabla A.5: Detalles de implementación de la representación CQT

Parámetro	Valor asignado
n fft	2048
n filters	64
window	hann
hop length	256
window length	1024

Tabla A.6: Detalles de implementación de la representación Gammatone

## Apéndice B

# Detalles de implementación: Configuración del HOG

CON	N puntos	Tamaño de las celdas	Tamaño de los bloques
Original	9	8x8	3x3
1	9	8x8	1x1
2	9	8x8	2x2
3	9	8x8	4x4
4	9	8x8	5x5
5	9	9x9	3x3
6	9	10x10	3x3
7	9	11x11	3x3
8	9	12x12	3x3
9	10	8x8	3x3
10	12	8x8	3x3
11	15	8x8	3x3
12	18	8x8	3x3

Tabla B.1: Detalles de implementación de cada configuración del HOG

### Apéndice C

# Detalles de implementación: Espacio de parámetros

Parámetro	Valores
Kernel	linear, rbf
С	0.01,  0.1,  1,  10,  100
Gamma	0.1, 1, 10, scale, auto

Tabla C.1: Espacio de parámetros del SVM

Parámetro	Valores
n neighbors	1-10
metric	manhattan, minkowski
algorithm	ball tree, kd tree
leaf size	10-60 (cada 10)

Tabla C.2: Espacio de parámetros del KNN

### 76APÉNDICE C. DETALLES DE IMPLEMENTACIÓN: ESPACIO DE PARÁMETROS

Parámetro	Valores
estimators	50, 100, 200, 500, 1000
Criterion	gini, entropy
max features	sqrt, log2
max depth	3, 5, 8, 10, None

Tabla C.3: Espacio de parámetros del RF

Apéndice D

# Detalles de implementación: Hiper-parámetros obtenidos

### 78APÉNDICE D. DETALLES DE IMPLEMENTACIÓN: HIPER-PARÁMETROS OBTENIDOS

Representación	Modelo	Valores asignados
	SVM	rbf (kernel), 100 (c), 0.1
STFT		(gamma)
	KNN	kd tree (algorithm), 50
		(leaf size), manhattan
		(metric), 2 (n neigh-
		bors)
	RF	1000 (n estimators), en-
		tropy (criterion), sqrt
		(max features), none
		(max depth)
01110	SVM	rbf (kernel), 10 (c), 0.1
CWT		(gamma)
	KNN	ball tree (algorithm), 50
		(leaf size), manhattan
		(metric), 5 (n neigh-
		bors)
	KF	1000 (n estimators),
		(man fastures) 10
		(max leatures), 10
	SVM	$\frac{1}{10000000000000000000000000000000000$
MECCa	5 V M	(gamma)
MITCOS	KNN	kd troe (algorithm) 40
		(leaf size) minkowski
		(netric) 4 $(n neigh-$
		bors)
	BF	1000 (n estimators) en-
		tropy (criterion) sort
		(max features), none
		(max depth)
		(max uepui)

Tabla D.1: Hiper-parámetros obtenidos en la base de datos sin aumento

Representación	Modelo	Valores asignados
STFT	SVM	linear (kernel), 100 (c), scale (gamma)
	KNN	kd tree (algorithm), 20 (leaf size), minkowski (metric), 3 (n neigh- bors)
	RF	200 (n estimators), en- tropy (criterion), sqrt (max features), none (max depth)
CWT	SVM	rbf (kernel), 10 (c), sca- le (gamma)
	KNN	ball tree (algorithm), 40 (leaf size), manhattan (metric), 2 (n neigh- bors)
	RF	100 (n estimators), gini (criterion), log2 (max features), 10 (max depth)
MFCCs	SVM	rbf (kernel), 10 (c), 0.1 (gamma)
	KNN	kd tree (algorithm), 60 (leaf size), minkowski (metric), 4 (n neigh- bors)
	RF	500 (n estimators), gi- ni (criterion), sqrt (max features), none (max depth)

Tabla D.2: Hiper-parámetros obtenidos en la base de datos con aumento



### • 11% Overall Similarity

Top sources found in the following databases:

Crossref database

Crossref Posted Content database

#### TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

(	Leen Hakki, Görkem Serbes. "Wheeze Events Detection Using Convolut Crossref	<1%
, (	"Breath Sounds", Springer Science and Business Media LLC, 2018 Crossref	<1%
(	Michele Cozzatti, Federico Simonetta, Stavros Ntalampiras. "Chapter 2	<1%
(	Jo Hanna Lindsey Serato, Rosula Reyes. "Automated lung auscultation Crossref	<1%
(	Bruno M Rocha, Dimitris Filos, Luís Mendes, Gorkem Serbes et al. "An Crossref	<1%
(	Cátia Pinho, Ana Oliveira, Cristina Jácome, João Manuel Rodrigues, Al Crossref	<1%
(	L.D. Mang, F.J. Canadas-Quesada, J.J. Carabias-Orti, E.F. Combarro, J Crossref	<1%
(	Sebastian Escobar-Pajoy, Juan P. Ugarte. "Computerized analysis of p Crossref	<1%
(	Juan P. Garcia-Mendez, Amos Lal, Svetlana Herasevich, Aysun Tekin e	<1%

10	Verónica Abreu, Ana Oliveira, José Alberto Duarte, Alda Marques. "Co Crossref	<1%
11	Idiano D'Adamo, Massimo Gastaldi, Caroline Hachem-Vermette, Riccar Crossref	<1%
12	Han Sung Kim, Hong Seong Park. "Ensemble Learning Model for Classi Crossref	<1%
13	Jian-Da Wu, Yu-Hung Huang. "Enhanced Identification of Internal Casti Crossref	<1%
14	Axiu MAO, Claire Giraudet, Kai LIU, Ines De Almeida Nolasco et al. "Aut Crossref posted content	<1%
15	Nikos Dimitris Fakotakis, Stavros Nousias, Gerasimos Arvanitis, Evang <sup>Crossref</sup>	<1%
16	"IV Latin American Congress on Biomedical Engineering 2007, Bioengi Crossref	<1%
17	Jonas Reeb, Theresa Wirth, Burkhard Rost. "Variant effect predictions Crossref	<1%
18	Argha Ghosh, A. Senthilrajan. "Comparison of machine learning techni Crossref	<1%
19	Kaifosh, Patrick, Jeffrey D. Zaremba, Nathan B. Danielson, and Attila L Crossref	<1%
20	Yuanjun Zhao, Roberto Togneri, Victor Sreeram. "Replay Anti-spoofing Crossref	<1%
21	Biyun Ding, Tao Zhang, Chao Wang, Ganjun Liu, Jinhua Liang, Ruimin H Crossref	<1%

(	Diogo Pessoa, Bruno Machado Rocha, Claas Strodthoff, Maria Gomes	<1%
(	Natalya Pya Arnqvist, Blaise Ngendangenzwa, Eric Lindahl, Leif Nilsson	<1%
;	S. Jayalakshmy, Gnanou Florence Sudha. "Conditional GAN based aug	<1%
(	Wei-Bang Ma, Xiang-Yuan Deng, Yang Yang, Wai-Chi Fang. "An Effectiv Crossref	<1%
(	<b>"Detección de objetos peligrosos en equipaje usando técnicas de deep</b> Crossref posted content	<1%
(	M. Fraiwan, L. Fraiwan, M. Alkhodari, O. Hassanin. "Recognition of pul	<1%
(	<b>Jizuo Li, Jiajun Yuan, Hansong Wang, Shijian Liu, Qianyu Guo, Yi Ma, Y</b> <sup>Crossref</sup>	<1%
(	Oscar Pulido Prieto, Ulises Juarez Martinez. "Naturalistic Programmin	<1%
(	Sibghatullah I. Khan, Ram Bilas Pachori. "Automated classification of I	<1%
(	Chandrasekhar Paseddula, Suryakanth V. Gangashetty. "Late fusion fra Crossref	<1%
•	"Precision Medicine Powered by pHealth and Connected Health", Sprin	<1%
	David Fernandes Neves Oliveira. "Interpretabilidade de modelos de apr	<1%

Selim Aras, Mehmet Cinel, Ali Gangal. "Investigation of an appropriate	<1%
Vera Ignatenko, Anton Surkov, Sergei Koltcov. "Random forests with pa	<1%
C. S. Sonali, John Kiran, B. S. Chinmayi, K. V. Suma, Muhammad Easa	<1%
Danyang Wang, Xiaoyu Luo, Peter S. Stewart. "Multiple Steady and Osc	<1%
Gabriel Jaime Zapata-Zapata, Julián David Arias-Londoño, Jesús Franc	<1%
Ngan Dao Hoang, Dat Tran-Anh, Manh Luong, Cong Tran, Cuong Pham	<1%
Zhuo Chen, Dazhi Gao, Kai Sun, Xiaojing Zhao, Yueqi Yu, Zhennan Wan	<1%
Jose R. Cruz. "Interacciones entre SARS-CoV-2 y el sistema de defens	<1%
Thanh Tam Nguyen, Thanh Cong Phan, Minh Hieu Nguyen, Matthias W	<1%
Truong Quang Vinh, Ngo Minh Chau, Truong Nguyen Nhat Nam, Ngo T	<1%
Waleska Zafred Ricci. "Mortalidade por doenças cardiovasculares em Crossref posted content	<1%
Aslı Boru İpek. "chapter 13 Analysis of System Marginal Price in the Tu	<1%

46	Handbook of Statistical Bioinformatics, 2011. Crossref	<1%
47	IFMBE Proceedings, 2013. Crossref	<1%
48	"Acoustic Technologies in Biology and Medicine", Wiley, 2023 Crossref	<1%
49	Ramiro de Vasconcelos dos Santos, João Vitor Venceslau Coelho, Neli Crossref	<1%
50	Dat Tran-Anh, Nam Hoai Vu, Khanh Nguyen-Trong, Cuong Pham. "Multi Crossref	<1%
51	Jose Neto, Nicksson Arrais, Tiago Vinuto, Joao Lucena. "Convolution-V Crossref	<sup>.</sup> <1%
52	Burcu A. Urgen, Selen Pehlivan, Ayse P. Saygin. "Distinct representatio Crossref	<1%
53	P Mayorga. "Quantitative models for assessment of respiratory diseas Crossref	<1%
54	Bruno Machado Rocha, Diogo Pessoa, Alda Marques, Paulo Carvalho, Crossref	<1%
55	Daniel A. Pérez-Aguilar, Jair. M Pérez-Aguilar, Andy P. Pérez-Aguilar, R Crossref	<1%
56	Francisco Javier Méndez Landa. "Cuerpo y violencia y en México (1988 Crossref posted content	·<1%
57	Nahit Emanet, Halil R Öz, Nazan Bayram, Dursun Delen. "A comparative Crossref	<1%



"World Congress on Medical Physics and Biomedical Engineering 2006 <1% Crossref
K. A. Sohrabi, D. Basu, F. Schudt, M. Scholtes, O. Seifert, U. Koehler, V <1% Crossref
Nicolás Santiago González, María de Lourdes García-Hernández, Patric <1% Crossref
Raines, Jim M., Daniel J. Gershman, Thomas H. Zurbuchen, Menelaos <1% Crossref
Cheng Wang, Jianqiang Li, Jie Chen, Heng Zhang, Li Wang, Zun Liu. "Int Crossref
Murad Megjhani, Kalijah Terilli, Soon Bin Kwon, Daniel Nametz et al. "A <sup>Crossref</sup>
Syed Zohaib Hassan Naqvi, Mohammad Ahmad Choudhry. "An Automa <1% Crossref
Bruno M. Rocha, Diogo Pessoa, Alda Marques, Paulo Carvalho, Rui Ped <1% Crossref
K. Calciolari, K. Gravena, N.S. Bernardi, D.J. Queiroz, J.C. Lacerda Neto <1% Crossref
Patrick Reisinger, Marlies Gillis, Nina Suess, Jonas Vanthornhout et al <1% Crossref posted content
"Design of a preprocessing system for sounds obtained from chest au <1% Crossref posted content
Haroldas Razvadauskas, Evaldas Vaičiukynas, Kazimieras Buškus, Luk <1%

## turnitin<sup>®</sup>

70	Ilyass Moummad, Nicolas Farrugia. "Pretraining Respiratory Sound Re Crossref	<1%
71	Manohar Swamynathan. "Chapter 4 Step 4 - Model Diagnosis and Tuni	<1%
72	Romuere R. V. e Silva, Kelson R. T. Aires, Rodrigo M. S. Veras, Thiago S Crossref	<1%
73	"Computational Analysis of Sound Scenes and Events", Springer Scien	<1%
74	<b>"Dynamic response of tall buildings with basement considering soil-str</b> Crossref posted content	<1%
75	"International Newspaper Librarianship for the 21st Century", Walter de	<1%
76	"Localización de estacionamientos de park and ride usando metodolog Crossref posted content	<1%
77	Biliana Alexandrova-Kabadjova, Sara G. Castellanos-Pascacio, Alma L	<1%
78	Carlos I. Cortés-Martínez, Adriana I. Rodríguez-Hernández, Mariana B	<1%
79	Mohammed Tawfik, Nasser M. Al-Zidi, Ibraheam Fathail, Sunil Nimbhor	<1%
80	S. Mesic. "Computer-controlled mechanical simulation of the artificiall	<1%
81	Tulia G. Falleti, Santiago L. Cunial. "Civic programmatic participation in	<1%


82

"Intelligent Systems and Applications", Springer Science and Business ... <1% Crossref



Dat Ngo, Lam Pham, Anh Nguyen, Ben Phan, Khoa Tran, Truong Nguye... <1% Crossref



Shengkun Xie, Feng Jin, Sridhar Krishnan, Farook Sattar. "Signal featur... <1% Crossref