

Clasificación de sonidos respiratorios adventicios a través de gradientes orientados

Classification of adventitious respiratory sounds through histogram of oriented gradients

Luis Pacheco Torres¹, Jesús E. Miranda Vega², Verónica Quintero Rosas³,
Julia Díaz Escobar⁴, Mario A. Camarillo Ramos⁵

Tecnológico Nacional de México, Instituto Tecnológico de Mexicali, Mexicali, MÉXICO

¹ ORCID: 0009-0004-1017-2559 | a17490487@itmexicali.edu.mx

² ORCID: 0000-0003-0618-0455 | elias.miranda@itmexicali.edu.mx

³ ORCID: 0000-0002-8508-7429 | veronicaquintero@itmexicali.edu.mx

⁴ ORCID: 0000-0001-6941-2730 | jdiaz@itmexicali.edu.mx

⁵ ORCID: 0000-0003-0700-1885 | mario.camarillo@itmexicali.edu.mx

Recibido 05-12-2023, aceptado 05-02-2024.

Resumen

Las enfermedades respiratorias representan una de las principales causas de muerte en el mundo. La prevención, el diagnóstico oportuno y el tratamiento efectivo son pilares fundamentales para reducir la propagación de enfermedades, así como su impacto negativo en la sociedad. El presente trabajo está enfocado en desarrollar un modelo computacional, a través del uso del algoritmo Histograma de Gradientes Orientados (HOG) en conjunto con técnicas de aprendizaje de máquina, capaz de clasificar sonidos respiratorios adventicios; para apoyar en el diagnóstico oportuno de enfermedades respiratorias. Para ello, se propone utilizar al HOG como extractor de características, el cual no ha sido explorado en la literatura actual sobre la clasificación de sonidos respiratorios. De igual manera, se plantea usar distintos algoritmos de aprendizaje de máquina, como: Máquina de Soporte Vectorial (SVM), K-Vecinos más Cercanos (KNN) y Bosques Aleatorios (RF). Asimismo, evaluar el modelo según el marco de trabajo proveído por la base de datos ICBH17. La principal contribución del presente trabajo es el análisis de diferentes configuraciones del histograma de gradientes orientados para optimizar los modelos de aprendizaje de máquina. En donde, después de una serie de experimentos, nuestro mejor modelo surgió con la configuración Alternativa 4 (ALT 4); que obtuvo como resultados: 55.07% ACC, 34.37% BAL, 51.52% SCC, 75.87% SPE y 27.18% SEN.

Palabras clave: Histograma de gradientes orientados, sonidos respiratorios adventicios, transformada de Fourier de tiempo corto, transformada continua en ondículas.

Abstract

Respiratory diseases represent one of the leading causes of death in the world. Prevention, timely diagnosis and effective treatment are fundamental pillars to reduce the spread of diseases as well as their negative impact on society. The present work is focused on developing a computational model, through the use of the Histogram of Oriented Gradients (HOG) algorithm in conjunction with machine learning techniques, capable of classifying adventitious respiratory sounds; to support the timely diagnosis of respiratory diseases. For this purpose, it is proposed to use the HOG as a feature extractor, which has not been explored in the current literature on the classification of respiratory sounds. Similarly, it is proposed to use different machine learning algorithms, such as: Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Random Forests (RF). In addition, evaluate the model according to the framework provided by the ICBH17 database. The main contribution of the present work is the analysis of different configurations of the histogram of oriented gradients for optimizing machine learning models. Where, after a series of experiments, our best model emerged with the Alternative 4 (ALT 4) configuration; which obtained as results: 55.07% ACC, 34.37% BAL, 51.52% SCC, 75.87% SPE and 27.18% SEN.

Index terms: Histogram of oriented gradients, adventitious respiratory sounds, short time Fourier transform, continuous wavelet transform.

I. INTRODUCCIÓN

Las enfermedades respiratorias representan una de las principales causas de muerte en el mundo. En 2019, de acuerdo a informes de la Organización Mundial de la Salud (OMS), 3 de las 10 principales causas de decesos en la población mundial eran enfermedades que atacaban al sistema respiratorio; ocupando el tercer puesto, las Enfermedades Pulmonares Obstructivas Crónicas (EPOC), 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 [1]. Se estima que, en ese año, murieron aproximadamente 7.7 millones de personas a causa de estas enfermedades. En otras palabras, en 2019, las enfermedades respiratorias fueron responsables de casi el 14% de las muertes en el mundo [1]. En 2021, según datos del Instituto Nacional de Estadística y Geografía (INEGI), 3 de las 10 principales causas de muerte en México eran enfermedades respiratorias; tomando el primer puesto, el coronavirus (COVID-19), seguido de la neumonía e influenza con el quinto lugar y finalizando con el décimo puesto, las EPOC [2]. Estas enfermedades representaron, en 2021, cerca del 28.5% de los decesos en México; alcanzando, en conjunto, un aproximado de 320,000 muertes [2]. Esto sin lugar a dudas convierte a las enfermedades respiratorias en un asunto de gran relevancia tanto a nivel global como a nivel nacional. Los pilares fundamentales para reducir tanto la propagación de enfermedades como su impacto negativo en la sociedad se centran principalmente en la prevención, el diagnóstico oportuno y el tratamiento efectivo [3].

Entre las principales herramientas de los especialistas en la salud, para evaluar la condición del sistema respiratorio, se encuentra la auscultación pulmonar [4], [5]. Este es un procedimiento médico centrado en escuchar, habitualmente a través de un instrumento como el estetoscopio, los sonidos generados por el sistema respiratorio [6]. Los sonidos respiratorios contienen información relevante acerca de la condición de los pulmones, vías respiratorias y otros órganos afines al sistema respiratorio; en especial, los sonidos respiratorios de carácter anormal (o sonidos adventicios) [4], [6] y [7]. Entre los sonidos respiratorios adventicios más comunes se encuentran: los sonidos crepitantes (*crackles*, en inglés) y los sonidos sibilantes (*wheezes*, en inglés). Los sonidos crepitantes usualmente indican un proceso patológico en el tejido pulmonar o en las vías respiratorias; mientras que, los sonidos sibilantes habitualmente están asociados a la obstrucción de las vías respiratorias [4], [7]. Sin embargo, es necesario aclarar que los sonidos respiratorios no están sujetos a una única enfermedad o condición; en cambio, pueden aparecer en múltiples padecimientos, y en ocasiones, hasta a la vez [4]. Es decir, los sonidos respiratorios sirven como un apoyo para llegar al correcto diagnóstico [4], [7].

Existen otros procedimientos, a parte de la auscultación pulmonar, 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 [6], [8] y [9]. No obstante, parte de estos métodos, como la espirometría; requieren tanto de la motivación como de la cooperación de los pacientes para obtener resultados óptimos [10]. 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 pulmonar resalta por su bajo costo, seguridad, fácil desempeño y su carácter no invasivo [4], [5]. Aun así, a pesar de sus cualidades, este es un procedimiento subjetivo que recae en gran medida en la experiencia, habilidad y capacidad auditiva del médico que lo desempeña [6], [7] y [8]. En otras palabras, depende de la interpretación particular que tenga el especialista de la salud sobre lo que está escuchando. Esto aumenta la probabilidad de confundir o clasificar erróneamente los distintos sonidos respiratorios, en consecuencia, se puede retrasar o dificultar el diagnóstico oportuno [6].

El análisis computarizado de sonidos respiratorios surge con la intención de sobrellevar las limitaciones del método tradicional, al trasladar el problema a un contexto más objetivo como lo es la computación [8], [11]. Para ello, incorpora dos componentes claves: (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 posterior de las señales de audio [6], [12] y [13]. Esto derivó en múltiples propuestas relacionadas al preprocesamiento, extracción de características y clasificación de las señales de audio [6], [12], [14], [15]. Esto motivó al desarrollo de diversas aplicaciones, siendo una de las más importantes: la clasificación automatizada de los sonidos respiratorios [14].

En este trabajo, se propone el desarrollo de un modelo computacional capaz de clasificar sonidos respiratorios adventicios, como los sonidos crepitantes y los sonidos sibilantes; para apoyar en el diagnóstico oportuno de enfermedades respiratorias. Para ello, se plantea el uso del algoritmo HOG como extractor de características; el cual no ha sido explorado en la literatura actual para la clasificación de sonidos respiratorios; y ha obtenido resultados favorables en la clasificación de escenas acústicas [16], [17] y [18]. De igual manera, se propone utilizar distintos algoritmos de aprendizaje de máquina, como: SVM, KNN y RF. Finalizando, con la evaluación del modelo a través de las métricas proveídas por el marco de trabajo de la base de datos ICBHI17 [9].

II. TRABAJOS RELACIONADOS

Múltiples enfoques se han presentado en la clasificación de sonidos respiratorios. Desde extractores de características comunes, en el análisis de señales de audio, como: Mel Frequency Cepstral Coefficients (MFCCs) [19], [20], Short Time Fourier Transform (STFT) [21], [22] o transformadas Wavelet [21]; hasta la implementación de técnicas de aprendizaje de máquina, particularmente aprendizaje profundo, utilizando: Redes Neuronales Convolucionales (CNN) [23], [24], [25], [26], [27], transformador de visión [28] o modelos de redes ya existentes, como: ResNet [26], [29], [30], [31], [32], [33], [34], [35], Inception [35], [36] o VGG [35], [37].

Chambres et al. [19] implementa un conjunto de características, en las cuales se incluyen: características con información espectral (p. ej. bandas de energía), características de ritmo, características de tono, características SFX, valores estadísticos de las previas características (desviación estándar, promedio, etc.) y MFCCs; utilizando un Árbol de Decisión (DT) como clasificador. Mientras que, [20] emplea únicamente los coeficientes MFCCs considerando la primera derivada con un Modelo Oculto de Márkov (HMM) combinado con un Modelo Gaussiano Mixto (GMM). Serbes et al. [21] plantea descomponer las señales de audio con una transformada Wavelet y aplicar a cada parte de la señal descompuesta una STFT; utilizando SVM como clasificador. Asimismo, Ntalampiras et al. [22] usa al STFT como extractor de características con HMM para la clasificación.

En el trabajo [23], se implementa una CNN en la cual incorpora la técnica Mezcla de Expertos (MoE) sobre diferentes Representaciones Tiempo-Frecuencia (RTF), como: Log-Mel, MFCCs, Gammatone y Constant-Q Transform (CQT). Por su parte, [24] utiliza una red neuronal convolucional combinada con un autocodificador (autoencoder, en inglés), sobre una STFT y un Gammatone. Mientras que, [25] combina una CNN con una Red Neuronal de Memoria a Corto Plazo (LSTM), para clasificar las señales de audio en formato imagen; obtenidas mediante una STFT. Asimismo, [26] emplea una CNN y un modelo ResNet, integrando a la estructura un bloque dilatación-atención; sobre un espectrograma Log-Mel. De igual manera, [27] usa una CNN con la técnica de aprendizaje contrastivo, sobre la representación tiempo-frecuencia Mel. Además, [28] implementa un transformador de visión convolucional sobre las RTF: Mel, MFCCs y CQT.

Ma et al. [29] plantea utilizar un modelo ResNet Bilineal (Bi-ResNet), sobre una STFT y una transformada Wavelet. Asimismo, [30] propone un modelo ResNet-18 anexando a la estructura un bloque de atención; sobre una STFT. Por su parte, [31] utiliza un modelo ResNet incorporando a la estructura un bloque no local, sobre

una STFT. Gairola et al. [32] presenta un modelo ResNet-34 sobre un espectrograma Mel. Al igual que, Nguyen y Pernkopf [33], que usan un modelo ResNet-50 sobre un espectrograma Log-Mel. Tong et al. [34] implementa un modelo ResNet anexando a la estructura un bloque Squeeze-and-Excitation, sobre un espectrograma Mel. Pham et al. [35] realiza una comparación entre distintos modelos, como: ResNet, VGG, MobileNet, Inception, Xception y DenseNet; sobre un escalograma y un espectrograma Log-Mel. De igual manera, Pham et al. [36] emplea un modelo Inception sobre un escalograma. Mientras que, Vinh et al. [37] utiliza un modelo VGG-12 combinado con un módulo de atención, sobre un Gammatone.

4

III. MÉTODOS Y MATERIALES

La base de datos ICBHI17 [9], fue seleccionada para la evaluación del presente trabajo. Esta 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. Consta de 920 grabaciones de 126 participantes, que reúnen 6896 ciclos respiratorios, de los cuales: 3640 no contienen sonidos adventicios (normal), 1864 contienen sonidos crepitantes, 886 sonidos sibilantes y 506 contienen los sonidos adventicios previos en un mismo ciclo respiratorio (both). Las grabaciones fueron recopiladas por dos equipos de investigación independientes en distintos países (Portugal y Grecia) durante varios años; en entornos clínicos y no clínicos. Los participantes eran pacientes de todas las edades (niños y adultos mayores), diagnosticados con: neumonía, EPOCs, asma, bronquiolitis, bronquiectasias, fibrosis quística, infecciones de las vías bajas e infecciones de las vías superiores. La Tabla 1 [9] muestra la información demográfica de esta base de datos.

A. Tarea propuesta y métricas de evaluación

ICBHI17 se creó con la intención de proveer a la comunidad científica de un punto de referencia, para la evaluación de algoritmos relacionados al análisis de sonidos respiratorios [9]. Para ello, propone una tarea en común: clasificar los ciclos respiratorios en las cuatro clases antes mencionadas (Normal, Crackle, Wheeze y Both). Asimismo, proporciona un conjunto de entrenamiento (60%) y un conjunto de prueba (40%). La Tabla 2 muestra en detalle el contenido de los conjuntos. Y define una serie de métricas: Average Score (SCC) (1), Sensitivity (SEN) (2) y Specificity (SPE) (3). Adicionalmente, se agregaron las métricas: Accuracy (ACC) y Balanced Accuracy (BAL).

$$SCC = \frac{SPE + SEN}{2} \quad (1)$$

$$SEN = \frac{\text{Sonidos adventicios clasificados correctamente}}{\text{total de sonidos adventicios}} \quad (2)$$

$$SPE = \frac{\text{Sonidos normales clasificados correctamente}}{\text{total de sonidos normales}} \quad (3)$$

B. Sistema propuesto

En la Fig. 1, se muestra el sistema propuesto el cual consta de tres etapas: (i) preprocesamiento, (ii) extracción de características y (iii) clasificación. En la etapa de preprocesamiento, las señales de audio son adecuadas para las etapas posteriores, a través de: técnicas de rellenado, filtrado y la conversión de las señales de audio a imágenes. Mientras que, en la etapa de extracción de características, se utiliza al HOG para generar nuevas características a partir de los datos originales (imágenes); de manera que, las características resultantes reúnan información relevante. Finalmente, en la etapa de clasificación, se entrenan modelos de aprendizaje de máquina, a partir de los datos obtenidos en la etapa previa, con el fin de discernir entre las distintas clases de las señales de audio (o sonidos respiratorios).

5

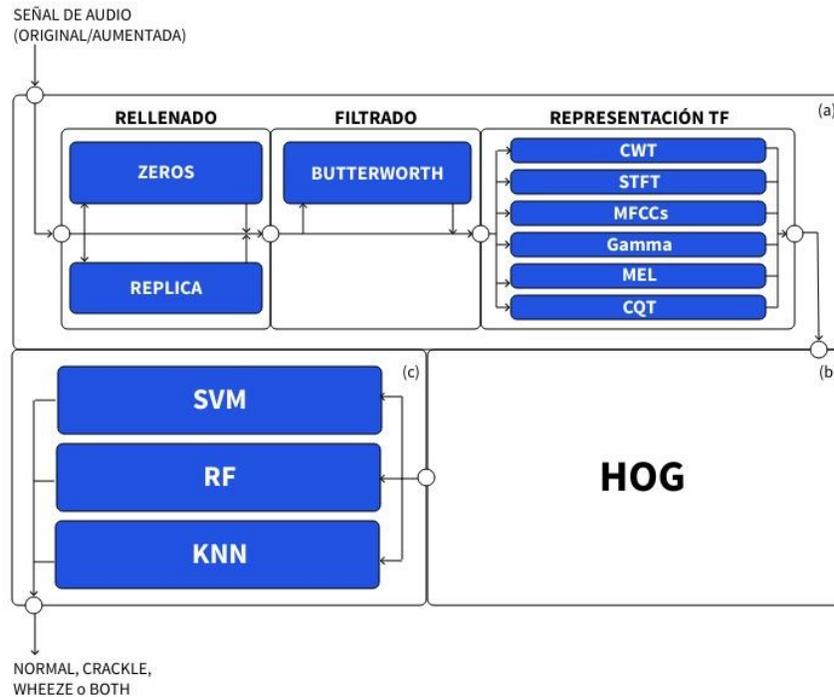


Fig. 1. Etapas del sistema: (a) preprocesamiento, (b) extracción de características, (c) clasificación.

1) Preprocesamiento

Todas las grabaciones fueron sub-muestreadas a 4 kHz con el objetivo de estandarizar los experimentos. Esta frecuencia corresponde a una de las tres frecuencias de muestreo que incluía ICBHI17, véase Tabla 1. Reduciendo el poder computacional necesario para procesar los datos, sin afectar las señales de interés (menores a 2000 Hz) [4]. A continuación, se segmentaron las grabaciones en sus respectivos ciclos respiratorios. Los segmentos diferían en duración: oscilando entre 0.2 y 16 segundos, con un promedio de 2.7 segundos. Se implementaron dos técnicas de rellenado, Zeros y Réplica, para igualar las longitudes de los ciclos respiratorios a 3 segundos. La técnica Zeros, como su nombre indica, agregaba el número cero como valor de relleno; mientras que, la técnica Réplica, duplicaba porciones de la señal. Además, se consideró mantener la discrepancia en las duraciones de los segmentos; puesto que, al convertir las señales de audio a imágenes, el tamaño de la imagen era independiente a la duración de las señales. La Fig. 2 muestra las técnicas de relleno utilizadas.

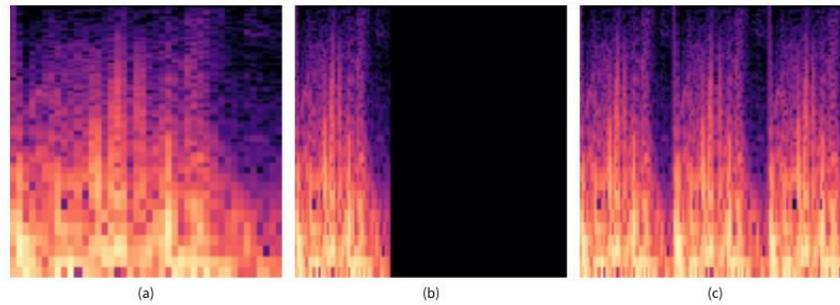


Fig. 2. Métodos de rellenado: (a) longitud original, (b) Zeros & (c) Réplica.

Después, se aplicó un filtro pasa-banda Butterworth de quinto orden, en el rango de frecuencias 100-1800 Hz; para suprimir los ruidos inherentes a las grabaciones, como: los latidos del corazón o el roce de los dispositivos con el cuerpo. También, se evalúa la omisión del filtro. Finalmente, se transformaron las señales de audio a imágenes, usando los métodos: CWT, Mel, STFT, MFCCs, Gammatone y CQT; para generar representaciones tiempo-frecuencia.

2) Extracción de características

Se eligió al histograma de gradientes orientados como extractor de características. Originalmente, este algoritmo fue propuesto dentro de la visión computacional para la detección de personas [38]. Sin embargo, no tardó en expandirse a otras áreas de aplicación como la clasificación de escenas acústicas, donde obtuvo resultados favorables [16], [17] y [18]. El HOG se basa en la evaluación de histogramas locales, normalizados, de la orientación del gradiente de la imagen en una cuadrícula densa [39]. Y funciona de acuerdo a los siguientes pasos: (i) se obtiene el gradiente de la imagen (magnitud y ángulo de cada píxel), (ii) se divide la imagen en regiones espaciales denominadas celdas ($n \times m$ píxeles), (iii) se crea un histograma de n puntos, espaciados uniformemente entre 0-180 grados, por cada celda; (iv) se acumula la magnitud de los píxeles en los n puntos del histograma en función al ángulo asociado al píxel, (v) se agrupan las celdas en bloques de $n \times m$ celdas, (vi) se aplica un método de normalización por cada bloque y (vii) se concatenan los histogramas de cada bloque para crear un vector de características final.

3) Clasificación

Diversos modelos de aprendizaje de máquina se han implementado en el ámbito de la clasificación de sonidos respiratorios. Tanto algoritmos tradicionales, como: SVM [21] o HMM [20], [22]; hasta algoritmos de aprendizaje profundo, tales como: CNN [23], [24], [25], [26], [2], ResNet [26], [29], [30], [31], [32], [33], [34], [35] o transformador de visión [28]. Para llevar a cabo esta tarea, se seleccionó a los algoritmos de aprendizaje de máquina tradicionales: SVM, KNN y RF; por su amplio uso en la literatura, así como por su desempeño en bases de datos pequeñas [40].

TABLA 1

INFORMACIÓN DEMOGRÁFICA DE LA BASE DE DATOS (NA: NO DISPONIBLE) [9]

LT que cuentan con CS		Estado
1	Número de grabaciones	920
2	Frecuencia de muestreo (número de grabaciones)	4 kHz (90); 10 kHz (6); 44.1 kHz (824)
3	Bits por muestra	16
4	Duración promedio de las grabaciones	21.5 s
5	Número de participantes	126: 77 adultos, 49 niños
6	Sexo	79 hombres, 46 mujeres (NA: 1)
7	Edad (promedio \pm desviación estándar)	43.0 \pm 32.2 años (NA: 1)
8	Edad de los adultos participantes	67.6 \pm 11.6 años (NA: 1)
9	Edad de los niños participantes	4.8 \pm 4.6 años (NA: 1)
10	IMC (Índice de Masa Corporal) de los adultos participantes	27.6 \pm 5.4 kg/m ² (NA: 2)
11	Peso de los niños participantes	21.4 \pm 17.2 kg (NA: 5)
12	Altura de los niños participantes	104.7 \pm 30.8 cm (NA: 7)

TABLA 2

DETALLES DE LA DIVISIÓN DE ICBHI17

Sonido respiratorio	Conjunto de entrenamiento	Conjunto de prueba
Normal	2063	1579
Crackles	1215	649
Wheezes	501	385
(Both) Wheezes+Crackles	363	143

IV. RESULTADOS Y DISCUSIONES

Se utilizó un servidor Linux, con un procesador Intel Xeon Gold y 64 GB de memoria RAM, para llevar a cabo los experimentos. La primera ronda de experimentos, consistió en reducir las RTF con el fin de disminuir tanto el tiempo como los recursos utilizados. Para ello, se evaluó cada representación sin la etapa de preprocesamiento, con el propósito de obtener los resultados más fieles al uso de cada tipo de imagen. Asimismo, se consideró a la métrica BAL como la principal guía en el desempeño de los modelos; puesto que, proporcionaba de manera más clara una pista acerca de la correcta clasificación tanto de los sonidos respiratorios normales como de los sonidos respiratorios adventicios. De esta manera, CWT, STFT y MFCC; en ese orden, consiguieron los resultados más altos. La Fig. 3 muestra los valores en la métrica BAL de cada una de las representaciones tiempo-frecuencia.

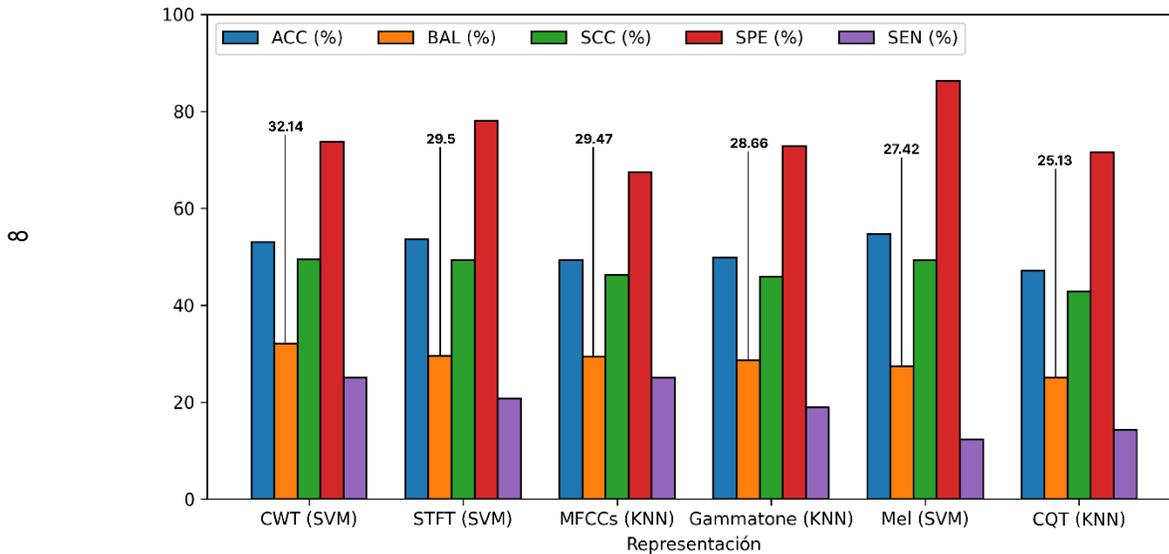


Fig. 3. Desempeño de los modelos en cada RTF considerando la métrica BAL.

En contraste, CQT, Mel y Gammatone; en ese orden, alcanzaron los peores resultados. Por lo cual, los siguientes experimentos, se realizaron con la primera tercia de RTF. Nótese que, a diferencia de los artículos relacionados a la aplicación del HOG en escenas acústicas, la representación CQT, no se adaptó bien a los sonidos respiratorios; al igual que Mel, que suele ser una de las representaciones más frecuentes en el estado del arte, junto con STFT y MFCC. Mientras que, CWT, obtuvo los mejores resultados; quizás por ello, últimamente, su aparición en el estado del arte esté en aumento.

Los experimentos posteriores, se centraron en evaluar distintos caminos para la etapa de preprocesamiento. Primero, se contempló utilizar únicamente las técnicas de rellenado sin el filtro pasa-banda. En donde, se observó que la técnica Réplica superaba a la técnica Zeros (véase Fig. 4 y Fig. 5). Esto, probablemente, debido a que Réplica proporcionaba información a la imagen; mientras que, Zeros generaba un vacío en la misma, denotado por la franja negra. Después, se experimentó con el caso opuesto; es decir, sin técnicas de rellenado y con el filtro pasa-banda. Esta combinación, alcanzó los resultados más favorables; atribuibles, posiblemente, al hecho de no modificar las señales de audio originales. Finalmente, se evaluó la combinación de las técnicas de rellenado con el filtro pasa-banda. En este caso, no se consiguieron resultados más grandes a los experimentos anteriores (solo el filtro pasa-banda); aunque, si se observó de nuevo, que la técnica Réplica obtenía resultados más altos que la técnica Zeros. Así, la combinación sin aplicar técnica de rellenado más filtro pasa-banda concretó la etapa de preprocesamiento.

9

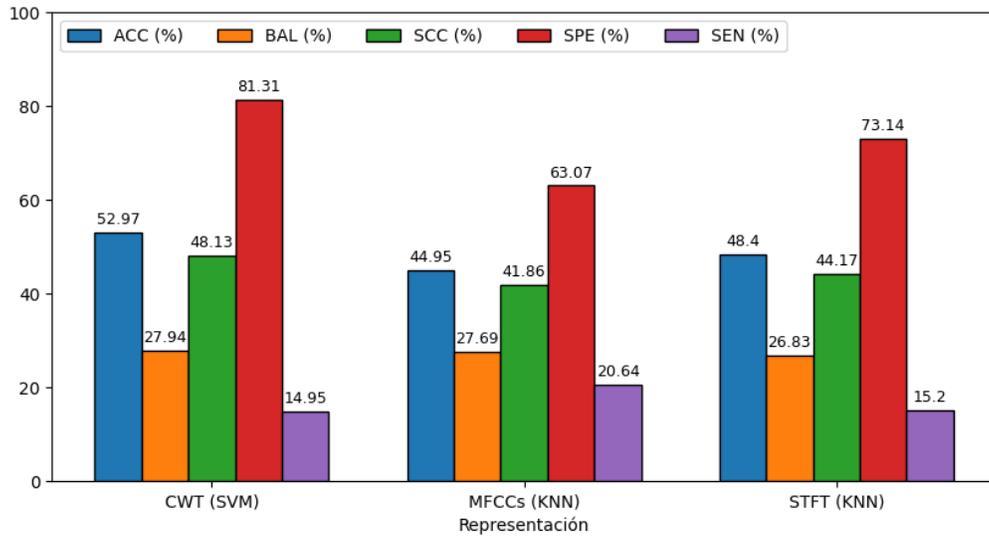


Fig. 4. Desempeño de los modelos en la combinación Zeros – sin Filtro.

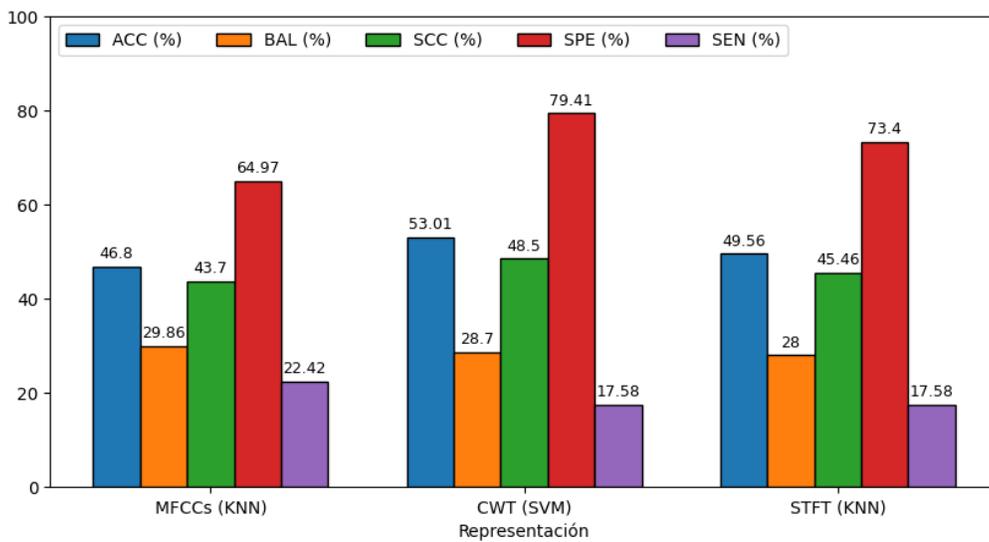


Fig. 5. Desempeño de los modelos en la combinación Réplica – sin Filtro

A continuación, se llevó a cabo una serie de configuraciones empíricas sobre los parámetros del HOG. Estas configuraciones, implicaban el incremento o decremento del valor original. Se realizaron 4 configuraciones por cada uno de los parámetros principales (3) del histograma de gradientes orientados. La Tabla 3 ilustra las modificaciones realizadas. La primera serie de configuraciones, se enfocaban en el tamaño de los bloques. En donde, se observó que al incrementar o al reducir el valor, también lo hacía el desempeño. Dentro de esta serie, la Alternativa 4 (ALT 4), alcanzó los resultados más altos. Las Tablas 4, 5 y 6 muestran los resultados de la configuración ALT 4.

TABLA 3

MODIFICACIONES DE LOS PARÁMETROS DEL HOG

Parámetro	Original	ALT 1	ALT 2	ALT 3	ALT 4	ALT 5	ALT 6	ALT 7	ALT 8	ALT 9	ALT 10	ALT 11	ALT 12
N puntos	9	-	-	-	-	9	-	-	-	10	12	15	18
Tamaño de las celdas	8, 8	-	-	-	-	9, 9	10, 10	11, 11	12, 12	8, 8	-	-	-
Tamaño de los bloques	3, 3	1, 1	2, 2	4, 4	5, 5	3, 3	-	-	-	3, 3	-	-	-

10

TABLA 4

CONFIGURACIÓN ALT 4 HOG (STFT)

Modelo	ACC (%)	BAL (%)	SCC (%)	SPE (%)	SEN (%)
SVM	51.52	30.99	47.96	72.38	23.53
KNN	52.28	29.72	48.1	76.75	19.45
RF	54.68	26.87	48.8	89.1	8.49

TABLA 5

CONFIGURACIÓN ALT 4 HOG (CWT)

Modelo	ACC (%)	BAL (%)	SCC (%)	SPE (%)	SEN (%)
SVM	55.07	34.37	51.52	75.87	27.18
KNN	50.25	31.28	46.8	70.42	23.19
RF	53.55	29.3	48.92	80.68	17.16

TABLA 6

CONFIGURACIÓN ALT 4 HOG (MFCCs)

Modelo	ACC (%)	BAL (%)	SCC (%)	SPE (%)	SEN (%)
SVM	49.27	30.32	45.78	69.72	21.83
KNN	41.32	30.75	39.54	51.74	27.35
RF	48.04	30.02	44.73	67.38	22.09

En contraste, este comportamiento no se observó en la serie de configuraciones relacionadas al tamaño de las celdas ni en la serie de configuraciones centradas en los n puntos del histograma local. En cambio, las modificaciones no parecían tener ningún efecto en el desempeño. Tanto ALT 6 como ALT 10 consiguieron los resultados más altos en sus respectivas series, tamaño de las celdas y n puntos del histograma local, respectivamente. No obstante, en general, el mejor resultado se obtuvo en la primera serie de configuraciones con ALT 4. Las Tablas 7, 8 y 9 muestran los mejores resultados de las configuraciones ALT 4, ALT 6 y ALT 10.

TABLA 7

MEJORES MODELOS DE LA CONFIGURACIÓN ALT 4

Modelo	ACC (%)	BAL (%)	SCC (%)	SPE (%)	SEN (%)
CWT (SVM)	55.07	34.37	51.52	75.87	27.18
STFT (SVM)	51.52	30.99	47.96	72.38	23.53
MFCC (KNN)	41.32	30.75	39.54	51.74	27.35

11

TABLA 8

MEJORES MODELOS DE LA CONFIGURACIÓN ALT 6

Modelo	ACC (%)	BAL (%)	SCC (%)	SPE (%)	SEN (%)
CWT (SVM)	54.42	32.74	50.64	76.56	24.72
STFT (KNN)	50.47	30.25	46.91	71.31	22.51
MFCC (KNN)	39.62	29.9	37.94	49.46	26.42

TABLA 9

MEJORES MODELOS DE LA CONFIGURACIÓN ALT 10

Modelo	ACC (%)	BAL (%)	SCC (%)	SPE (%)	SEN (%)
CWT (SVM)	54.71	32.94	51.01	76.37	25.65
MFCCs (KNN)	41.1	30.31	39.5	50.53	28.46
STFT (KNN)	52.17	30.15	48.35	74.54	22.17

Por último, se implementó un ajuste de hiper-parámetros, utilizando tanto la base de datos original como una base de datos aumentada (por medio de time stretching). En ambos casos, los resultados en la métrica BAL eran superiores a los obtenidos hasta el momento. Sin embargo, este aumento era a costa de la métrica SPE. Por lo cual, el mejor modelo obtenido con el algoritmo HOG, era la combinación de SVM con la representación CWT de ALT 4; puesto que, presentaba un mejor balance entre las métricas SPE y SEN.

V. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo se presentó un sistema centrado en la clasificación de sonidos respiratorios, que utilizaba al algoritmo histograma de gradientes orientados como extractor de características; descriptor, originario del campo de la visión computacional, que ha mostrado resultados favorables en la clasificación de escenas acústicas. La contribución principal de este trabajo se enfoca en análisis de diferentes configuraciones del HOG para la optimización de los modelos de aprendizaje de máquina. En donde, después de una serie de experimentos, la combinación de SVM con la representación CWT de la configuración ALT 4, alcanzo los mejores resultados con: 55.07% ACC, 34.37% BAL, 51.52% SCC, 75.87% SPE y 27.18% SEN; consiguiendo el mejor balance entre las métricas SPE y SEN, de las configuraciones. Como trabajos a futuro, se plantea implementar mejoras en la etapa de preprocesamiento, con la inclusión de distintos filtros digitales como: filtro de la media, mediana, Savitzky-Golay, Kalman, entre otros. Al igual que, en la segmentación de las grabaciones, al considerar la entropía para eliminar partes donde no existe señal respiratoria. Con el fin de aumentar el rendimiento de las métricas antes mencionadas.

AGRADECIMIENTOS

Estos autores agradecen al Consejo Nacional de Humanidades, Ciencias y Tecnologías (Conahcyt) y al Tecnológico Nacional de México (TecNM) por el apoyo brindado para realizar este proyecto. De la misma manera se extiende este agradecimiento al Tecnológico Nacional de México / IT de Mexicali por usar los recursos informáticos e instalaciones.

REFERENCIAS

12

- [1] “The top 10 causes of death,” *Who.int*. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [Accessed: 24-Nov-2023].
- [2] “Comunicado de Prensa Núm. 600/22 26 de octubre de 2022 página 1/92,” *Org.mx*. [Online]. Available: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2022/EDR/EDR2021_10.pdf [Accessed: 24-Nov-2023].
- [3] D. Perna, A. Tagarelli, “Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks”, 2019 *IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 50-55, 2019, doi: <https://doi.org/10.48550/arXiv.1907.05708>
- [4] M. Sarkar, I. Madabhavi, N. Niranjani, M. Dogra, “Auscultation of the respiratory system,” *Ann. Thorac. Med.*, vol. 10, no. 3, p. 158, 2015, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4518345/pdf/ATM-10-158.pdf>
- [5] A. Bohadana, G. Izbicki, S. S. Kraman, “Fundamentals of lung auscultation,” *N. Engl. J. Med.*, vol. 370, no. 8, pp. 744–751, 2014, <https://www.nejm.org/doi/full/10.1056/nejmra1302901>
- [6] R. Palaniappan, K. Sundaraj, N. Ahamed, A. Arjunan, S. Sundaraj, “Computer-based respiratory sound analysis: A systematic review,” *IETE Tech. Rev.*, vol. 30, no. 3, p. 248, 2013, <https://www.tandfonline.com/doi/abs/10.4103/0256-4602.113524>
- [7] S. Reichert, R. Gass, C. Brandt, E. Andrès, “Analysis of respiratory sounds: State of the art,” *Clin. Med. Circ. Respirat. Pulm. Med.*, vol. 2, p. CCRPM.S530, 2008, <https://doi.org/10.4137/CCRPMP>
- [8] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, N. Malmurugan, “Neural classification of lung sounds using wavelet coefficients,” *Comput. Biol. Med.*, vol. 34, no. 6, pp. 523–537, 2004, [https://doi.org/10.1016/S0010-4825\(03\)00092-1](https://doi.org/10.1016/S0010-4825(03)00092-1)
- [9] B. M. Rocha *et al.*, “An open access database for the evaluation of respiratory sound classification algorithms,” *Physiol. Meas.*, vol. 40, no. 3, p. 035001, 2019, <https://doi.org/10.1088/1361-6579/ab03ea>
- [10] D. Hayes Jr, S. S. Kraman, “The physiologic basis of spirometry,” *Respir. Care*, vol. 54, no. 12, pp. 1717–1726, 2009, <https://rc.rcjournal.com/content/54/12/1717>
- [11] A.R.A. Sovijärvi, J. Vanderschoot, J.E. Earis, “Standardization of computerized respiratory sound analysis,” *European Respiratory Review*, 2000, <https://www.ers-education.org/lr/show-details/?idP=37502>
- [12] A. Gurung, C. G. Scrafford, J. M. Tielsch, O. S. Levine, W. Checkley, “Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis,” *Respir. Med.*, vol. 105, no. 9, pp. 1396–1403, 2011, <https://doi.org/10.1016/j.rmed.2011.05.007>
- [13] A. Marques, A. Oliveira, C. Jácome, “Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: A systematic review,” *Respir. Care*, vol. 59, no. 5, pp. 765–776, 2014, <https://doi.org/10.1016/j.rmed.2011.05.007>
- [14] R. X. A. Pramono, S. Bowyer, E. Rodriguez-Villegas, “Automatic adventitious respiratory sound analysis: A systematic review,” *PLoS One*, vol. 12, no. 5, p. e0177926, 2017, <https://doi.org/10.1371/journal.pone.0177926>
- [15] J. E. Earis, B. M. G. Cheetham, “Current methods used for computerized respiratory sound analysis”, *European Respiratory Review*, vol. 10, no. 77, pp. 586-590, 2000, https://www.researchgate.net/publication/327528906_Current_methods_used_for_computerized_respiratory_sound_analysis
- [16] A. Rakotomamonjy, G. Gasso, “Histogram of gradients of time-frequency representations for audio scene classification,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 142–153, 2014, <https://doi.org/10.1109/TASLP.2014.2375575>
- [17] S. Abidin, R. Togneri, F. Sohel, “Spectrotemporal analysis using local binary pattern variants for acoustic scene classification,” *IEEE ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 11, pp. 2112–2121, 2018, <https://doi.org/10.1109/TASLP.2018.2854861>
- [18] V. Bisot, S. Essid, G. Richard, “HOG and subband power distribution image features for acoustic scene classification,” in *2015 23rd European Signal Processing Conference (EUSIPCO)*, 2015, <https://doi.org/10.1109/EUSIPCO.2015.7362477>
- [19] G. Chambres, P. Hanna, M. Desainte-Catherine, “Automatic detection of patient with respiratory diseases using lung sound analysis,” in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, 2018, <https://doi.org/10.1109/CBMI.2018.8516489>

- [20] N. Jakovljevic, T. Šcar-Turukalo, “Hidden Markov Model Based Respiratory Sound Classification,” in *Precision Medicine Powered by pHealth and Connected Health*, vol. 66. Singapore; Singapore: Springer, 2018, <https://link.springer.com/book/10.1007/978-981-10-7419-6>
- [21] G. Serbes, S. Ulukaya, Y. P. Kahya, *An Automated Lung Sound Preprocessing and Classification System Based On Spectral Analysis Methods*,” in *Precision Medicine Powered by pHealth and Connected Health (N. Maglaveras, I. Chouvarda, and)*, vol. 66. Singapore; Singapore: Springer, 2018.
- [22] I. Ampiras, “Classification of Sounds Indicative of Respiratory Diseases,” in *Engineering Applications of Neural Networks*, vol. 1000. Cham: Springer International Publishing, 2019.
- [23] H. Pham, R. Phan, A. Palaniappan, I. Mertins, “CNNMoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 2938–2947, 2021, <https://doi.org/10.48550/arXiv.2004.04072>
- [24] D. Ngo, L. Pham, A. Nguyen, B. Phan, K. Tran, T. Nguyen, “Deep learning framework applied for predicting anomaly of respiratory sounds,” in *2021 International Symposium on Electrical and Electronics Engineering (ISEE)*, 2021, <https://doi.org/10.1109/ISEE51682.2021.9418742>
- [25] G. Petmezas *et al.*, “Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function,” *Sensors (Basel)*, vol. 22, no. 3, p. 1232, 2022, <https://doi.org/10.3390/s22031232>
- [26] Y. Chang, Z. Ren, T. T. Nguyen, W. Nejdil, B. W. Schuller, “Example based Explanations with Adversarial Attacks for Respiratory Sound Analysis,” *Proc. Interspeech 2022*, pp. 4003–4007, 2022, <https://doi.org/10.21437/Interspeech.2022-11355>
- [27] I. Moummad, N. Farrugia, “Pretraining respiratory sound representations using metadata and contrastive learning,” *arXiv [cs.SD]*, 2022, <https://doi.org/10.48550/arXiv.2210.16192>
- [28] J. Neto, N. Arrais, T. Vinuto, J. Lucena, “Convolution-Vision Transformer for Automatic Lung Sound Classification,” in *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Natal, Brazil: IEEE, 2022, pp. 97–102, <https://doi.org/10.1109/SIBGRAPI55357.2022.9991756>
- [29] Y. Ma *et al.*, “LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using biResNet Deep Learning Algorithm,” in *2019 IEEE Biomedical Circuits and Systems Conference*, Nara, Japan: IEEE, 2019, pp. 1–4, <https://doi.org/10.1109/BIOCAS.2019.8919021>
- [30] Z. Yang, S. Liu, M. Song, E. Parada-Cabaleiro, B. W. Schuller, “Adventitious respiratory classification using attentive residual neural networks,” in *Interspeech 2020*, 2020, <https://doi.org/10.21437/Interspeech.2020-2790>
- [31] Y. Ma, X. Xu, Y. Li, “LungRN+NL: An improved adventitious lung sound classification using non-local block ResNet neural network with mixup data augmentation,” in *Interspeech 2020*, 2020, <https://doi.org/10.21437/Interspeech.2020-2487>
- [32] S. Gairola, F. Tom, N. Kwatra, M. Jain, “RespireNet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2021, pp. 527–530, 2021, <https://doi.org/10.48550/arXiv.2011.00196>
- [33] T. Nguyen, F. Pernkopf, “Lung sound classification using co-tuning and stochastic normalization,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 9, pp. 2872–2882, 2022, <https://doi.org/10.48550/arXiv.2108.01991>
- [34] F. Tong, L. Liu, X. Xie, Q. Hong, L. Li, “Respiratory sound classification: From fluid-solid coupling analysis to feature-band attention,” *IEEE Access*, vol. 10, pp. 22018–22031, 2022, <https://doi.org/10.1109/ACCESS.2022.3151789>
- [35] L. Pham, D. Ngo, T. Hoang, A. Schindler, I. McLoughlin, “An ensemble of deep learning frameworks applied for predicting respiratory anomalies,” *arXiv [cs.SD]*, 2022, <https://doi.org/10.48550/arXiv.2201.03054>
- [36] L. Pham, H. Phan, A. Schindler, R. King, A. Mertins, I. McLoughlin, “Inception-based network and multi-spectrogram ensemble applied to predict respiratory anomalies and lung diseases,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2021, pp. 253–256, 2021, <https://doi.org/10.1109/EMBC46164.2021.9629857>
- [37] T. Q. Vinh, N. M. Chau, T. N. N. Nam, N. T. Long, “Implementation of an irregular lung sound diagnostic system,” in *2022 IEEE Ninth International Conference on Communications and Electronics (ICCE)*, 2022, <https://doi.org/10.1109/ICCE55644.2022.9852035>
- [38] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2005, <https://doi.org/10.1109/CVPR.2005.177>
- [39] N. Dalal, B. Triggs, C. Schmid, “Human detection using oriented histograms of flow and appearance”, in *Lecture Notes in Computer Science*, pp. 428–441, 2006, <https://inria.hal.science/inria-00548587/document>
- [40] Q. Zhang *et al.*, “SPRSound: Open Source SJTU Pediatric Respiratory Sound Database,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 99, no. 99, 2022, <https://doi.org/10.36227/techrxiv.20422479.v1>