Audio Cough Analysis by Parametric Modelling of Weighted Spectrograms to Interpret the Output of Convolutional Neural Networks

P. Amado-Caballero*, J. R. Garmendia-Leiza, M.D. Aguilar-García, C. Martínez-Fernández-de-Septiem, L. M. San-José-Revuelta, A. García-Ruano, C. Alberola-López, P. Casaseca-de-la-Higuera

***Abstract*—This study explores the feasibility of employing eXplainable Artificial Intelligence XAI methodologies for the analysis of cough patterns in respiratory diseases. A cohort of 20 adult patients, all presenting persistent cough as a symptom of respiratory disease, was monitored for 24 hours using a smartphone. The audio signals underwent frequency domain transformation to yield 1-second spectrograms, subsequently processed by a CNN to detect cough events. Quantitative analysis of spectrogram regions relevant for cough detection highlighted by occlusion maps, revealed significant differences between patient groups. Notably, distinctions were observed between the Chronic Obstructive Pulmonary Disease (COPD) patient group and groups with other respiratory pathologies, both chronic and non-chronic. In conclusion, interpretability analysis methods applied to neural networks offer insights into cough-related distinctions among patients with varying respiratory conditions.**

***Index Terms*—Respiratory diseases, cough, audio analysis, CNN, XAI, occlusion maps.**

I. Introduction

Respiratory diseases, like COPD and cancer, are one of the major causes of death [1]. These chronic conditions often lead to dependence and disability. Recent studies [2] emphasize the prolonged need for care and home monitoring after COVID-19, regardless of hospitalization. The impact of COVID-19 on chronic respiratory diseases, such as COPD or cancer, heightens the risk of critical hospitalization and death [3]. Continuous monitoring of respiratory conditions is thus crucial for identifying and managing exacerbations in these patients.

The European Commission's telemedicine study [4] highlighted the potential of telemedicine in managing respiratory diseases but emphasizes a pressing need for more research in this area. Despite its promise for affordable monitoring [5], telemedicine has fallen short due to the lack of reliable objective measures for symptoms. Improved remote consultations depend on obtaining such metrics, leading to early diagnoses and real-time monitoring for respiratory patients [6].

There has been recently a surge in utilizing deep learning methods to address the challenge of cough detection and analysis for diagnostic purposes. Methods such as those in [7], [8], demonstrate significant outcomes in both detection and diagnosis. Despite outperforming conventional machine learning approaches, they operate like black boxes, hindering the ability to interpret whether coughs exhibit specific characteristics linked to particular diseases or stages thereof.

This paper presents a methodology based on eXplainable Artificial Intelligence (XAI) to interpret specific spectral signatures of cough sounds in chronic respiratory diseases. The initial step involves employing a convolutional neural network (CNN) to detect cough sounds through spectrograms. Occlusion maps representing activated regions of the time-frequency patterns after being processed by the CNN are extracted and used to weight the spectrograms. After that, a Gaussian mixture model is applied to the emphasized spectrograms. The model parameters are further analysed to identify significant differences among diseases and interpret the underlying reasons. To our knowledge, the proposed methodology has not been so far applied to cough patterns.

II. Materials

An observational study of cough evolution during 24 hours of a patient's normal life was carried out. Twenty-four hours of audio from ambulatory patients in the Palencia Health Area (Spain) were prospectively recorded. The database consists of approximately 15,000 cough events corresponding to 20 patients aged between 23 and 87 years (9 women, 11 men) with the following respiratory pathologies: Acute respiratory disease (ARD, 3), pneumonia (4), COPD (6), lung cancer (3), and others such as asthma, bronchiectasis or sarcoidosis (remaining patients). A Sony Xperia Z2 Android smartphone was used to collect the data using 16-bit WAV format at 44.1 kHz. The following sets were defined for comparison:

- G1: Chronic vs. Non-chronic patients.
- G2: COPD patients vs. other diseases.
- G3: COPD patients vs. other diseases excluding cancer.
- G4: COPD patients vs. ARD and pneumonia patients.
- G5: COPD patients vs. other chronic diseases.
- G6: COPD patients vs. lung cancer patients.

This work is part of the project TED2021-131536B-I00, funded by Spanish MCIN/AEI/10.13039/501100011033 and EU's "NextGenerationEU"/PRTR.

P. Amado, L.M. San-José, C. Alberola and P. Casaseca are with the Laboratory of Image Processing (LPI), E.T.S. Ingenieros de Telecomunicación, Universidad de Valladolid, Valladolid, Spain.

J. R. Garmendia and M. D. Aguilar are with Centro de Salud Los Jardinillos, SACYL, Palencia, Spain.

C. Martínez is with Hospital Universitario de Burgos, SACYL, Burgos, Spain.

A. García is with Centro de Salud de Alaejos, SACYL, Alaejos (Salamanca), Spain.

(*) Corresponding author: pamacab@lpi.tel.uva.es

This is the final manuscript (post-print) accepted and published in the Proceedings of the 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Orlando, Florida, USA, July 15-19, 2024. ISBN: 979-8-3503-7149-9, ISSN: 2694-0604. DOI: 10.1109/EMBC53108.2024.10781781.

III. Methods

A. Audio Signal Preprocessing

The 44.1 kHz audio signal is resamplet at 5x, to reach a final frequency of 8.82 kHz. The power spectral density (PSD) is then calculated in 10 ms windows, employing a non-overlapping Hanning window. Afterward, these PSDs are concatenated over 1 s intervals, forming a set of 45×100 spectrograms. These time-frequency representations undergo logarithmic normalization, creating input images tailored for the convolutional neural network.

B. Cough Window Identification

To distinguish spectrograms corresponding to cough events, we used a convolutional neural network designed from scratch. This network is formed by a convolutional layer of 32 filters, 2×2 *kernels* and a ReLu activation function. A 2×2 *Max-Pool* layer is incorporated followed up by a dropout layer to respectively reduce dimensionality and overfitting. This arquitecture is repeated, doubling the number of filters. The last four layers before the output are a convolutional layer (128 filters) followed by a dropout layer and a convolutional layer (256 filters) followed by a *Max-Pool* layer. The output of this arquitecture is resized to feed the output architecture, formed by two *Fully-Connected* layers, one of 512 neurons followed by a ReLU activation and the other with two neurons and *softmax* activation. Figure 1 shows a block diagram of the employed CNN.

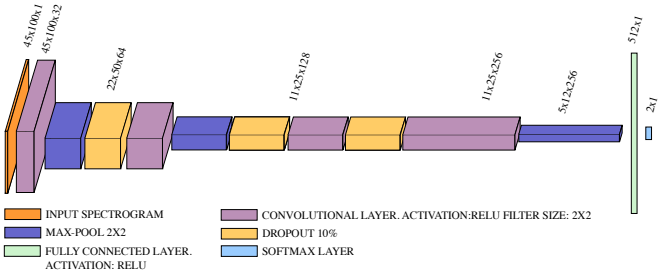

Fig. 1. CNN architecture for cough detection.

Training was carried out using the AdaMax optimizer ($\alpha = 0.002$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) [9], batch size$= 128$ and 50 epochs. To fine-tune the hyperparameters and determine the optimal stopping point for training, we employed a validation set comprising 20% of the training dataset, which itself constituted 70% of the entire audio clips collection.

C. XAI methodology for cough analysis

The foundation of the here proposed XAI methodology lies on our recent proposal [10] for ADHD diagnosis. This method consists in:

1) Occlusion maps: For each patient, we select those spectrograms for which the network output identified cough with confidence levels higher than 90%. After that, heat maps were created using occlusion maps [11] to highlight relevant regions in cough spectrograms. This technique defines a mask which is placed over the input spectrograms hiding part of their information to the CNN. The class probability of the CNN fed by this input allows estimating the importance of the hidden area for classification (higher probability means lower importance). The process is repeated after sliding the mask until the whole spectrogram is covered. The output values are stored in a matrix that is resized to have the same dimensions as the original input. After all the maps were created for the selected cough spectrograms, they were normalized and pixel-averaged so that each patient has a unique occlusion map associated with his/her cough.

2) Weighted Spectrograms: Differently to [10], where the occlusion maps were directly analysed, we weighted each patient's pixel-averaged spectrogram with their corresponding occlusion map. This way, only the relevant regions were further analysed (areas with occlusion map values below 0.7 were ignored). Figure 2 show examples of the median weighted spectrograms for each of the studied groups.

3) Parametric modelling and hypothesis testing: Quantitative analysis is carried out by fitting a Gaussian Mixture Model (GMM) to the weighthed spectrograms. Let $ws(x, y)$ be the weighted spectrogram at point $(x, y) \in \chi \subset \mathbb{Z}^2$ of the 2D grid. We define $f(x, y; \boldsymbol{\theta}_i)$ as the 2-dimensional Gaussian density function where $\boldsymbol{\theta_i}$ denote the function parameters. Two main activation regions can be seen in the occlusion maps (one at lower frequencies and the other at higher values, see), so a 2-Gaussian model was chosen. The model is fitted using the following expression:

$$\min_{p_1,p_2,\boldsymbol{\theta_1},\boldsymbol{\theta_2}} \sum_{(x,y)\in\chi} \left(ws(x,y) - \sum_{i=1}^{2} p_i f(x,y;\boldsymbol{\theta}_i) \right)^2 \quad (1)$$

being

$$\boldsymbol{\theta}_i = \left[\begin{array}{ccccc} \eta_{x_i} & \sigma_{x_i} & \eta_{y_i} & \sigma_{y_i} & \rho_{x_i y_i} \end{array} \right], \quad (2)$$

the parameters of the Gaussian function, where η denotes the means, σ the standard deviations and ρ the Pearson correlation coefficient of the two components. Subscripts refer to the direction, with x for the horizontal axis (time) and y for the vertical (frequency).

To check for statistically significant differences between the groups defined in section 2, the following procedure was applied:

- First, the two Gaussians were sorted according to the value of p_i in equation (1). The one with lowest p_i value was defined as non-dominant.
- G1–G6 sets were compared using hypothesis testing. Initially, a Gaussianity test was executed on each sample; if Gaussianity persisted in both, we resorted to an unpaired Student's t-test. Alternatively, if Gaussianity was dismissed, we turned to the Mann-Whitney U-test.
- Once the corresponding test was performed for all the parameters in (2), we obtained *boxplots* for the cases in which significant differences ($p-$value lower than 0.05) were detected.

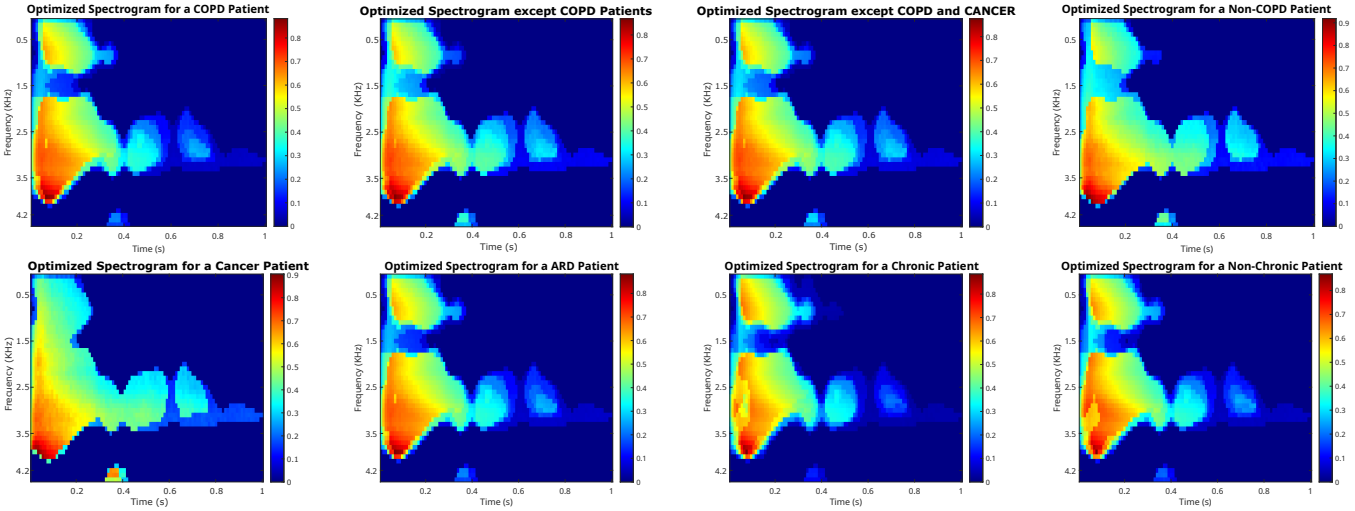

Fig. 2. Median Weighted Spectrogram for the study groups

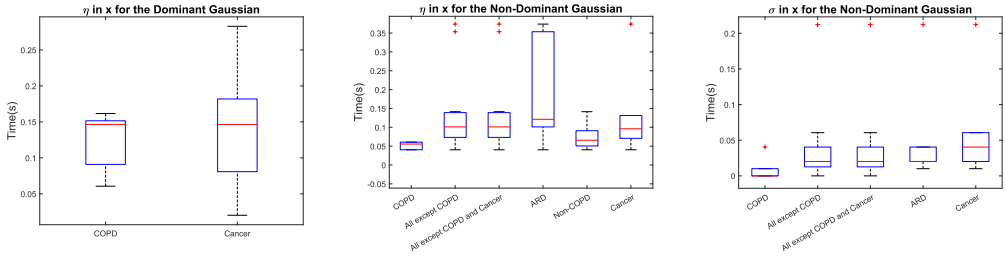

Fig. 3. Boxplots obtained for parameters with significant differences.

IV. Results and Discussion

Tables I (dominant Gaussian) and II (non-dominant Gaussian) show parameters' test results for the different compared groups G1–G6. Results are presented in terms of mean ± SEM (Standard Error of the Mean) or [median, IQR (interquartile range)] depending on wether the Student's t-test or the Mann-Whitney U-test were performed. Comparisons for which $p < 0.05$ are highlighted and the corresponding illustrated using boxplots in Figure 3.

Significant differences were found for the following groups and parameters:

- Dominant Gaussian's η_x in G6 ($\eta_{x-COPD} < \eta_{x-cancer}$, $p = 0.0427$). The dominant Gaussian can be identified in the examples in Figure 2 as the region with highest energy, encompassing frequency and time ranges of 1.75–2.75 kHz and 0–0.8s respectively. Examples in Figure 2 show that the two separate regions after 0.4s have lower energy for COPD patients than for cancer patients, thus shifting the estimated time mean to be lower. Therefore, the higher frequency's cough energy can be interpreted as being more concentrated at the beginning of the event.
- Non-dominant Gaussian's η_x is significantly lower for COPD patients in groups G2–G6. The same applies to σ_x except for group G5. The non-dominant Gaussian can be found in the examples in Figure 2 at the lower frequency regions (0–1 kHz). COPD patients have higher energy in this region at the beginning of the event, which shifts both mean and std estimation to lower time values. The boxplots in Figure 3 show these meaningful differences, which can be interpreted as lower frequency energy being concentrated at the beginning of the cough event for COPD patients.

V. Conclusion

This study showed the feasibility of using XAI techniques to understand cough differences in various patient groups. Significant differences between chronic COPD patients and those with other chronic and non-chronic conditions were found. The methods used help identify which parts of a cough sound are crucial for the neural network in each disease, revealing specific temporal patterns in COPD patients. However, it's important to note that the study is in the early stages, as the patient groups for some diseases were small. Despite this, the techniques show potential. With larger groups of patients, there's hope for more significant results, providing a better understanding of the symptoms in different respiratory conditions.

References

[1] World Health Organisation, *The Global Impact of Respiratory Disease*. European Respiratory Society, 2017.

TABLE I
RESULTS OBTAINED FOR THE p-VALUES OF THE DOMINANT GAUSSIAN IN ALL STUDY GROUPS

		p	η_x	η_y	σ_x	σ_y	ρ_{xy}
G1	**Chronic**	[0.9540,0.1351]	[0.1296,0.0905]	[1.3722,0.7692]]	[0.0410,0.0608]	[0.1259,0.2124]	-0.0203±0.1760
	Non-Chronic	[0.8877,0.2877]	[0.1400,0.0520]	[2.0055,0.3513]	[0.0167,0.0414]	[0.2734,0.2785]	0.2088±0.1871
	p	0.47472	0.8868	0.1331	0.1613	0.1331	0.3966
G2	**COPD**	[0.9334,0.1236]	[0.1567,0.0616]	[1.9162,2.4617]	[0.0335,0.0292]	[0.1438,0.2111]	[0.0259,0.4867]
	Other diseases	[0.9461,0.2917]	[0.1385,0.0864]	[1.4895,0.8478]	[0.0265,0.0537]	[0.2129,0.3043]	[0.2302,0.6752]
	p	0.3011	0.6605	0.5249	0.8836	0.5249	0.8196
G3	**COPD**	[0.0666,0.1236]	[0.0647,0.0222]	[1.9262,2.0573]	[0.0016,0.0135]	[0.1328,1.4760]	[0.0434,1.1864]
	Other except Cancer	[0.9461,0.3240]	[0.0930,0.0770]	[1.6199,0.8920]	[0.0265,0.0532]	[0.2734,0.2887]	[2177,0.6180]
	p	0.3277	0.2238	0.6070	0.6889	0.2238	0.8977
G4	**COPD**	0.9001±0.0503	[0.1567,0.0616]	[1.9162,2.4617]	[0.0335,0.0292]	[0.1438,0.2111]	-0.0127±0.3115
	ARD	0.8206±0.0695	[0.1393,0.0646]	[2.0199,0.4512]	[0.0216,0.0508]	[0.3057,0.2896]	0.0771±0.1572
	p	0.3763	0.5887	0.9372	0.4848	0.1797	0.8821
G5	**COPD**	0.9001±0.0503	[0.1567,0.0616]	[1.9162,2.4617]	[0.0335,0.0292]	[0.1438,0.2111]	[0.0259,0.4867]
	Non-COPD	[0.9463,0.2877]	[0.08,0.004710]	[1.0806,0.1910]	[0.0432,0.0744]	[0.2129,0.3169]	[0.2177,0.9587]
	p	0.5341	0.0952	0.1667	0.9048	0.7143	0.6718
G6	**COPD**	0.9001±0.0503	0.1348±0.0173	2.2711±0.5165	0.0379±0.0103	0.1411±0.0536	0.0325±0.2477
	Cancer	0.9081±0.0640	0.2372±0.0518	1.4798±0.0097	0.1509±0.1289	0.0466±0.0466	0.5525±0.6439
	p	0.9364	**0.0427**	0.1861	0.5413	0.3825	0.3175

TABLE II
RESULTS OBTAINED FOR THE p-VALUES OF THE NON-DOMINANT GAUSSIAN IN ALL STUDY GROUPS

		p	η_x	η_y	σ_x	σ_y	ρ_{xy}
G1	**Chronic**	[0.0460,0.1351]	[0.0786,0.0347]	[1.3722,2.0573]	[0.0410,0.0453]	[0.0909,1.5079]	-0.1255±0.2388
	Non-Chronic	[0.1123,0.2877]	[0.14,0.2259]	[1.2739,1.0620]	[0.0221,0.0224]	[0.0320,0.1351]	0.2716±0.1957
	p	0.4747	0.1088	0.0702	0.4747	0.9623	0.2479
G2	**COPD**	[0.0666,0.1236]	[0.0647,0.0222]	[1.9262,2.0573]	[0.0016,0.0135]	[0.1328,1.4760]	[0.0434,1.1864]
	Other diseases	[0.0539,0.2917]	[0.1104,0.0594]	[1.4059,0.6726]	[0.0223,0.0268]	[0.0114,0.1397]	[0.1104,0.6752]
	p	0.3011	**0.0019**	0.4623	**0.0145**	0.5249	0.8285
G3	**COPD**	[0.0666,0.1236]	[0.0647,0.0222]	[1.9262,2.0573]	[0.0016,0.0135]	[0.1328,1.4760]	[0.0434,1.1864]
	Other except Cancer	[0.0539,0.3240]	[0.1156,0.1070]	[1.4059,0.8429]	[0.0221,0.0263]	[0.1252,0.7602]	[0.1104,0.8362]
	p	0.3277	**0.0048**	0.52870	**0.0360**	0.7756	0.6416
G4	**COPD**	0.0999±0.0503	[0.0647,0.0222]	[1.9262,2.0573]	[0.0016,0.0135]	[0.1328,1.4760]	[0.0434,1.1864]
	ARD	0.1794±0.0695	[0.1278,0.2504]	[1.3399,12410]	[0.0222,0.0204]	[0.0643,0.1395]	[0.0435,0.4292]
	p	0.3763	**0.0260**	0.3939	**0.0411**	0.8301	0.6576
G5	**COPD**	[0.0666,0.1236]	[0.0647,0.0222]	[1.9262,2.0573]	[0.0016,0.0135]	[0.1328,1.4760]	[0.0434,1.1864]
	Non-COPD	0.1724±0.1265	[0.0979,0.0471]	[1.5918,1.8025]	[0.0178,0.0319]	0.9206±0.8519	[0.5195,1.5]
	p	0.5341	**0.0238**	1	0.3686	0.8894	0.7668
G6	**COPD**	0.0999±0.0503	0.0626±0.0045	2.1500±0.5503	0.0097±0.0067	0.7897±0.4849	-0.0127±0.3115
	Cancer	0.0919±0.0640	0.0894±0.0054	1.3770±0.3970	0.0540±0.0083	0.0058±0.0057	-0.3558±0.6439
	p	0.9364	**0.0199**	0.1478	**0.0139**	0.3170	0.3614

[2] S. B. *et al.*, "Low physical functioning and impaired performance of activities of daily life in COVID-19 patients who survived hospitalisation," *European Respiratory Journal*, vol. 56, no. 4, 2020.

[3] World Health Organisation, "WHO Coronavirus (COVID-19) Dashboard," https://covid19.who.int/, 2021.

[4] European Commission, "Market study on telemedicine," European Union, Tech. Rep., 2018.

[5] Audit Scotland, "A review of Telehealth in Scotland," Audit Scotland, Tech. Rep., 2011.

[6] H. Pinnock *et al.*, "Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial," *BMJ*, vol. 347, 2013.

[7] Mingyu You *et al.*, "Automatic cough detection from realistic audio recordings using C-BiLSTM with boundary regression," *Biomedical Signal Processing and Control*, vol. 72, p. 103304, 2022.

[8] A. Tena, F. Clarià, and F. Solsona, "Automated detection of COVID-19 cough," *Biomedical Signal Processing and Control*, vol. 71, p. 103175, 2022.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[10] P. Amado-Caballero *et al.*, "Insight into ADHD diagnosis with deep learning on actimetry: Quantitative interpretation of occlusion maps in age and gender subgroups," *Artificial Intelligence in Medicine*, vol. 143, p. 102630, 2023.

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.