

Audio-Based Detection of COPD Exacerbations via Two-Stage Cough Detection and Classification

Marcin Kolakowski

0000-0003-0068-8784

Warsaw University of Technology

00-665, Warsaw, Poland

Email: marcin.kolakowski@pw.edu.pl

Tiberiu Seceleanu

0000-0003-1996-1234

Mälardalen University

721 23 Västerås, Sweden

Email: tiberiu.seceleanu@mdu.se

Irina Mocanu

0000-0001-5176-9344

CITST; National University of Science
and Technology POLITEHNICA Bucharest

060042, Bucharest, Romania

Email: irina.mocanu@upb.ro

Abstract—Chronic obstructive pulmonary disease (COPD) is a chronic respiratory condition characterized by frequent exacerbation episodes. The paper presents an approach to detecting COPD exacerbations based on audio recordings performed by patients. In the proposed algorithm, the audio signal is processed in two stages. First, it is converted to a mel spectrogram and analyzed by a neural network to extract cough segments. Next, the detected coughs are passed to a cough classification model that determines whether the coughs are induced by a COPD exacerbation. The proposed approach was cross-validated using two datasets for cough detection and classification, and further evaluated on a synthetic dataset of recorded user narratives. The results show that effective cough detection is feasible, achieving an AUC of 0.989 and 0.967 on the synthetic dataset. COPD cough classification remains more challenging, with an AUC of 0.794, particularly due to the limited size of publicly available COPD-focused datasets.

I. INTRODUCTION

CHRONIC obstructive pulmonary disease (COPD) is a chronic lung condition characterized by burdensome respiratory symptoms, including shortness of breath, cough, and increased mucus production. The disease is marked by frequent exacerbations during which the symptoms intensify, which often lead to hospitalization. COPD is prevalent worldwide and is a third leading cause of death following cardiovascular diseases [1].

Effective COPD management is a challenging task [2] because it requires careful monitoring of symptoms to enable early detection of exacerbations and timely interventions, thereby reducing hospital admissions and improving long-term outcomes. The use of modern data-driven ICT (Information and Communication Technologies) solutions can be especially helpful in both COPD patient monitoring and in reducing the strain on healthcare systems [3].

This work was performed within the framework of the THCS OREO project, Advancing personalized prevention strategies for chronic obstructive pulmonary disease and associated comorbidities that utilize existing IT technologies and novel AI-based tools. In Poland, this work was funded by the Polish National Centre for Research and Development, grant number THCS/II/100/OREO/2025. In Sweden, the work was funded by FORTE, Sweden with Project ID (diary number): 2024-01599. In Romania, this work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI - UEFISCDI, project number 97/2025 COFUND-THCS-OREO, within PNCDI IV.

In clinical settings, the COPD exacerbations and other lung conditions can be accurately diagnosed based on lung imaging [4], [5]. Such examinations, however, are costly and cannot be performed daily in a home environment. In early home-based COPD treatment support platforms, COPD exacerbations were usually detected based on the outcomes of digital questionnaires [6] or by using mobile applications for symptom tracking [7]. The more recent trend is to employ AI-based (Artificial Intelligence) predictive algorithms detecting exacerbation based on multimodal data [8].

The proposed algorithms utilize a wide range of data types, including physiological measurements, questionnaires, audio recordings, and environmental parameters and air pollution measurements. Of the above, the audio recordings show much promise for providing accessible, cost-effective COPD management platforms, as they can be easily collected with commonly owned smartphones and do not require purchasing additional devices.

In the literature, several algorithms for audio-based detection of COPD exacerbations have been presented. The proposed solutions utilize different signal types, recording techniques, and processing algorithms.

The most common type of audio recording used for COPD evaluation is respiratory sound recording. Coughs, which are the most audible symptom, are the most popular [9], [10] being closely followed by breathing sounds [9], [11]. The other, less commonly used, types of signals are human speech and vowel sounds. Their collection usually requires reading a fixed sequence or holding a single vowel for several seconds [12], which is monotonous and may elicit the patient's resistance when performed frequently.

The audio can be recorded using several devices. The most prevalent, due to their accessibility, are smartphones [9], [13]. Although they can easily capture cough events and human speech, they usually struggle to detect low-volume breathing sounds. To record breathing, studies suggest using either digital stethoscopes [11] or custom devices [14]. Those are typically hardly accessible and often require recording respiratory sounds in specific areas (e.g., at the suprasternal notch [15]). Given the above, recording breathing sounds might not be a viable approach for at-home COPD management.

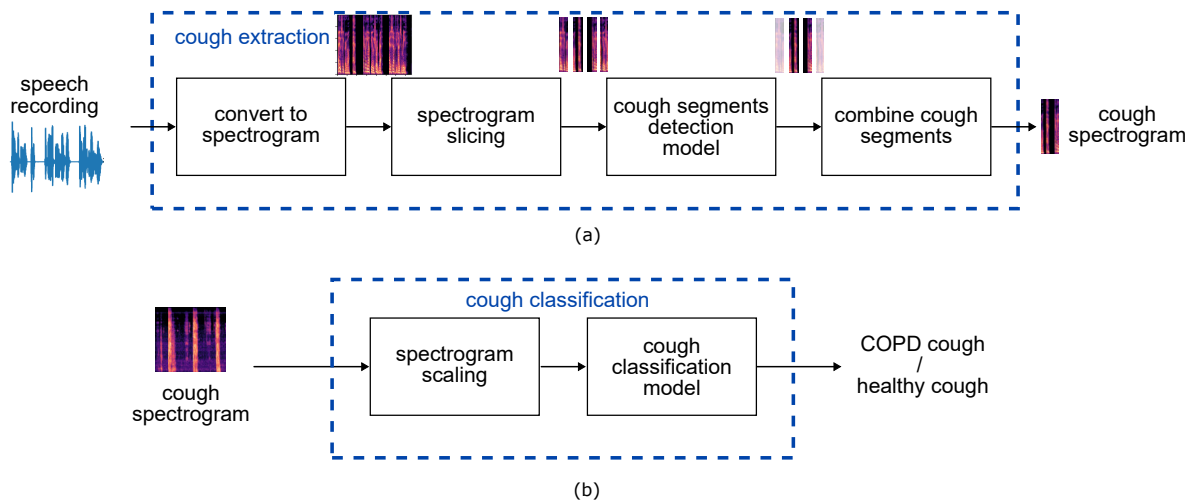


Fig. 1. The proposed COPD exacerbation detection algorithm workflow: a) cough segments extraction, b) cough type classification

From a signal processing perspective, the presented works use different approaches to feature extraction. Some studies use specially engineered signal features [14][8], enabling their easy and more meaningful integration with other multimodal data. They can also be efficiently used with smaller datasets. The neural-network-based approaches [11][10] enable the capture of underlying dependencies; however, larger architectures may be challenging to train due to a lack of large COPD-targeted datasets.

In this paper, we present an approach to detecting COPD exacerbations based on patients' sound recordings. The concept assumes recording audio for extended periods and extracting cough segments using a CNN-based algorithm. The extracted audio segments are then processed by a separate network to determine whether the recorded cough is healthy or may be caused by a current or imminent COPD exacerbation. The proposed approach was validated on two publicly available datasets for cough detection and COPD cough classification, and a custom, synthetically generated dataset of recorded users narratives.

The rest of the paper is organized as follows. The proposed approach is outlined in Section II. The used datasets and the data augmentation procedures are described in Section III. The descriptions of the employed models and experimental results are presented in Section IV. Section V concludes the paper.

II. COPD EXACERBATION DETECTION

In the proposed concept, the COPD exacerbations are detected by analyzing cough recordings collected by the patient. Contrary to several solutions described in the literature [9], [12], the concept does not require the patient to record the coughs or fixed sound sequences specifically. In the proposed method, recording segments containing cough sounds are automatically extracted and processed to classify the cough as healthy or COPD-induced. The workflow of the proposed algorithm is presented in Fig. 1.

The first step of the exacerbation detection process is cough extraction. To better capture the spectral features of the recorded sounds, the audio signal is converted to a spectrogram. In the proposed approach, we use a mel spectrogram, as mel-frequency representations have been shown to enable more effective audio classification [12]. To emphasize smaller values, the spectrogram is converted to the decibel scale.

Next, the spectrogram is sliced into 0.6 s segments of the audio signal. According to studies, such time is sufficient to capture all cough-related sound artifacts [16]. The spectrogram segments are then passed to the Cough Segments Detection model, which classifies them as either 'cough' or 'no cough'. Finally, the detected segments are combined to produce a spectrogram of the entire coughing sequence.

The obtained mel spectrogram of the coughing sequence is resampled and passed to the cough classification model. The model's output indicates whether the recorded cough is healthy or may be associated with a COPD exacerbation.

The proposed approach can be implemented both on the edge and in a centralized manner. Edge implementation would eliminate the need to transmit complete sound recordings containing potentially sensitive information. Moreover, it would not require a stable Internet connection, which, for many users, especially those living in remote rural areas, can be excluding. However, due to constrained resources and energy-use considerations, it may significantly limit the use of more advanced and larger models.

III. DATASET

Despite COPD being a disease that is prevalent worldwide, the number of publicly available audio recording collections including samples from COPD patients is very low. The majority of open audio data on respiratory diseases and cough detection were curated during the global pandemic and primarily concern COVID-19 patients. It raises a significant challenge in developing AI-based solutions for COPD prediction and

evaluation. Most published works rely on moderately sized in-house datasets collected under clinical conditions, which are often not publicly available due to internal regulations and privacy concerns [9]. Therefore, training the proposed method on a single consistent dataset is currently not feasible. In this work, we used two separate datasets to train the cough detection and cough classification models. We also employed a synthetically generated dataset to further validate the cough detection algorithm in a patient-narrative scenario.

A. Cough detection dataset

For the cough-detection algorithm training, we used recordings from the Multimodal Dataset for Automatic Edge-AI Cough Detection [14]. The publicly available collection includes inertial signals and sound recordings from 15 subjects, collected with a chest-mounted device.

The dataset comprises recordings of cough and similar sounds: deep breathing, laughing, and throat clearing. The recordings were simultaneously performed using two microphones (facing towards and outwards the body) under controlled conditions. During the recordings, the users were asked to cough while sitting or moving. To simulate real conditions, typical environmental sounds (traffic, conversation, and music) were played in the background.

The dataset includes 2686 recordings (672 containing coughs) from 1433 captures (each scenario was recorded with two microphones). For training purposes, we have used only the recordings from the body-facing microphone. The recordings are approximately 10 seconds long. They were partitioned into 0.6 s non-overlapping segments, which were converted into mel spectrograms, yielding a dataset of 21297 samples (4294 coughs).

B. Cough detection synthetic test dataset

In addition to cross-validating the cough segment detection algorithm, we decided to test its cough-extraction capabilities on a synthetic dataset comprising 30 COPD patient narratives. The narratives were two-sentence statements concerning the user's current health status and sleep conditions the previous night. The statement contents were artificially generated using ChatGPT (GPT-5.3) [17], which was prompted to generate

two-sentence answers interrupted by tagged coughs, as if they were responses from a patient experiencing a COPD exacerbation. The prompt used for that purpose is presented in Example 3.1. The obtained narratives were then a basis for artificial voice generation using ElevenLabs' Eleven V3 model [18]. Eleven V3, as of the time of writing, is the most recent ElevenLabs Text-to-Speech model, allowing the generation of expressive statements with additional sound effects by injecting specific tags into the text prompt. The relevant effects available include coughs and wheezing. For this test, we have generated 30 statements using 5 voices. Each narrative was generated twice, yielding 60 samples. Three exemplary user narratives are presented in Example 3.2.

C. Cough classification dataset

The cough classification algorithm was trained on the Cough-COPD (CC) Dataset [19]. The dataset includes 221 cough samples from 78 COPD patients (57 male, 21 female) and 632 samples from 221 healthy individuals (83 male, 138 female). The study participants were of Asian descent and aged 25 to 75 years [13].

The Cough-COPD dataset is small. Therefore, to improve model generalization, we performed data augmentation using the following techniques:

- volume adjustment - by multiplying the signal by a random constant from 0.5 to 1.5,
- time stretch - speeding up / slowing down the audio by a random factor from 0.8 to 1.2 range,
- time shift - moving the signal along the time axis by a maximum of one second,
- noise addition - addition of Gaussian noise to the signal,

Although pitch change is a common augmentation technique, it was not applied. Pitch is one of the critical indicators used by medical professionals, and its augmentation may significantly impact the model's diagnostic capabilities.

For each training example, a random subset of the above techniques was applied. To improve data variability, the subsets were selected at random. Each of the training examples was augmented three times, enlarging the dataset to a total of 3412 cough samples (884 COPD-induced cough).

Example 3.1: A ChatGPT prompt used to generate user narratives with coughs

```
Generate ten sets of naturally sounding
answers by people suffering from COPD
exacerbations to those three questions:
1. How did you sleep last night?
2. How have you been since we last talked
3. Did you feel short of breath recently?
```

In every answer, insert at least one or two cough tags (as [cough]), which would simulate natural coughs.
Format the output as CSV.

Example 3.2: Three exemplary user statements converted into speech recordings using ElevenLabs V3.

```
1. I didn't sleep very well... [cough]
kept waking up feeling short of breath.
Had to sit up a few times to breathe.

2. Not doing too well [cough cough].
I get tired really quickly now.

3. Yeah, even just moving around a bit
makes me breathless [cough]. It's
definitely worse than usual.
```

IV. EXPERIMENTAL VERIFICATION

A. Models and training

Both cough segment classification and cough classification tasks were solved using convolutional neural networks with attention mechanisms. The general architecture of the used networks is presented in Fig. 2.

The neural networks used in this study comprise a CNN-based feature extractor (detailed structure in Fig. 2.b) and a classification head composed of dense layers. The proposed architectures are intentionally simple to enable deployment on more constrained devices and to enable edge processing of the recorded audio signals.

The convolutional block includes a Squeeze-and-Excitation (SE) block providing the attention mechanism. After performing the convolution, the resulting features are passed to SE, where the most significant channels are determined and amplified in the resulting feature set via direct multiplication. The features multiplication is followed by optional Batch Normalization, ReLU activation, and MaxPooling.

The classification head consists of Dense blocks (dense layers with ReLU activations and additional dropout to improve generalization). The output is passed through the sigmoidal activation function. All neural network models were implemented with PyTorch.

In the experimental verification of the concept, we have performed a parameter search over the space presented in Tab. I. The most promising architectures were validated using a 5-fold cross-validation scheme. The parameters for which the most favorable results were achieved are presented in Tab. II.

For cough classification, we additionally implemented the approach presented in [12], in which exacerbations are detected using hand-crafted audio features. The features used for that purpose include 30 first Mel-Frequency Cepstral

TABLE I
NEURAL NETWORK ARCHITECTURE PARAMETERS SEARCH SPACE

Parameter	Search Space
Number of CNN blocks (N)	1, 2, 3
CNN layer size	8, 16, 32, 64, 128, 256
Number of Dense blocks (M)	1, 2, 3
CNN layer size	8, 16, 32, 64
Batch Normalization	yes / no
Dropout	0.2, 0.5

TABLE II
PARAMETERS OF ARCHITECTURES OBTAINED AFTER THE PARAMETER SEARCH

Parameter	Cough Detector	Cough Classifier
Input shape	128×110	128×512
CNN layer sizes	[256, 512, 256]	[128, 256, 128]
Dense layer sizes	[16, 16]	[64, 16]
BatchNormalization	no	no
Dropout	0.2	0.2
Batch size	512	128
Optimizer	Adam	
Learning rate	1×10^{-4}	
Epochs	60	120

Coefficients, pitch, local jitter, shimmer, Formant 1, 2, and 3 frequencies, Loudness, and Harmonics-to-noise ratio. For all parameters, average and standard deviation were calculated across the window, yielding a total of 76 parameters. The audio features were computed using the librosa [20] and openSMILE 3.0 [21] Python libraries. The cough classification was performed using Logistic Regression (LR) and a Random Forest (RF) Classifier (with 200 trees).

B. Results discussion

The performance metrics for both of the models are presented in Table III. The receiver operating characteristic (ROC) curves for cough segment detection and cough classification are presented in Fig. 3 and Fig. 4, respectively. In the analyses performed, positive labels were assigned to signal segments containing coughs and COPD-induced cough samples, respectively.

During cross-validation, the cough segment detection model achieved an average accuracy of around 0.966 (0.941 balanced), which is similar to results reported in other works [10]. It shows that effective cough detection and extraction are certainly possible. When validated on the synthetic dataset,

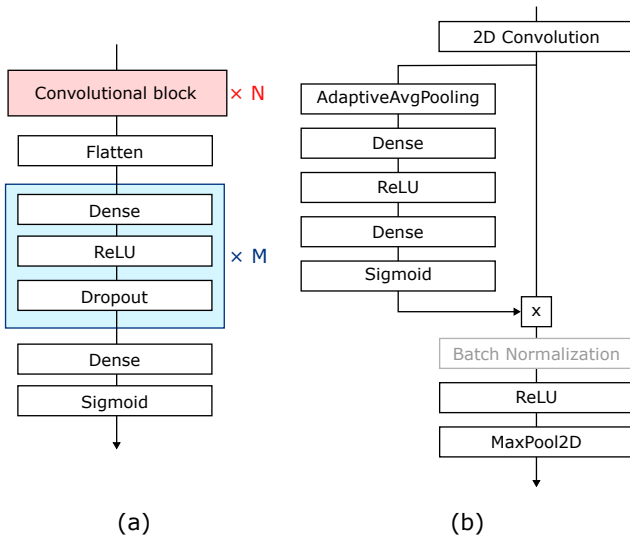


Fig. 2. Neural network architecture, a) the general network architecture b) the convolutional block

TABLE III
MODEL PERFORMANCE METRICS

Metric	CSD (cv)	CSD (st)	CC (cv)	CC (LR)	CC (RF)
Accuracy	0.966	0.914	0.757	0.804	0.846
Bal. Acc.	0.941	0.769	0.698	0.708	0.719
Precision	0.928	0.918	0.544	0.662	0.911
Recall	0.900	0.549	0.575	0.507	0.453
F1	0.914	0.687	0.548	0.572	0.599
Specificity	0.982	0.990	0.821	0.908	0.984
AUC	0.989	0.967	0.794	0.708	0.719

CSD - Cough Segments Detection; CC - cough classification; cv - cross-validation result; st - validation on synthetic data, LR - Logistic Regression; RF - Random Forest

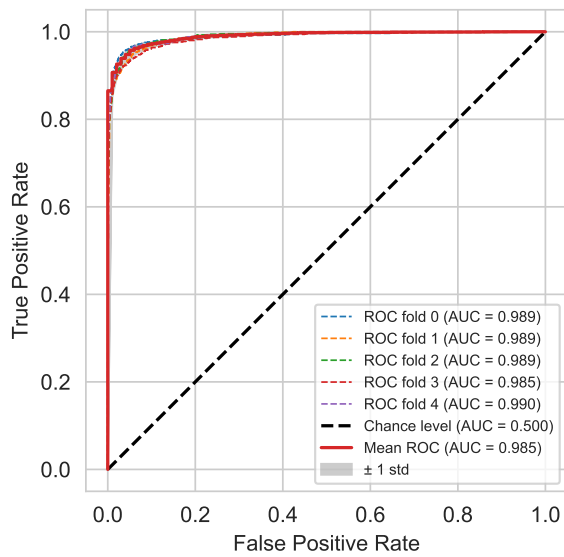


Fig. 3. Receiver operating characteristic curve for the cough segment detector

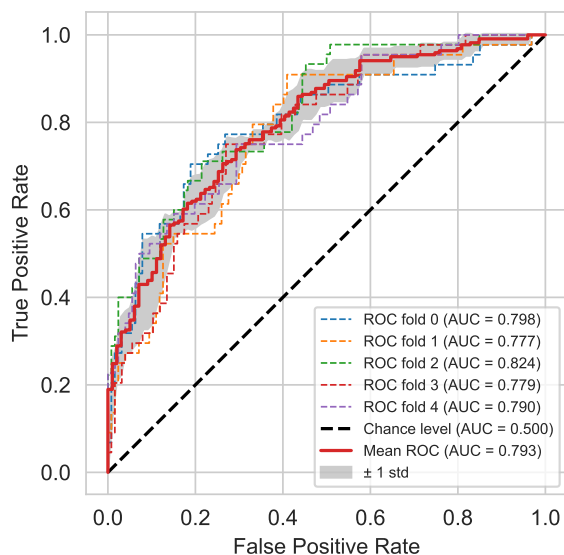


Fig. 4. Receiver operating characteristic curve for the cough classifier

the accuracy is lower because the model struggles to detect artificially generated coughs. One reason for this lower performance may be that, in the training dataset, the coughs were particularly forced by the person, making them more pronounced than usual. The synthetic coughs generated by the model were audibly different. They resemble dry coughs that appear when speaking for a longer time, rather than illness-induced coughs. Generating clinically relevant samples would require using dedicated models.

The results obtained for the cough classifier are on par with those reported in the original work using the Cough COPD dataset (0.635 recall, achieved with a much more advanced

model in [13]). The average AUC is approximately 0.8, which can be considered a good value given the model's simplicity and the dataset's size. The results for LR and RF were slightly better in terms of accuracy, but the AUC was significantly lower. It must be kept in mind that the COPD CC dataset is relatively small and imbalanced, making it difficult to draw definitive conclusions about the method's accuracy.

For all tested methods, recall was relatively low, which points future research towards identifying more discriminating features. Analyzing cough spectrograms jointly with other data types may yield more accurate results. Given that the proposed approach uses voice recordings, it opens a possibility to extract relevant information from their contents. Asking the patient targeted questions and mining the answers for information on COPD exacerbation symptoms may provide additional features for more effective prediction. Such an approach has already been successfully tested in other disease treatments [22].

V. CONCLUSIONS

The paper presents a cough-analysis-based approach to COPD exacerbation detection. In the proposed method, the audio recordings made by the patients are processed to extract cough segments and then classify them as healthy or COPD-induced.

To verify the concept's feasibility, we employed CNNs with an attention mechanism. The results show that although cough segments can be easily localized and extracted, classifying COPD coughs remains a much greater challenge, largely due to a lack of large, balanced datasets.

The proposed concept shows much promise for future development. A possible direction is to collect multimodal COPD-directed datasets, in which, alongside recording audio signals, patients perform physiological measurements and report perceived symptoms.

ACKNOWLEDGMENT

During the preparation of this work, the authors used Chat GPT (GPT-5.3) for generating the users' narratives and Eleven V3 for converting them into speech. The authors take full responsibility for the content of this publication.

REFERENCES

- [1] A. Lindberg et al., "Large underreporting of COPD as cause of death-results from a population-based cohort study," *Respiratory Medicine*, vol. 186, p. 106518, Sep. 2021, <http://dx.doi.org/10.1016/j.rmed.2021.106518>.
- [2] A. Anzueto, H. Hoy, and C. Gainer, "Challenges and Opportunities for the Management of COPD: A Narrative Review of Patient and Health Care Professional Perspectives," *J COPD F*, vol. 12, no. 4, pp. 339–356, 2025, <http://dx.doi.org/10.15326/jcopdf.2024.0597>.
- [3] L. Beckman, C. D. Muylder, and O. Reinhold, "Effects and challenges of Governance, Lean Healthcare and Digital Transformation in Outpatient Management: An integrative review," *Annals of Computer Science and Information Systems*, vol. 45, Polish Information Processing Society, pp. 33–40, Oct. 15, 2025, <http://dx.doi.org/10.15439/2025f2505>.
- [4] D. Zhu et al., "Identification of frequent acute exacerbations phenotype in COPD patients based on imaging and clinical characteristics," *Respiratory Medicine*, vol. 209, p. 107150, Apr. 2023, <http://dx.doi.org/10.1016/j.rmed.2023.107150>.

- [5] G. S. Araújo, O. A. C. Cortes, O. Reinhold, and F. M. F. Lobato, "Hybrid Approaches for Pneumonia Detection in X-rays: Combining CNNs and ML Classifiers," *Annals of Computer Science and Information Systems*, vol. 43, Polish Information Processing Society, pp. 461–470, Oct. 15, 2025. <http://dx.doi.org/10.15439/2025f4310>.
- [6] University Hospitals of Leicester NHS Trust, "i-IMPACT | An online programme for patients with a long-term cardiac, respiratory or Covid condition." <https://www.i-impact.co.uk/> (accessed Jun. 25, 2026)
- [7] H. Davies et al., "myCOPD App for Managing Chronic Obstructive Pulmonary Disease: A NICE Medical Technology Guidance for a Digital Health Technology," *Appl Health Econ Health Policy*, vol. 21, no. 5, pp. 689–700, May 2023, <http://dx.doi.org/10.1007/s40258-023-00811-x>.
- [8] M. Atzeni, G. Cappon, J. K. Quint, F. Kelly, B. Barratt, and M. Vettoretti, "A machine learning framework for short-term prediction of chronic obstructive pulmonary disease exacerbations using personal air quality monitors and lifestyle data," *Sci Rep*, vol. 15, no. 1, Jan. 2025, <http://dx.doi.org/10.1038/s41598-024-85089-2>.
- [9] Y. Gong et al., "AI-driven smartphone screening for acute COPD exacerbations: enhancing health equity in developing regions," *npj Digit. Med.*, vol. 8, no. 1, Nov. 2025, <http://dx.doi.org/10.1038/s41746-025-02086-z>.
- [10] R. T. Bhowmik and S. P. Most, "A Personalized Respiratory Disease Exacerbation Prediction Technique Based on a Novel Spatio-Temporal Machine Learning Architecture and Local Environmental Sensor Networks," *Electronics*, vol. 11, no. 16, p. 2562, Aug. 2022, <http://dx.doi.org/10.3390/electronics11162562>.
- [11] C. R. Khandhan and G. M. Kaviya, "Deep Learning Model for Chronic Obstructive Pulmonary Disease through Breathing Sound," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, pp. 1514–1521, Jul. 06, 2023. <http://dx.doi.org/10.1109/icesc57686.2023.10193372>.
- [12] Y. Yan, L. van Bommel, F. M. E. Franssen, S. O. Simons, and V. Urovi, "Developing a multi-feature fusion model for exacerbation classification in asthma and COPD," *Computer Methods and Programs in Biomedicine*, vol. 268, p. 108796, Aug. 2025, <http://dx.doi.org/10.1016/j.cmpb.2025.108796>.
- [13] W. Sun, G. Wu, M. Ming, J. Zhang, C. Shi, and L. Qin, "Self-supervised learning for intelligent disease diagnosis using audio signals: beyond copd to a spectrum of diseases," *Appl Intell*, vol. 55, no. 7, Feb. 2025, <http://dx.doi.org/10.1007/s10489-024-06028-2>.
- [14] L. Orlandic, J. Thevenot, T. Teijeiro, and D. Atienza, "A Multimodal Dataset for Automatic Edge-AI Cough Detection," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp. 1–7, Jul. 24, 2023. <http://dx.doi.org/10.1109/embc40787.2023.10340413>.
- [15] M. A. Fernandez-Granero, D. Sanchez-Morillo, and A. Leon-Jimenez, "An artificial intelligence approach to early predict symptom-based exacerbations of COPD," *Biotechnology & Biotechnological Equipment*, vol. 32, no. 3, pp. 778–784, Feb. 2018, <http://dx.doi.org/10.1080/13102818.2018.1437568>.
- [16] A. B. Chang, "The physiology of cough," *Paediatric Respiratory Reviews*, vol. 7, no. 1, pp. 2–8, Mar. 2006, <http://dx.doi.org/10.1016/j.prrv.2005.11.009>.
- [17] OpenAI, ChatGPT (GPT-5.3), large language model. [Online]. Available: <https://chat.openai.com>. (accessed Apr. 7, 2026)
- [18] ElevenLabs, Eleven v3 — Most Expressive AI Voice Model." <https://elevenlabs.io/v3> (accessed Jun. 25, 2026)
- [19] University of Science and Technology of China and Hangzhou slant-health Co, Ltd, "Cough-COPD(CC) Dataset". Zenodo, Feb. 17, 2025. <http://dx.doi.org/10.5281/zenodo.14881818>.
- [20] B. McFee et al., "librosa: Audio and Music Signal Analysis in Python," *Proceedings of the Python in Science Conference. SciPy*, pp. 18–24, 2015. <http://dx.doi.org/10.25080/majora-7b98e3ed-003>.
- [21] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile," *Proceedings of the 18th ACM international conference on Multimedia. ACM*, pp. 1459–1462, Oct. 25, 2010. <http://dx.doi.org/10.1145/1873951.1874246>.
- [22] K. Matsumoto, M. Yoshida, and C. Karino, "Mining Patient Narratives to Analyze Lifestyle–Blood Glucose Relationships: An LLM-Based Text Mining Framework," *J*, vol. 9, no. 2, p. 14, May 2026. <http://dx.doi.org/10.3390/j9020014>.