Discover Artificial Intelligence

Case Study

Classification for long-term monitoring of cough

Albertus C. den Brinker[1] · Ronald Rietman[2] · Okke Ouweltjes[3] · Matthijs van Marion[4] · Susannah Thackray-Nocera[5] · Michael G. Crooks[5] · Alyn H. Morice[5]

Received: 18 October 2024 / Accepted: 10 April 2025
Published online: 12 May 2025

Abstract
For management of chronic respiratory diseases, unobtrusive longitudinal monitoring of cough has been proposed. Such a monitoring system was developed using a classifier trained on an initial observation period. After this initial period, a personalized system is available being optimized for the patient and the particular acoustic environment. Long-term deployment of the system requires that the extracted features and learned model characterizing the coughs (and its environment) are time-invariant. This is studied by an example using annotation of two largely different epochs. The results suggest that time-invariance of the cough sound is sufficiently guaranteed for practical deployment, but that changing acoustic environmental conditions may be a factor to reckon with. Cues for detecting changing situations are discussed.

Keywords Telehealth · COPD · Cough · Unobtrusive monitoring · Long-term acoustic monitoring · Classification · Stationarity

1 Introduction

Cough is a symptom characteristic of many respiratory diseases. It is also a phenomenon that can be captured in an unobtrusive way by using a microphone system. It could thus become an effective ingredient of telehealth solutions [1–4]. Since day- and night-time cough counts are correlated [5], a truly hassle-free, unobtrusive monitoring system can be created as a stationary device situated in the sleeping quarters of the patient. The night-time monitoring has the additional advantage that in almost all cases it presents a less adverse acoustic environment for signal extraction. A privacy-preserving system is achievable by extracting features in the home of the patient and transferring only these from the patient's home. Transferring short snippets for a limited number of acoustic events next to the features allows to check the acoustic conditions and to develop personalised classifiers (based on annotation of the snippets) without being able to listen in to any conversation. The cough classifier can thus be optimized using data from its initial period; a generic classifier would perform considerably worse due to the fact that it is not trained for the patient's cough sounds

✉ Albertus C. den Brinker, acdenbrinker@kpnmail.nl; Ronald Rietman, ronald.rietman@philips.com; Okke Ouweltjes, okke.ouweltjes@philips.com; Matthijs van Marion, matthijs.van.marion@philips.com; Susannah Thackray-Nocera, susannah.thackray-nocera@nhs.net; Michael G. Crooks, michael.crooks@nhs.net; Alyn H. Morice, a.h.morice@hull.ac.uk | [1]Independent Researcher, Formerly with Philips Research, Philips, High Tech Campus, NL-5656AE Eindhoven, The Netherlands. [2]Innovation Engineering, Data Science and AI, Philips, High Tech Campus, NL-5656AE Eindhoven, The Netherlands. [3]Digital Standardization and Licensing Research, Philips, High Tech Campus, NL-5656AE Eindhoven, The Netherlands. [4]HPM R&D, SW Platform Applications, Philips, High Tech Campus, NL-5656AE Eindhoven, The Netherlands. [5]Department of Academic Respiratory Medicine, Centre for Cardiovascular and Metabolic Research, Hull York Medical School, Castle Hill Hospital, Hull HU16 5JQ, UK.

nor for the pertinent acoustic environment. We note that most common cough monitors do not deploy personalisation, see e.g., [6, 7].

For a system that is trained from data in an initial observation period, the prerequisite is that the character of the cough remains stable over time. However, it is presumed that cough sounds change with age as the vocal and respiratory system also change over time. This would argue against an assumption of time-invariant cough sounds, though maybe not at short time-scales. In previous studies, we did not observe clear issues with the assumption of stationarity of the cough sounds over a prolonged period (several months), but in one of the patients in a more recent study we observed a long-term decreasing trend in the cough data, see Fig. 1. An obvious explanation for such behaviour would be a change in character of the cough sound such that the classifier trained on data of the start of the trial does not adequately detect coughs in later days. This hypothesis is considered in this paper. Since the effort of annotation is relatively large, the verification of the stationarity was done for one patient only.

The relevance of the paper is twofold and relate to findings in practical usage of cough monitoring over long periods of time. Firstly, it is verified that the long-term trend (over multiple months) of a specific patient is not caused by changes in cough sound over time, thus underpinning the assumption of long-term stationarity of the cough sound character. With results of previous trials in mind, it is assumed that this finding carries over to other patients. Secondly, it was found that variations in the environment occur that are more likely to disturb the results. This is related to the Out-of-Distribution (OOD) problem: feature combinations insufficiently present in the training are often not learned. In our case, it was observed that depending on the non-cough data, substantially different separability is obtained during its application. It is shown that several indicators can be used to identify such a situation and thus to take appropriate action, e.g., by ignoring the outcome at these days or retraining the classifier taking the new environmental sounds into account. All of these can be done in the context of privacy-preserving set-up of the cough monitoring system. Though solutions for the OOD problem have emerged also for cough classification they are not yet mature [8], and tuning them for personalised classifiers may be problematic in large scale deployment. Therefore simple indicators, potentially outside of the feature set, to monitor whether deployment and training conditions match remain valuable.

A larger analysis of the full trial data addressing the primary study aim is ongoing: overviews of cough classifier performance, cough trending and its relation to medical condition are not part of this paper.

2 Data collection and analysis

2.1 Clinical trial

A double blind clinical trial was started to validate an earlier designed COPD exacerbation alert mechanism [9, 10]. It aims at verification of the hypothesis that tracking the cough count by an automated system provides medically relevant information about the respiratory status of a patient. Forty mild to severe COPD patients were recruited and monitored for at most 6 months. The monitor is active 12 h per day from 9 pm to 9 am implying that it presumably operates not only during sleeping hours but also at times when the patient goes to bed or rises, i.e., being active in the bedroom.

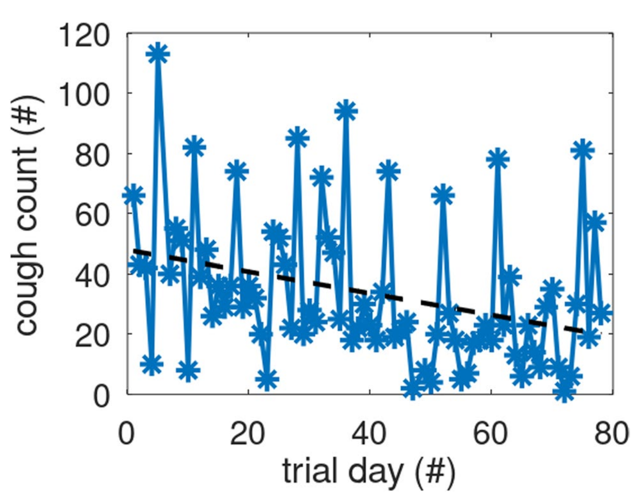

Fig. 1 Cough count trending. Results from classifier using samples from start days for training. The line represents the trend obtain by linear regression

The present paper reports on one patient where it was desired to test whether results were in some way affected by non-stationarity of the cough sound as this could be considered a plausible explanation for the observed gradually decreasing amount of coughs over the monitoring period shown in Fig. 1. We note that due to the privacy preserving methodology, it is not possible to create a ground truth of the number of coughs in any of the monitored days. The study was reviewed and approved by the North East-York Research Ethics Committee (REC Ref.: 21/YH/0203), the United Kingdom Health Research Authority and the Internal Committee Biomedical Experiments of Philips Research. The REC is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK. Informed consent was obtained from all individual participants included in the study.

2.2 Monitor

A research prototype was created consisting of a single board computer with a USB measurement microphone and a cellular dongle, see Fig. 2. Feature extraction and type of cough classifier were based on earlier trials [10–12]. The whole set-up of this system was designed in a first trial where complete recordings were made in the patients' home. That trial provided the fundament for designing a system in which privacy-preservation was key. The recap of the processing system is as follows. A first mechanism detects an audio transition with a method based on linear prediction analysis. When a transition is detected, a fine-grain transition position search is executed and spectral parameters (MFFC-like) around the transition are calculated as well as several time-domain ones (energy before and after transition and density of acoustic events). All together, it means that a time stamp is generated and a very limited number of features is available for classification. From the very start, annotators noted clearly audible differences between coughs of different COPD patients which, consequently, should somehow be reflected in the MFCCs, making personalisation of the classifier relevant. This is further supported by the results from the previous trial. While a leave-one-patient-out validation resulted in AUCs ranging from 0.878 to 0.957 (for a subset of 7 patients [12]), the personalised approach gave AUCs in the range 0.93 to 0.99 (all 25 patients [13]).

Since furthermore new acoustic environments are expected with each new patient, some access to the prevailing acoustic environment was deemed essential. This is achieved by capturing an audio excerpt of 1 s length for a limited number of timestamps. These excerpts prevent listening in to conversations, yet enable checking the audio quality of incoming sound, detection of issues in the acoustic environment (e.g., a ticking clock next to cough monitor) and personalisation of the classifier.

All together, the processing set-up of the monitor is classical in its nature, driven by the axioma of collecting as little data as possible. Nevertheless, the earlier investigations indicated that such a trained system is able to provide an early warning for deterioration of respiratory condition by tracking the number of detected coughs [9, 10].

2.3 Annotation

For annotation, an audio-visual interface was used. It presents the sound of a snippet over the speaker or headphone combined with a visual representation of the waveform and its spectrogram. The audio is essential to annotate if the features belong to a cough, and not to an environmental sound or a vocal sound of the patient which is not a cough (like: throat clearance, sigh, moan, sneeze, burp, speech, laughter). The signal waveform is provided because a typical cough consists of three phases: an explosive part, an intermediate stage and a voiced phase. Not only is visualization

Fig. 2 Bedside cough monitor. Left: adapter, mid: single-board computer and microphone, right: dongle

Table 1 Training data taken from start or end of monitoring period. Number of annotation days, total number of acoustic events, coughs and non-coughs in these days

	Days	Events	Coughs	Non-coughs
Start	7	6686	223	1000
End	9	1662	180	573

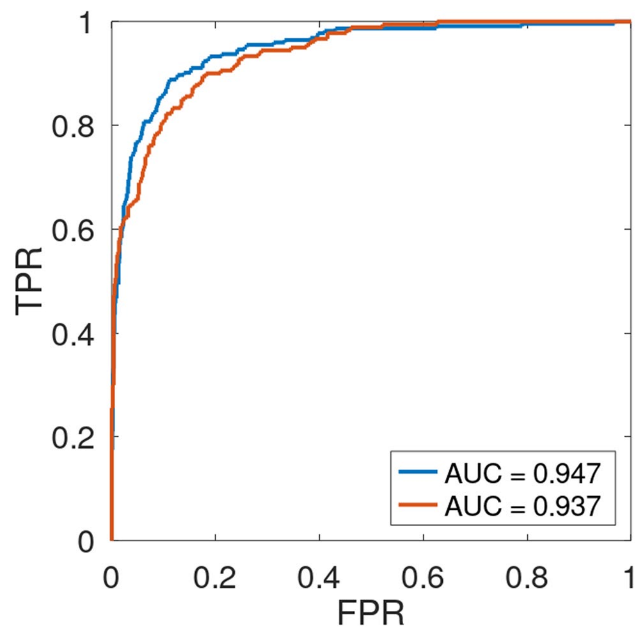

Fig. 3 ROC curves from 10-fold cross-validation. TPR: true positive rate, FPR: false positive rate

helpful because of this specific pattern, it is also instrumental to obtain a single identification of the cough over time and not multiple. The explosive phase of the cough is defined as the target as this is the acoustic response to the opening of the vocal chords after pressure build-up: a requirement for a cough by definition. It is also the most easily identified part of the cough: the intermediate and voiced phase are not always clearly present in both sound and visuals. To obtain a unique signature for a cough, the central position of the feature extraction is shown in the graph and the annotator checks if this corresponds to the explosive phase if the sound resembles a cough. The annotation for the current analysis were created by one person; for more details on the annotation, see [14].

2.4 Machine learning

Two models were trained using an extreme gradient boosted decision tree classifier (XGBoost) [9, 12] where default settings were used for model parameters and the number of boost rounds was manually optimized preventing performance drops with increased rounds. In order to study the effect of time dependence, the training was based on annotations from either the first or last days of the monitoring period. The interval between these two periods is 60 days.

From experience, it was known that having about 200 samples for both cough and non-cough categories provides a stable model. In Table 1, the data amount of training data is shown. The first annotation period provides more training data than the last 9 days. The number of acoustic events in the first monitoring days was clearly larger and this is reflected in the number of snippets and in the prevalence of coughs in the annotation. The difference between the number of acoustic events and the number of *coughs* plus *non-coughs* demonstrates that only part of the data of the annotation period can be actually used for training due to the fact that only for a limited amount of acoustic events the required sound snippets are available. Applying the trained models on the pertinent annotation days means that the classifier is dominantly running on features not contained in the training. The classifiers were evaluated using 10-fold cross validation with ROC curves in Fig. 3 and AUC above 0.93.

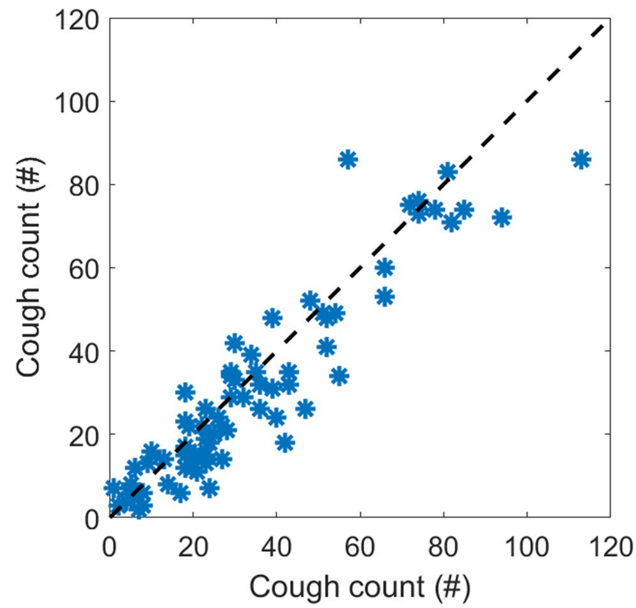

Fig. 4 Scatter plot of the raw cough counts of both classifiers based on annotations from start or end days of the trial. Each asterisk presents a day. Horizontally the detected counts are based on a classifier trained with annotations from sounds during start days, vertically from last days

Table 2 Pearson correlation (C) and cosine similarity (S) of the raw cough counts and cough trends (B-scale) of the two constructed classifiers

	C	S
Detected coughs	0.930	0.975
B-scale	0.930	0.994

3 Results

Correlation, similarity, scatter and trending plots of the data are considered. After the classification of all features, the detected coughs were counted per night. This is called the raw cough count. A non-linear mapping was applied (B-scale [9]) to facilitate the alert functionality as it accounts for the increasing spread with increased counts. The B-representation enables a more direct comparison of cough data among subjects having a totally different baseline, see [14]. In Fig. 4, the raw cough counts are compared. The dashed line gives the ideal situation of equal counts from both classifiers. The actual samples agree well with this line. This is quantitatively shown in Table 2, where the Pearson correlation coefficient C and the uncentered correlation coefficient (cosine similarity) S are provided. In Fig. 5, the trend graphs of the cough count (B-scale) are shown as obtained by both classifiers. All these results demonstrate a high degree of similarity. It highlights that the decreasing trend is not a result of a change in cough character over the monitoring period.

That both classifiers return more or less the same result does not necessarily mean that the underlying mechanisms work the same. To get more insight into the difference we compare the results of the classifiers (excluding training days) in the form of a confusion matrix shown in Table 3. The amount of data that both classifiers agree upon is 87%. We also observe that the majority of acoustic events (75%) is labelled as a non-cough by both classifiers.

These results are dependent on the threshold setting. To better understand what drives this difference and to be independent of the threshold setting we looked at the probabilities generated by the classifiers also including the training periods, where we note that the classifiers were trained on only a limited part of the data in either beginning or end days, i.e., only for those data with an associated audio snippet. Per monitoring session, we added probabilities assigned to acoustic events and made scatter plot Fig. 6. If the acoustic events were highly separable in coughs and non-coughs, meaning only probabilities close to 0 or 1, the plots should resemble the scatter plot in Fig. 4. For most days, both classifiers do generate a similar sum, yet there are 2 sessions clearly different. These outliers were considered in more detail.

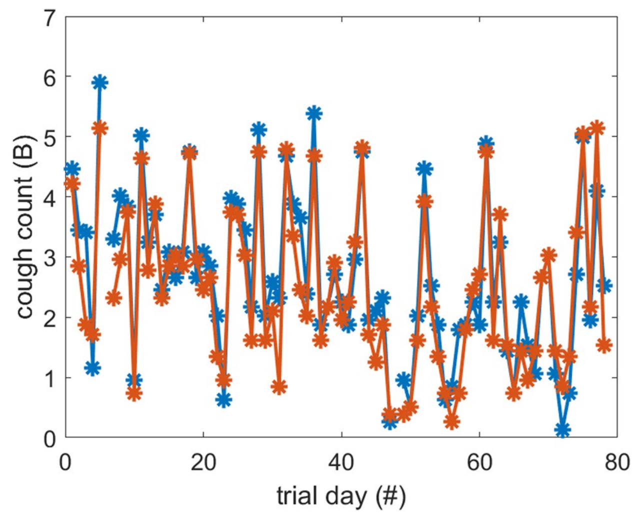

Fig. 5 Cough count trending on the B-scale. Results from classifier using start days (blue line) and end days (red line) annotation

Table 3 Confusion matrix showing number of events with identical and unequal classification, with threshold set to 0.9. Columns C_1 and N_1 contain the coughs and non-coughs of the classifier trained on the start days; rows C_2 and N_2 those of the classifier trained on data from end days

	$C_1 = 2314$	$N_1 = 9329$
$C_2 = 2064$	1441	623
$N_2 = 9579$	873	8706

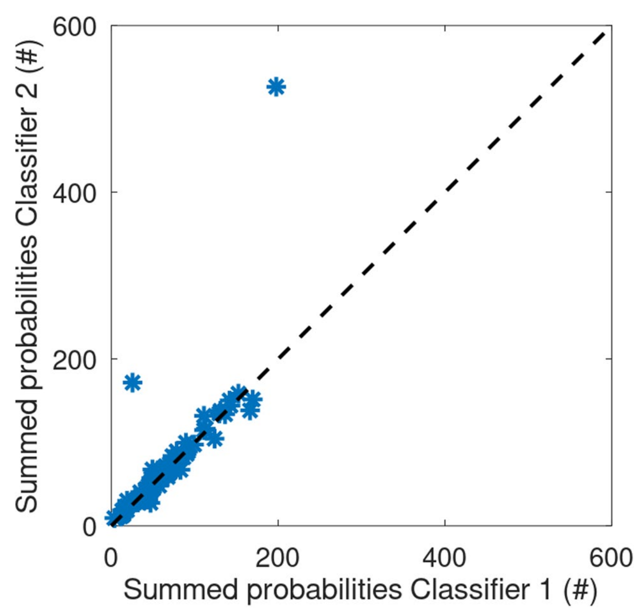

Fig. 6 Scatter plot of sum of non-categorical classifier output from both classifiers. Each asterisk presents a day. Horizontally the data are based on annotations from sounds during start days, vertically those from last days

Both outliers are large for the case of the classifier trained on the end annotation. It was also immediately clear that the number of acoustic events on the associated days is extremely large: these are 1485 and 4182 where the maximum for any of the other days is less than 560.

We report further on the largest discrepancy; the other one is similar in nature and cause. In Fig. 7, the assigned cough probabilities of session 5 are plotted in increasing order excluding all data that was used for training. We observe a clear difference in behaviour between the two classifiers. Whereas the blue curve shows a sharp transition, the red curve

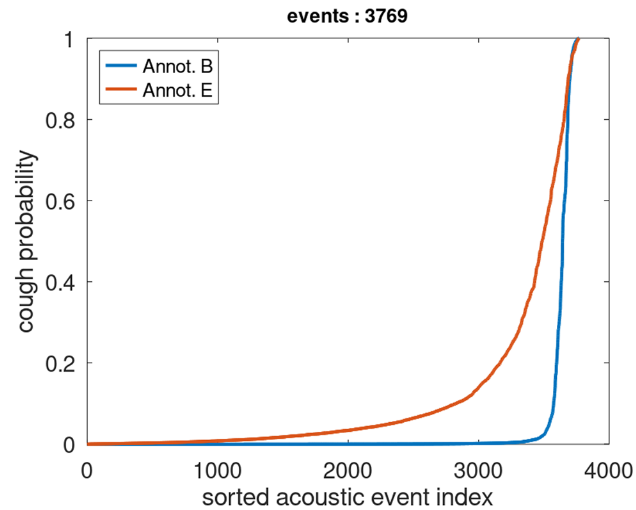

Fig. 7 Sorted classifier outputs (probabilities) for trial session 5. Results from classifier using start days (blue line) and end days (red line)

presents a much more gradual transition from low to high probabilities. It implies that the separation between the *cough* and *non-cough* events on this day is better captured in the first classifier.

The outliers are on two consecutive days: days 4 and 5 of the trial and part of the data of both days are used in the training set for the classifier based on the start annotation. The lack of the particular acoustic events occurring on days 4 and 5 in the training of the second classifier is reason for the poorer separability observed in Fig. 7 (red line). Inspection of the available snippets revealed that on these days the television was turned on. It is clear that not having these specific environmental sounds in the training set provides less well-trained classifiers and that the increased number of acoustic events in these sessions is indicative for the increased richness in the *non-cough* sounds.

4 Discussion

The results highlight that the classifier is a two-class system: the cough class as well as the environmental sounds need proper and sufficient samples in the training set. Unless one is prepared to use a one-class classifier (i.e., only characterizing the cough class), it means that a full picture of the environmental sounds is required. This is an issue as acoustic environments tend to change over time. In [13] it was argued that a high specificity should be preferred over a high sensitivity in view of the fact that the prevalence of cough events tends to be low relative to other acoustic events. Here we see a second reason why it makes sense to prioritize high specificity over high sensitivity: it makes the system less vulnerable to environmental sounds that were not present in the training set: the OOD issue. It is obvious that training including all potential adverse environmental sounds is impossible.

Fortunately, the results also point at possibilities for detecting potential incorrect or questionable classifier outputs. Two factors were explicitly covered: an unexpected large amount of acoustic events (larger than normal and deviant from that used in the training set) suggests that the acoustic environment has changed. Secondly, the classifier output can be monitored. Days with an atypical slow transition from high to low outputs (see Fig. 7) may need to be treated with more care. Retraining of the classifier is an option; a minimum action for real clinical deployment would be to signal the classifier output on those days as less trustworthy.

5 Conclusion

We addressed the issue if cough counts would be affected largely by the epoch used for annotation. This might occur if cough sounds change in character over time and the extracted features and deployed classifier are sensitive to this. Data from a single patient was taken where a long-term trend was observable in the data and tested the hypothesis that its cause was a change in cough character. A classifier was developed based on annotation from two well-separated epochs (2 months). The cough count agreement was high (Pearson correlation of 0.93). From this singular case there is no clear evidence that the annotation epoch largely matters for the cough class, suggesting that the cough sounds (or at least the used parameters) may be considered as time-invariant. This is in line with results from earlier trials.

What is important, however, is that the training contains sufficient cough *and* environmental (non-cough) data. The data revealed particular days with non-cough acoustic events for which the classifier was not adequately trained in one of the two test conditions. A properly functioning monitoring system would require such periods to be identified and adequate measures to be taken. Capturing meta data (e.g., the amount of acoustic events per time unit or session) and comparing these with those from the training data is considered a promising option for identifying possible performance issues caused by discrepancies between training and deployment conditions.

Author contributions A.C. den Brinker: Conceptualization, Formal analysis, Writing - original draft & review & editing. R. Rietman: Investigation, Software, Writing - review & editing. O. Ouweltjes: Investigation, Software, Writing - review & editing. M. van Marion: Software. S. Thackray-Nocera: Data curation, Investigation. M.G. Crooks: Methodology, Writing - review & editing. A.H. Morice: Methodology, Writing - review & editing.

Funding A. H. Morice, M. G. Crooks, S. Thackray-Nocera report financial support provided by Philips Innovation Hub Eindhoven.

Data availability Not applicable.

Materials availability Not applicable.

Code availability Not applicable.

Declarations

Ethics approval and consent to participate The study was reviewed and approved by the North East-York Research Ethics Committee (REC Ref.: 21/YH/0203), the United Kingdom Health Research Authority and the Internal Committee Biomedical Experiments of Philips Research. The REC is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK. Informed consent was obtained from all individual participants included in the study.

Consent for publication The authors affirm that human research participants provided informed consent for publication.

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2000;161(5):1609–13. https://doi.org/10.1164/ajrccm.161.5.9908022.
2. Calverley P, Pauwels R, Löfdahl C-G, Svensson K, Higenbottam T, Carlsson L-G, Ståhl E. Relationship between respiratory symptoms and medical treatment in exacerbations of COPD. Eur Respir J. 2005;26(3):406–13. https://doi.org/10.1183/09031936.05.00143404.
3. Aaron SD, Donaldson GC, Whitmore GA, Hurst JR, Ramsay T, Wedzicha JA. Time course and pattern of COPD exacerbation onset. Thorax. 2012;67(3):238–43. https://doi.org/10.1136/thoraxjnl-2011-200768.
4. Rezende LC, Ribeiro EG, Parreiras LC, Guimarães RA, Reis GM, Carajá AF, Franco TB, Souza Mendes LP, Augusto VM, Silva KL. Telehealth and telemedicine in the management of adult patients after hospitalization for COPD exacerbation: a scoping review. J Bras Pneumol. 2023;49:3.
5. Crooks MG, Hayman Y, Innes A, Williamson J, Wright CE, Morice AH. Objective measurement of cough frequency during COPD exacerbation convalescence. Lung. 2016;194(1):117–20. https://doi.org/10.1007/s00408-015-9782-y.
6. Hall JI, Lozano M, Estrada-Petrocelli L, Birring S, Turner R. The present and future of cough counting tools. J Thoracic Dis 2020; https://doi.org/10.21037/jtd-2020-icc-003.

7. Serrurier A, Neuschaefer-Rube C, Röhrig R. Past and trends in cough sound acquisition, automatic detection and automatic classification: a comparative review. Sensors. 2022. https://doi.org/10.3390/s22082896.
8. Chen Y, Attri P, Barahona J, Hernandez ML, Carpenter D, Bozkurt A, Lobaton E. Robust cough detection with out-of-distribution detection. IEEE J Biomed Health Inform. 2023;27(7):3210–21. https://doi.org/10.1109/JBHI.2023.3264783.
9. den Brinker AC, van Dinther R, Crooks MG, Thackray-Nocera S, Morice AH. Alert system design based on experimental findings from long-term unobtrusive monitoring in COPD. Biomed Signal Process Contr 63, 2021;02205 https://doi.org/10.1016/j.bspc.2020.102205.
10. Crooks MG, Brinker AC, Thackray-Nocera S, Dinther R, Wright CE, Morice AH. Domiciliary cough monitoring for the prediction of COPD exacerbations. Lung. 2021;199(2):131–7. https://doi.org/10.1007/s00408-021-00435-9.
11. Crooks MG, Brinker A, Hayman Y, Williamson JD, Innes A, Wright CE, Hill P, Morice AH. Continuous cough monitoring using ambient sound recording during convalescence from a COPD exacerbation. Lung. 2017;195(3):289–94. https://doi.org/10.1007/s00408-017-9996-2.
12. Di Perna L, Spina G, Thackray-Nocera S, Crooks MG, Morice AH, Soda P, den Brinker AC. An automated and unobtrusive system for cough detection. In: 2017 IEEE Life Sciences Conference (LSC), 2017;p. 190–193 . https://doi.org/10.1109/LSC.2017.8268175.
13. den Brinker AC, Coman M, Ouweltjes O, Crooks MG, Thackray-Nocera S, Morice AH. Performance requirements for cough classifiers in real-world applications. In: 2020 28th European Signal Processing Conference (EUSIPCO); 2021. p. 96–100. https://doi.org/10.23919/Eusipco47968.2020.9287860.
14. den Brinker AC, Ouweltjes O, Rietman R, Thackray-Nocera S, Crooks MG, Morice AH. Nighttime cough characteristics in chronic obstructive pulmonary disease patients. Sensors. 2025;25(2):404. https://doi.org/10.3390/s25020404.